Sir Stanley Davidson (1894–1981)

This famous textbook was the brainchild of one of the great Professors of Medicine of the 20th century. Stanley Davidson was born in Sri Lanka and began his medical undergraduate training at Trinity College, Cambridge; this was interrupted by World War I and later resumed in Edinburgh. He was seriously wounded in battle, and the carnage and shocking waste of young life that he encountered at that time had a profound effect on his subsequent attitudes and values.

In 1930 Stanley Davidson was appointed Professor of Medicine at the University of Aberdeen, one of the first full-time Chairs of Medicine anywhere and the first in Scotland. In 1938 he took up the Chair of Medicine at Edinburgh and was to remain in this post until retirement in 1959. He was a renowned educator and a particularly gifted teacher at the bedside, where he taught that everything had to be questioned and explained. He himself gave most of the systematic lectures in Medicine, which were made available as typewritten notes that emphasised the essentials and far surpassed any textbook available at the time.

Principles and Practice of Medicine was conceived in the late 1940s with its origins in those lecture notes. The first edition, published in 1952, was a masterpiece of clarity and uniformity of style. It was of modest size and price, but sufficiently comprehensive and up to date to provide students with the main elements of sound medical practice. Although the format and presentation have seen many changes in 21 subsequent editions, Sir Stanley’s original vision and objectives remain. More than half a century after its first publication, his book continues to inform and educate students, doctors and health professionals all over the world.

Readers may be interested to listen to an interview with Sir Stanley Davidson, which can be found on the Royal College of Physicians of Edinburgh website at: www.rcpe.ac.uk/library-archives/sir-stanley-davidson-1894-1981.

Content Strategist: Laurence Hunter
Content Development Specialist: Wendy Lee
Project Manager: Louisa Talbott
Designer/Design Direction: Miles Hitchen
Illustration Manager: Jennifer Rose
Since Davidson's Principles and Practice of Medicine was first published in 1952, over two million copies have been sold and the book has acquired a large following of medical students, doctors and other health professionals all over the world. It has been translated into many languages, most recently Japanese, Russian, Italian and Polish, and has won numerous prizes, the last edition being highly commended in the British Medical Association Book Awards. Davidson’s has endured because with each new edition it has evolved to provide comprehensive updated information in a concise and easy-to-read format.

From its beginnings, Davidson’s has sought to explain the basis for medical practice. The integration of ‘preclinical’ science with clinical practice is now a feature of many undergraduate medical curricula, and many students use Davidson’s from the outset of their medical course. In recognition of this, the first part of the book, ‘Principles of Medicine’, highlights the mechanisms of health and disease, along with the professional and ethical principles underlying medical practice. Many examples of clinical problems are included to bring the medical sciences to life for the new student and to rejuvenate the interest of the experienced clinician. The second part of the book, ‘Practice of Medicine’, covers the major medical specialties. Every chapter has been thoroughly revised for this edition to ensure that it reflects the ‘cutting edge’ of medical knowledge and practice and is pitched at a level of detail to meet the needs of candidates preparing for examination for Membership of the Royal College of Physicians or its equivalents. In recognition of the emerging specialty of Stroke Medicine, this topic is now covered in a separate chapter from Neurological Disease. Surgical approaches to disease management are mentioned in Davidson's, but readers are encouraged to consult the sister book, Principles and Practice of Surgery, for more details.

Many of the innovations introduced in recent editions have been warmly received. We have retained both a patient-orientated approach, in the ever-popular ‘Clinical Examination’ overviews and ‘Presenting Problems’ sections, alongside practical content, in ‘Emergency’ and ‘Practice Point’ boxes. Embedding horizontal themes within the book – for example, with the ‘In Old Age’ and ‘In Pregnancy’ boxes – has been applauded, and we have extended this approach by adding ‘In Adolescence’ boxes in relevant chapters; these emphasise key points in managing the transition of patients between paediatric and adult services.

We are proud of Davidson’s international heritage. As well as recruiting authors from around the globe, particularly for topics such as Infectious Diseases and HIV, we have welcomed new members on to our International Advisory Board. These leading experts from 16 countries provide useful comments that, along with the feedback received from our global readership, are crucial to our planning of every chapter in each new edition. We have also visited several medical schools on the Indian subcontinent and received invaluable feedback from students and teachers. We have addressed as many of these suggestions as possible in this edition.

Education is achieved by assimilating information from many sources and readers of this book can enhance their learning experience by using complementary resources. The StudentConsult platform continues to provide online access to the text and illustrations of the main edition. The book is also available in various eBook formats. Davidson’s has had a long-standing association with its sister books, Macleod’s Clinical Examination (now in its 13th Edition) and Principles and Practice of Surgery (now in its 6th Edition). The Davidson’s ‘family’ has expanded with the publication of Davidson’s Essentials of Medicine, a long-requested pocket-size version of the main text; Davidson’s Foundations of Clinical Practice, a guide to starting work as a junior doctor; Davidson’s 100 Clinical Cases, which contains cases directly based on the ‘Presenting Problems’ in the main text; and Macleod’s Clinical Diagnosis, which describes a systematic approach to differential diagnosis of symptoms and signs. We congratulate the editors and authors of these books for continuing the tradition of concise, easily read and beautifully illustrated texts.

The regular introduction of new authors and editors to Davidson’s is important to maintain the freshness of each new edition. On this occasion, Dr Ian Penman has joined the editorial team and 18 new authors have contributed material. We all take immense pride in producing an outstanding book for the next generation of doctors, and in continuing the great tradition first established by Sir Stanley Davidson and passed on by all the previous editors and authors, for what remains one of the world’s leading textbooks of medicine.

BRW, NRC, SHR, IDP
Edinburgh
**Contents**

| List of presenting problems | viii |
| Contributors | x |
| International Advisory Board | xiv |
| Introduction | xvi |
| Acknowledgements | xviii |
| Figure acknowledgements | xix |

## PART 1  PRINCIPLES OF MEDICINE

<table>
<thead>
<tr>
<th>1</th>
<th>Good medical practice</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.D. Cumming • S.I.R. Noble</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Therapeutics and good prescribing</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. Maxwell</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>Molecular and genetic factors in disease</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.R. FitzPatrick • J.R. Seckl</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
<th>Immunological factors in disease</th>
<th>71</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.E. Marshall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Environmental and nutritional factors in disease</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. Hanlon • M. Byers • J.P.H. Wilding • H.M. Macdonald</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>Principles of infectious disease</th>
<th>133</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.P. Hobson • D.H. Dockrell</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>Ageing and disease</th>
<th>165</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.D. Witham</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## PART 2  PRACTICE OF MEDICINE

<table>
<thead>
<tr>
<th>8</th>
<th>Critical illness</th>
<th>179</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.R. Nimmo • T. Walsh</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9</th>
<th>Poisoning</th>
<th>205</th>
</tr>
</thead>
<tbody>
<tr>
<td>S.H.L. Thomas • J. White</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10</th>
<th>Medical psychiatry</th>
<th>231</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.C. Sharpe • S.M. Lawrie</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11</th>
<th>Oncology</th>
<th>259</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.G. Dark • A.R. Abdul Razak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>12</td>
<td>Palliative care and pain</td>
<td>283</td>
</tr>
<tr>
<td>13</td>
<td>Infectious disease</td>
<td>293</td>
</tr>
<tr>
<td>14</td>
<td>HIV infection and AIDS</td>
<td>387</td>
</tr>
<tr>
<td>15</td>
<td>Sexually transmitted infections</td>
<td>411</td>
</tr>
<tr>
<td>16</td>
<td>Clinical biochemistry and metabolism</td>
<td>427</td>
</tr>
<tr>
<td>17</td>
<td>Kidney and urinary tract disease</td>
<td>461</td>
</tr>
<tr>
<td>18</td>
<td>Cardiovascular disease</td>
<td>525</td>
</tr>
<tr>
<td>19</td>
<td>Respiratory disease</td>
<td>643</td>
</tr>
<tr>
<td>20</td>
<td>Endocrine disease</td>
<td>733</td>
</tr>
<tr>
<td>21</td>
<td>Diabetes mellitus</td>
<td>797</td>
</tr>
<tr>
<td>22</td>
<td>Alimentary tract and pancreatic disease</td>
<td>837</td>
</tr>
<tr>
<td>23</td>
<td>Liver and biliary tract disease</td>
<td>921</td>
</tr>
<tr>
<td>24</td>
<td>Blood disease</td>
<td>989</td>
</tr>
<tr>
<td>25</td>
<td>Rheumatology and bone disease</td>
<td>1057</td>
</tr>
<tr>
<td>26</td>
<td>Neurological disease</td>
<td>1137</td>
</tr>
<tr>
<td>27</td>
<td>Stroke disease</td>
<td>1231</td>
</tr>
<tr>
<td>28</td>
<td>Skin disease</td>
<td>1249</td>
</tr>
<tr>
<td>29</td>
<td>Laboratory reference ranges</td>
<td>1307</td>
</tr>
</tbody>
</table>

Index 1313
List of presenting problems

These presentations represent the most common reasons for referral to each medical specialty and are described in the ‘Presenting Problems’ sections of all system-based chapters. The same approach has also been employed in several of the chapters in the ‘Principles of Medicine’ section, reinforcing the close connection between clinical problems and fundamental mechanisms of disease.

### Abnormal investigation results

<table>
<thead>
<tr>
<th>Abnormal investigation result</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid–base, 443</td>
<td></td>
</tr>
<tr>
<td>Metabolic acidosis, 445</td>
<td></td>
</tr>
<tr>
<td>Metabolic alkalosis, 446</td>
<td></td>
</tr>
<tr>
<td>Mixed abnormalities, 447</td>
<td></td>
</tr>
<tr>
<td>Respiratory acidois, 447</td>
<td></td>
</tr>
<tr>
<td>Respiratory alkalosis, 447</td>
<td></td>
</tr>
<tr>
<td>Blood culture</td>
<td></td>
</tr>
<tr>
<td>Positive, 303</td>
<td></td>
</tr>
<tr>
<td>Electrolytes</td>
<td></td>
</tr>
<tr>
<td>Hypercalcaemia, 767</td>
<td></td>
</tr>
<tr>
<td>Hypocalcaemia, 768</td>
<td></td>
</tr>
<tr>
<td>Hyperkalaemia, 442</td>
<td></td>
</tr>
<tr>
<td>Hypokalaemia, 440</td>
<td></td>
</tr>
<tr>
<td>Hypermagnesaemia, 448</td>
<td></td>
</tr>
<tr>
<td>Hypomagnesaemia, 448</td>
<td></td>
</tr>
<tr>
<td>Hypernatraemia, 439</td>
<td></td>
</tr>
<tr>
<td>Hyponatraemia, 437</td>
<td></td>
</tr>
<tr>
<td>Hyperphosphataemia, 449</td>
<td></td>
</tr>
<tr>
<td>Hypophosphataemia, 448</td>
<td></td>
</tr>
<tr>
<td>Erythrocyte sedimentation rate</td>
<td></td>
</tr>
<tr>
<td>Raised, 85</td>
<td></td>
</tr>
<tr>
<td>Full blood count</td>
<td></td>
</tr>
<tr>
<td>Anaemia, 1001</td>
<td></td>
</tr>
<tr>
<td>Leucocytosis, 1005</td>
<td></td>
</tr>
<tr>
<td>Leucopenia, 1004</td>
<td></td>
</tr>
<tr>
<td>Pancytopenia, 1008</td>
<td></td>
</tr>
<tr>
<td>Polycythaemia, 1003</td>
<td></td>
</tr>
<tr>
<td>Thrombocytopenia, 1007</td>
<td></td>
</tr>
<tr>
<td>Thrombocytosis, 1008</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
</tr>
<tr>
<td>Hyperglycaemia, 808, 818</td>
<td></td>
</tr>
<tr>
<td>Hypoglycaemia, 783, 814</td>
<td></td>
</tr>
<tr>
<td>Hormones</td>
<td></td>
</tr>
<tr>
<td>Hyperprolactinaemia, 790</td>
<td></td>
</tr>
<tr>
<td>Hypogonadism, male, 760</td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism, 743</td>
<td></td>
</tr>
<tr>
<td>Thyroid function tests</td>
<td></td>
</tr>
<tr>
<td>Asymptomatic abnormalities, 745</td>
<td></td>
</tr>
<tr>
<td>Thyrotoxicosis, 740</td>
<td></td>
</tr>
<tr>
<td>Lipids, 453</td>
<td></td>
</tr>
<tr>
<td>Hypercholesterolaemia, 453</td>
<td></td>
</tr>
<tr>
<td>Hypertriglyceridaemia, 455</td>
<td></td>
</tr>
<tr>
<td>Mixed hyperlipidaemia, 455</td>
<td></td>
</tr>
<tr>
<td>Rare hyperlipidaemia, 455</td>
<td></td>
</tr>
<tr>
<td>Liver function tests, 935</td>
<td></td>
</tr>
<tr>
<td>Proteinuria, 476</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td></td>
</tr>
<tr>
<td>Adrenal mass, 779</td>
<td></td>
</tr>
<tr>
<td>Incidental pulmonary nodule, 660</td>
<td></td>
</tr>
<tr>
<td>Pituitary tumour, 789</td>
<td></td>
</tr>
</tbody>
</table>

### Symptoms and signs

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amenorrhoea, 759</td>
<td></td>
</tr>
<tr>
<td>Ascites, 938</td>
<td></td>
</tr>
<tr>
<td>Blackouts, 554, 1157</td>
<td></td>
</tr>
<tr>
<td>Bleeding</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal, 853, 942</td>
<td></td>
</tr>
<tr>
<td>Generalised, 201, 1006</td>
<td></td>
</tr>
<tr>
<td>Breathlessness, 289, 543, 655</td>
<td></td>
</tr>
<tr>
<td>Chest pain, 539, 658</td>
<td></td>
</tr>
<tr>
<td>Coma, 198, 1159, 1237</td>
<td></td>
</tr>
<tr>
<td>Constipation, 860</td>
<td></td>
</tr>
<tr>
<td>Cough, 289, 654</td>
<td></td>
</tr>
<tr>
<td>Deafness, 1173</td>
<td></td>
</tr>
<tr>
<td>Diarrhoea, 306, 857</td>
<td></td>
</tr>
<tr>
<td>Dizziness, 173, 554, 1157, 1167</td>
<td></td>
</tr>
<tr>
<td>Dyspepsia, 852</td>
<td></td>
</tr>
<tr>
<td>Dysphagia, 851, 1173</td>
<td></td>
</tr>
<tr>
<td>Erectile dysfunction, 474</td>
<td></td>
</tr>
<tr>
<td>Falls, 172, 554, 1157</td>
<td></td>
</tr>
<tr>
<td>Fever, 296</td>
<td></td>
</tr>
<tr>
<td>In immunocompromised host, 301, 396</td>
<td></td>
</tr>
<tr>
<td>In injection drug user, 299</td>
<td></td>
</tr>
<tr>
<td>In neutropenic patient, 302</td>
<td></td>
</tr>
<tr>
<td>With weight loss, 271</td>
<td></td>
</tr>
<tr>
<td>Finger clubbing, 271</td>
<td></td>
</tr>
<tr>
<td>Gait abnormalities, 1168</td>
<td></td>
</tr>
<tr>
<td>Galactorrhea, 790</td>
<td></td>
</tr>
<tr>
<td>Genital problems, 415, 417</td>
<td></td>
</tr>
<tr>
<td>Itch/pain in women, 418</td>
<td></td>
</tr>
<tr>
<td>Itch/rash in men, 415</td>
<td></td>
</tr>
<tr>
<td>Lumps in men, 416</td>
<td></td>
</tr>
<tr>
<td>Lumps in women, 418</td>
<td></td>
</tr>
<tr>
<td>Ulceration in men, 415</td>
<td></td>
</tr>
<tr>
<td>Ulceration in women, 418</td>
<td></td>
</tr>
<tr>
<td>Urethral discharge, 415</td>
<td></td>
</tr>
<tr>
<td>Vaginal discharge, 417</td>
<td></td>
</tr>
<tr>
<td>Goitre, 746</td>
<td></td>
</tr>
<tr>
<td>Gynaecomastia, 762</td>
<td></td>
</tr>
<tr>
<td>Haemoptysis, 658</td>
<td></td>
</tr>
<tr>
<td>Hair and nail problems, 1264</td>
<td></td>
</tr>
<tr>
<td>Heart murmurs and abnormal heart sounds, 560</td>
<td></td>
</tr>
<tr>
<td>Heartburn, 852</td>
<td></td>
</tr>
<tr>
<td>Hepatomegaly, 938</td>
<td></td>
</tr>
<tr>
<td>Hirsutism, 763</td>
<td></td>
</tr>
<tr>
<td>Hypertension, 478</td>
<td></td>
</tr>
<tr>
<td>Incontinence</td>
<td></td>
</tr>
<tr>
<td>Faecal, 1174</td>
<td></td>
</tr>
<tr>
<td>Urinary, 175, 472, 1174</td>
<td></td>
</tr>
<tr>
<td>Infertility, 761</td>
<td></td>
</tr>
<tr>
<td>Jaundice, 936</td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy, 395, 1005</td>
<td></td>
</tr>
<tr>
<td>Memory loss, 1161</td>
<td></td>
</tr>
<tr>
<td>Movement and coordination problems, 1165, 1237</td>
<td></td>
</tr>
</tbody>
</table>
Nausea, 289, 306, 853
Oedema, 478
Pain, 284
  Abdominal, 418, 861
  Ankle and foot, 1076
  Back, 1072
  Elbow, 1075
  Generalised musculoskeletal, 1071, 1076
  Hand and wrist, 1075
  Hip, 1075
  Knee, 1075
  Loin, 471
  Multiple joints, 1069
  Neck, 1074
  Regional musculoskeletal, 1074
  Shoulder, 1074
  Single joint, 1069
Palpable mass, 270
Palpitation, 556
Pleural effusion, 661
Psychological
  Abnormal perception, 1167
  Anxiety, 234, 290
  Confusion, agitation, delirium, 173, 237, 238, 290, 1161, 1175
  Delusions, 236, 1175
  Depressed mood, 235, 290, 1175
  Elated mood, 235
  Hallucinations, 236, 1175
  Personality change, 1175
Puberty
  Delayed, 758
Seizures, 1159, 1237
Sensory disturbance, 1164
Skin problems, 1256
  Blisters, 1258
  Colour change, 1263
  Itch (pruritus), 415, 418, 1258
  Leg ulcers, 1262
  Lumps and changing lesions, 1256
  Papulosquamous rashes, 1257
  Photosensitivity, 1260
Sleep disturbance, 1175
Smell disturbance, 1169
Speech disturbance, 1168, 1236
Splenomegaly, 1006
Syncope see Falls and Blackouts
Urinary symptoms, 471
  Dysuria, 471
  Frequency, 472
  Haematuria, 474
  Incontinence, 175, 472, 1174
  Nocturia, 472
  Oliguria/anuria, 471
  Polyuria, 472
Visual disturbance, 1169, 1237
Vomiting, 289, 306, 853
Weakness, 290, 1076, 1162, 1236
Weight loss, 271, 290, 396, 859

Syndromes

Adrenal insufficiency, 777
Ageing problems, 176
Alcohol misuse, 240
Allergy, 90
Anaphylaxis, 91, 190
Angioedema, 93
Cardiac arrest, 557
Circulatory failure
  Anaphylaxis, 91, 190
  Shock, 190, 544
Cushing’s syndrome, 773
Diabetes mellitus, 808
  Complications, 820
  Hyperosmolality, 814
  In pregnancy, 817
  In young patients, 818
  Ketoadidosis, 811
  Long-term supervision, 811
  Newly discovered, 808
  Peri-operative, 818
Disseminated intravascular coagulation, 201, 1007
Drug reactions
  Adverse reactions, 175
  Glucocorticoids/corticosteroids, 776
Ectopic hormone production, 271
Fracture, 1071
Gastrointestinal and liver abnormalities in critical illness, 198
Gastrointestinal obstruction in terminal illness, 290
Heart failure, 546
Hepatic encephalopathy, 941
HIV/AIDS manifestations, 395
  Cardiac, 405
  Gastrointestinal, 399
  Haematological, 404
  Liver, 400
  Mucocutaneous, 396
  Neoplasms, 405
  Nervous system, 402
  Ophthalmic, 402
  Renal, 405
  Respiratory, 400
  Rheumatological, 404
Hypopituitarism, 787
Infection manifestations
  In adolescence, 313
  In blood disease, 1008
  In pregnancy, 313
  In the tropics, 308
  Recurrent, 79
Liver failure
  Acute, 932
Malabsorption, 857
Paraneoplastic syndromes, 271
Proctitis, 417
Psychological
  Medically unexplained symptoms, 236
  Psychological factors affecting medical conditions, 240
Renal failure
  Acute, 197, 478
  Chronic, 483
Respiratory failure, 191, 663
Self-harm, 238
Sepsis, 200, 304
Skin
  Acute failure, 1264
  Skin manifestations of cancer, 272
  Sodium depletion, 432
  Sodium excess, 434
  Substance misuse, 240
  Sudden death, 557
  Venous thrombosis, 1008
Contributors

Albiruni Ryan Abdul Razak
MRCPi
Consultant Medical Oncologist, Princess Margaret Cancer Centre, Toronto; Assistant Professor, University of Toronto, Canada

Brian J. Angus
BSc DTM&H FRCP MD FFTM(Glas)
Reader in Infectious Diseases, Nuffield Department of Medicine, University of Oxford; Director, Oxford Centre for Tropical Medicine, UK

Quentin M. Anstee
BSc MBBS PhD MRCP(UK)
Senior Lecturer, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne; Honorary Consultant Hepatologist, Freeman Hospital, Newcastle upon Tyne, UK

Andrew W. Bradbury
BSc MBChB(Hons) MD MBA FEBVS(Hon) FRCS
Sampson Gamgee Professor of Vascular Surgery, Director of Quality Assurance and Enhancement, College of Medical and Dental Sciences, University of Birmingham, UK

Leslie Burnett
MBBS PhD FRCPA FHGSA
Consultant Pathologist, NSW Health, PaLMS Pathology North, Royal North Shore Hospital, Sydney; Clinical Professor in Pathology and Genetic Medicine, Sydney Medical School, University of Sydney, Australia

Mark Byers
OBE FRCP MCEM MFSEM DA(UK)
General Practitioner, Ministry of Defence, UK

Jenny I.O. Craig
MD FRCP FRCPath
Consultant Haematologist, Addenbrooke’s Hospital, Cambridge, UK

Allan D. Cumming
BSc MD FRCP
Dean of Students, College of Medicine and Veterinary Medicine, University of Edinburgh, UK

Graham Dark
MBBS FRCP FHEA
Senior Lecturer in Cancer Education, Newcastle University; Consultant Medical Oncologist, Freeman Hospital, Newcastle upon Tyne, UK

Richard J. Davenport
FRCPE DM
Consultant Neurologist, Royal Infirmary of Edinburgh and Western General Hospital, Edinburgh; Honorary Senior Lecturer, University of Edinburgh, UK

Robert S. Dawe
MD FRCP
Consultant Dermatologist, Ninewells Hospital and Medical School, Dundee; Honorary Clinical Reader, University of Dundee, UK

David Dockrell
MD FRCP FRCPG FACP
Professor of Infectious Diseases, University of Sheffield, UK

Michael J. Field
AM MD FRACP
Emeritus Professor, Sydney Medical School, University of Sydney, Australia

David R. FitzPatrick
MD FRCP
Consultant in Clinical Genetics, Royal Hospital for Sick Children, Edinburgh; Professor, University of Edinburgh, UK

Jane Goddard
PhD FRCP
Consultant Nephrologist, Royal Infirmary of Edinburgh; Part-time Senior Lecturer, University of Edinburgh, UK

Neil R. Grubb
MD FRCP
Consultant Cardiologist, Edinburgh Heart Centre; Honorary Senior Lecturer, University of Edinburgh, UK
Phil Hanlon
BSc MD MPH
Professor of Public Health, University of Glasgow, UK

Richard P. Hobson
PhD MCRP(UK), FRCPath
Consultant Microbiologist, Leeds Teaching Hospitals NHS Trust; Honorary Senior Lecturer, Leeds University, UK

Sally H. Ibbotson
BSc(Hons) MD FRCPE
Clinical Senior Lecturer in Photobiology, University of Dundee; Honorary Consultant Dermatologist, Ninewells Hospital and Medical School, Dundee, UK

J. Alastair Innes
BSc MBChB PhD FRCPE
Consultant Physician, Western General Hospital, Edinburgh; Honorary Reader in Respiratory Medicine, University of Edinburgh, UK

David E. Jones
MA BM BCh PhD FRCP
Professor of Liver Immunology, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne; Consultant Hepatologist, Freeman Hospital, Newcastle upon Tyne, UK

Peter Langhorne
PhD FRCPG
Professor of Stroke Care, University of Glasgow; Honorary Consultant, Royal Infirmary, Glasgow, UK

Stephen M. Lawrie
MD(Hons) FRCPsych FRCP(Hon)
Head, Division of Psychiatry, School of Clinical Sciences, University of Edinburgh; Honorary Consultant Psychiatrist, Royal Edinburgh Hospital, UK

John Paul Leach
MD FRCPG FRCP
Consultant Neurologist, Institute of Neuroscience, Southern General Hospital, Glasgow; Honorary Associate Clinical Professor, University of Glasgow, UK

Charlie W. Lees
MBBS FRCP PhD
Consultant Gastroenterologist, Western General Hospital, Edinburgh; Honorary Senior Lecturer, University of Edinburgh, UK

Gary Maartens
MBChB MMed
Consultant Physician, Department of Medicine, Groote Schuur Hospital, Cape Town; Professor of Clinical Pharmacology, University of Cape Town, South Africa

Helen M. Macdonald
BSc(Hons) PhD MSc RNutr(Public Health)
Chair in Nutrition and Musculoskeletal Health, University of Aberdeen, UK

Lynn M. Manson
MD FRCP FRCPath
Consultant Haematologist, Scottish National Blood Transfusion Service; Honorary Clinical Senior Lecturer, University of Edinburgh, UK

Sara E. Marshall
FRCP FRCPath PhD
Honorary Consultant Immunologist, NHS Tayside; Professor of Clinical Immunology, University of Dundee, UK

Simon Maxwell
MD PhD FRCP FRCPE FBPharmacoIS FHEA
Professor of Student Learning (Clinical Pharmacology and Prescribing), University of Edinburgh; Honorary Consultant Physician, Western General Hospital, Edinburgh, UK

Rory J. McCrimmon
MD FRCP
Professor of Experimental Diabetes and Metabolism, University of Dundee; Honorary Consultant, Ninewells Hospital and Medical School, Dundee, UK

Iain B. McInnes
FRCP PhD FRSE FMedSci
Muirhead Professor of Medicine and Director of Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK

David E. Newby
FESC FACC FMedSci FRSE
British Heart Foundation John Wheatley Professor of Cardiology, University of Edinburgh; Consultant Cardiologist, Royal Infirmary of Edinburgh, UK

John D. Newell-Price
MA PhD FRCP
Reader in Endocrinology and Honorary Consultant Endocrinologist, Department of Human Metabolism, School of Medicine and Biomedical Science, Sheffield, UK

Graham R. Nimmo
MD FRCP FFArCSI FFICM
Consultant Physician in Intensive Care Medicine and Clinical Education, Western General Hospital, Edinburgh, UK
Contributors

Simon I.R. Noble
MBBS FRCP Dip Pal Med PGCE
Clinical Senior Lecturer in Palliative Medicine, Cardiff University; Honorary Consultant, Palliative Medicine, Royal Gwent Hospital, Newport, UK

David R. Oxenham
MRCP
Consultant in Palliative Care, Marie Curie Hospice, Edinburgh; Honorary Senior Lecturer, University of Edinburgh, UK

Ewan R. Pearson
PhD FRCP
Professor of Diabetic Medicine, University of Dundee; Honorary Consultant in Diabetes, Ninewells Hospital and Medical School, Dundee, UK

Ian D. Penman
BSc MD FRCP
Consultant Gastroenterologist, Royal Infirmary of Edinburgh; Honorary Senior Lecturer, University of Edinburgh, UK

Stuart H. Ralston
MD FRCP FMedSci FRSE
Arthritis Research UK Professor of Rheumatology, University of Edinburgh; Honorary Consultant Rheumatologist, Western General Hospital, Edinburgh, UK

Peter T. Reid
MD FRCP
Consultant Physician, Western General Hospital, Edinburgh; Honorary Senior Lecturer in Respiratory Medicine, University of Edinburgh, UK

Gordon R. Scott
BSc FRCP
Consultant in Genitourinary Medicine, Chalmers Sexual Health Centre, Edinburgh; Honorary Senior Lecturer, University of Edinburgh, UK

Jonathan R. Seckl
BSc MBBS PhD FRCP FMedSci FRSE
Professor of Molecular Medicine, Executive Dean (Medicine) and Vice-Principal (Research), University of Edinburgh; Honorary Consultant Physician, Royal Infirmary of Edinburgh, UK

Michael Sharpe
MA MD FRCP FRCPsych
Professor of Psychological Medicine, University of Oxford; Honorary Professor, University of Edinburgh; Honorary Consultant in Psychological Medicine, Oxford University Hospitals NHS Trust and Oxford Health NHS Foundation Trust, UK

Peter Stewart
FRACP FRCPA MBA
Clinical Director, Sydney South West Pathology Service; Clinical Associate Professor, University of Sydney, Australia

Mark W.J. Strachan
MD FRCP
Consultant in Diabetes and Endocrinology, Western General Hospital, Edinburgh; Honorary Professor, University of Edinburgh, UK

David Sullivan
FRACP FRCPA FCSANZ
Clinical Associate Professor, Central Clinical School, Sydney Medical School, University of Sydney, Australia

Shyam Sundar
MD FRCP FNA
Professor of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

Simon H.L. Thomas
BSc MD FRCP FRCP
Consultant Physician, Newcastle Hospitals NHS Foundation Trust; Professor of Clinical Pharmacology and Therapeutics, Newcastle University, Newcastle upon Tyne, UK

A. Neil Turner
PhD FRCP
Professor of Nephrology, University of Edinburgh; Honorary Consultant Physician, Royal Infirmary of Edinburgh, UK

Simon W. Walker
DM FRCPth FRCPath FRCP
Senior Lecturer in Clinical Biochemistry, University of Edinburgh; Honorary Consultant Clinical Biochemist, Royal Infirmary of Edinburgh, UK

Tim Walsh
BSc(Hons) MBChB(Hons) FRCA FRCP FFICM MD MRes
Professor of Critical Care, University of Edinburgh; Honorary Consultant, Royal Infirmary of Edinburgh, UK
Henry Watson
MD FRCPE FRCPath
Consultant Haematologist, Aberdeen Royal Infirmary; Honorary Senior Lecturer, University of Aberdeen, UK

Julian White
MD FACTM
Consultant Clinical Toxinologist and Head of Toxinology, Women’s and Children’s Hospital, Adelaide; Associate Professor, Department of Paediatrics, University of Adelaide, Australia

John PH. Wilding
DM FRCP
Professor of Medicine, Head of Department of Obesity and Endocrinology, Institute of Ageing and Chronic Disease, University of Liverpool; Honorary Consultant Physician, University Hospital Aintree, Liverpool, UK

Miles D. Witham
BM BCh PhD FRCPE
Clinical Senior Lecturer in Ageing and Health, University of Dundee; Honorary Consultant Geriatrician, NHS Tayside, Dundee, UK
International Advisory Board

O.C. Abraham
Professor, Department of Medicine, Christian Medical College, Vellore, India

Tofayel Ahmed
Professor of Medicine, Somaj Vittik Medical College, Gono Bishwabidyalay, Dhaka; Professor of Medicine, Pioneer Dental College, Dhaka, Bangladesh

Samar Banerjee
Professor, Department of Medicine, Vivekananda Institute of Medical Sciences and Ramakrishna Mission Seva Pratishthan, Kolkata, India

Matthew A. Brown
Professor of Immunogenetics and Director, University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, Australia

Khalid I. Bzeizi
Senior Consultant and Head of Hepatology, Prince Sultan Military Medical City, Riyadh, Saudi Arabia

M.K. Daga
Director, Professor of Medicine and In-Charge ICU, Maulana Azad Medical College, New Delhi, India

D. Dalus
Professor and Head, Department of Internal Medicine, Medical College and Hospital, Trivandrum, India

Tapas Das
Professor, Department of Medicine, KPC Medical College and Hospital, Jadavpur, Kolkata, India

Tarun Kumar Dutta
Head, Division of Clinical Haematology; Professor and Head, Department of Medicine, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Puducherry, India

M. Abul Faiz
Professor of Medicine (Retired), Sir Salimullah Medical College, Mitford, Dhaka, Bangladesh

Albert G. Frauman
Professor of Clinical Pharmacology and Therapeutics, University of Melbourne; Medical Director, Austin Centre for Clinical Studies, Austin Health, Heidelberg, Victoria, Australia

Tsuguya Fukui
President, St Luke’s International Hospital, Tokyo, Japan

Hadi A. Goubran
Haematologist, Saskatoon Cancer Centre and Adjunct Professor, College of Medicine, University of Saskatchewan, Canada; Professor of Medicine and Haematology, Cairo University, Egypt

Saman B. Gunatilake
Professor of Medicine, Faculty of Medical Sciences, University of Sri Jayewardenepura, Sri Lanka

Rajiva Gupta
Director and Head, Rheumatology and Clinical Immunology, Medanta – The Medicity, Gurgaon, India

S.M. Wasim Jafri
Alticharn Professor of Medicine, Associate Dean, Aga Khan University, Karachi, Pakistan

Saroj Jayasinghe
Professor, Department of Clinical Medicine, Faculty of Medicine, University of Colombo; Honorary Consultant Physician, National Hospital of Sri Lanka, Colombo, Sri Lanka

A.L. Kakrani
Professor and Head, Department of Medicine, Padmashree Dr D.Y. Patil Medical College and Hospital, Pimpri, Pune, India

Vasantha Kamath
Director-Dean, Karnataka Institute of Medical Sciences (KIMS), Hubli, Karnataka, India

Piotr Kuna
Professor of Medicine, Department of Internal Medicine, Asthma and Allergy, Barlicki University Hospital, Medical University of Lodz, Poland
Ammar F. Mubaidin
Associate Professor of Neurology, King Hussein Medical Center, Amman, Jordan

Milind Nadkar
Professor of Medicine (Emergency Medicine) and Chief, Rheumatology Services, Seth G.S. Medical College and KEM Hospital, Mumbai, India

Saraladevi Naicker
Professor and Head, Division of Nephrology; Academic Head, Department of Internal Medicine, University of the Witwatersrand, Johannesburg, South Africa

Nardeep Naithani
Professor and Head, Department of Internal Medicine, Armed Forces Medical College, Pune, India

Matthew Ng
Honorary Clinical Professor, University of Hong Kong, Tung Wah Hospital, Hong Kong

Tommy Olsson
Professor of Medicine, Department of Medicine, Umeå University Hospital, Umeå, Sweden

Prem Pais
Dean and Professor of Medicine, St John’s Medical College, Bangalore, India

A. Ramachandran
President, India Diabetes Research Foundation, and Chairman, Dr A. Ramachandran’s Diabetes Hospitals, Chennai, India; Visiting Professor, Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Imperial College, London, UK

Medha Y. Rao
Professor of Internal Medicine and Associate Dean, M.S. Ramaiah Medical College and Hospital, Bangalore, India

N.R. Rau
Professor of Medicine, Department of Medicine, Kasturba Hospital, Manipal, India

Harsha R. Salkar
Professor and Head, Department of Medicine, NKP Salve Institute of Medical Sciences, Nagpur, India

K.R. Sethuraman
Vice-Chancellor, Sri Balaji Vidyapeeth University, Pondicherry, India

Surendra K. Sharma
Chief, Division of Pulmonary and Critical Care Medicine; Professor and Head, Department of Medicine, All India Institute of Medical Sciences, New Delhi, India

Ibrahim Sherif
Emeritus Professor of Medicine, Tripoli University; Consultant Endocrinologist, Alafia Clinic, Tripoli, Libya

Ian J. Simpson
Emeritus Professor of Medicine, Faculty of Medical and Health Sciences, University of Auckland; Nephrologist, Auckland City Hospital, Auckland, New Zealand

Dr S.G. Siva Chidambaram
Professor of Medicine, Institute of Internal Medicine, Madras Medical College and Rajiv Gandhi Government General Hospital, Chennai, India

Arvind K. Vaish
Professor and Head, Department of Medicine, King George’s Medical University, Lucknow, India

Subhash Varma
Professor and Head, Department of Internal Medicine, Postgraduate Institute of Medical Education and Research, Chandigarh, India
Introduction

The first section of the book, ‘Principles of Medicine’, describes the basis on which medicine is practised and the fundamental mechanisms determining health and disease which are relevant to all medical specialties. The second section, ‘Practice of Medicine’, is devoted to individual medical specialties. Each chapter has been written by experts in the field to provide the level of detail expected of trainees in their discipline. To maintain the book’s virtue of being concise, care has been taken to avoid unnecessary duplication between chapters.

The system-based chapters follow a standard format, beginning with an overview of relevant clinical examination, followed by an account of functional anatomy, physiology and investigations, then the common presentations of disease, and details of the individual diseases and treatments of that system. Where appropriate, the chapters in the first section follow a similar format; in chapters which describe the immunological, cellular and molecular basis of disease, this problem-based approach brings the close links between modern medical science and clinical practice into sharp focus.

The methods used to present information are described below.

Boxes

Boxes are a popular way of presenting information and are particularly useful for revision. They are classified by the type of information they contain, using specific symbols.

General Information

These include causes, clinical features, investigations, treatments and other useful information.

Evidence-based Medicine

Clinicians base their practice on the best available evidence, which needs to be up to date, relevant, authoritative and easily accessible. Over 120 evidence-based medicine (EBM) boxes are included in this edition. They contain recommendations that are supported by evidence obtained from meta-analysis of several randomised controlled trials (RCTs) or one (or more) high-quality RCT, and therefore conform to ‘Grade A’ criteria, as described in Chapter 1 (p. 8).

Practice Point

There are many practical skills that students and doctors must learn. These vary from inserting a nasogastric tube to reading an ECG or X-ray, or interpreting investigations such as arterial blood gases or thyroid function tests. ‘Practice Point’ boxes provide straightforward guidance on how these and many other skills can be acquired and applied.

Emergency

These boxes describe management of many of the most common emergencies in medicine.
In Old Age

In most developed countries, older people comprise 20% of the population and are the chief users of health care. While they contract the same diseases as those who are younger, there are often important differences in the way they present and how they are best managed.

Chapter 7, ‘Ageing and Disease’, concentrates on the principles of managing the frailest group who suffer from multiple comorbidity and disability, and who tend to present with non-specific problems such as falls or delirium. However, many older people also suffer from specific single-organ pathology. ‘In Old Age’ boxes are thus included in each chapter and describe common presentations, implications of physiological changes of ageing, effects of age on investigations, problems of treatment in old age, and the benefits and risks of intervention in older people.

In Pregnancy

Many conditions are different in the context of pregnancy, while some arise only during or shortly after pregnancy. Particular care must be taken with investigations (for example, to avoid radiation exposure to the fetus) and treatment (to avoid the use of drugs which harm the fetus). These issues are highlighted by ‘In Pregnancy’ boxes distributed throughout the book.

In Adolescence

Although Paediatric Medicine is not covered in Davidson’s, many chronic disorders begin in childhood and adult physicians often contribute to multidisciplinary teams that manage young patients ‘in transition’ between paediatric and adult health-care services. This group of patients often presents a particular challenge, due to the physiological and psychological changes that occur in adolescence and which can have a major impact on the disease and its management. Adolescents can be encouraged to take over responsibility from their parents/carers in managing their disease, but are naturally rebellious and often struggle to adhere to the impositions of chronic treatment. To highlight these issues, we have introduced this new box format in the 22nd Edition.

Terminology

Recommended International Non-proprietary Names (rINNs) are used for all drugs, with the exception of adrenaline and noradrenaline. However, British spellings have been retained for drug classes and groups (e.g. amphetamines not amfetamines).

Units of measurement

The International System of Units (SI units) is the recommended means of presentation for laboratory data and has been used throughout Davidson’s. However, we recognise that many laboratories around the world continue to provide data in non-SI units, so these have been included in the text for the commonly measured analytes. Both SI and non-SI units are also given in Chapter 29, which describes the reference ranges used in Edinburgh’s laboratories. It should be appreciated that these reference ranges may vary from those used in other laboratories.

Finding what you are looking for

A contents list is given on the opening page of each chapter. In addition, the book contains numerous cross-references to help readers find their way around, along with an extensive index of over 15000 subject entries. The online text available on StudentConsult (www.studentconsult.com) allows for detailed searches of the content by keyword. A list of up-to-date reviews and useful websites with links to management guidelines appears at the end of each chapter.
We are indebted to former authors who have stepped down from this edition. They include Dr Chris M.C. Allen, Dr Jeffrey K. Aronson, Dr Jane Collier, Professor Martin Dennis, Professor Michael Doherty, Professor B. Miles Fisher, Professor Brian M. Frier, Dr Ian Grant, Professor Christian J. Lueck, Dr Brian McClelland, Dr Kelvin Palmer, Professor Jonathan Rees, Dr Olivia Schofield, Mr Laurence H. Stewart, Dr George Webster and Dr Edmund Wilkins.

We are grateful to members of the International Advisory Board, all of whom provided detailed suggestions which have improved the book. Several members have stepped down and we would like to thank them for their support during the preparation of previous editions. They include Professor Jan D. Bos, Professor Y.C. Chee, Professor W.F. Mollentze, Professor Dato’ Tahir Azhar, Professor Alan Jardine, Dr Uwe Kornak, Dr Stuart McLellan, Dr Drak Nandwani, Dr David Patch, Professor Donald Salter, Professor John Simpson, Mr Grant Stewart and Professor Ian Weller.

As part of the publisher’s review, students and doctors from medical schools in the UK, Europe, Africa and Asia have provided valuable feedback on this textbook and their comments have helped shape this new edition. We hope we have listed all those who assisted, including Dr Sam Alfred, Dr Rustam Al-Shahi Salman, Professor Harry Campbell, Dr Richard Casasola, Dr Gavin Clunie, Professor Michael Eddie, Dr Catherine Elliot, Professor David Gawkrodger, Professor Jeremy Hall, Dr Amy Hughes, Professor Alan Jardine, Dr Uwe Kornak, Dr Stuart McLellan, Dr Drak Nandwani, Dr David Patch, Professor Donald Salter, Professor John Simpson, Mr Grant Stewart and Professor Ian Weller.

We are especially grateful to all those working for Churchill Livingstone, in particular Laurence Hunter, Wendy Lee and Robert Britton, for their endless support and expertise in the shaping, collation and illustration of this edition. We would also like to thank Louisa Talbott for her efficient project management, Susan Boobis for her labours in compiling the extensive index and Ruth Noble for expert proofreading.

BRW, NRC, SHR, IDP
Edinburgh
Figures reproduced with the publisher’s permission are listed at the end of each chapter. We are also grateful to the following individuals and organisations for the loan of illustrations:

Fig. 5.18A Institute of Ophthalmology, Moorfields Eye Hospital, London.

Fig. 6.9 Disc kindly supplied by Charlotte Symes.

Figs 10.4, 10.5AB Dr J. Xuereb.

Fig. 11.5 Dr J. Wilsdon, Freeman Hospital, Newcastle upon Tyne.

Page 294 insets (splinter haemorrhages) Dr Nick Beeching, Royal Liverpool University Hospital; (Roth’s spots) Prof. Ian Rennie, Royal Hallamshire Hospital, Sheffield. Page 295 (streptococcal toxic shock syndrome, meningococcal sepsis, shingles), Fig. 13.1 inset (cellulitis of the leg), Figs 13.6ABD, 13.20, 13.44B, 13.45B, 13.51 Dr Ravi Gowda, Royal Hallamshire Hospital, Sheffield. Fig. 13.1 insets (pulmonary tuberculosis) Dr Ann Chapman, Royal Hallamshire Hospital, Sheffield; (empyema, pyogenic liver abscess, diverticular abscess, tuberculous osteomyelitis) Dr Robert Peck, Royal Hallamshire Hospital, Sheffield. Fig. 13.3C Dr Julia Greig, Royal Hallamshire Hospital, Sheffield. Fig. 13.6C Dr Ratanaphone Phetsouvanh, Mahosot Hospital, Vientiane, PDR Laos. Fig. 13.14 Prof. Goura Kudesia, Northern General Hospital, Sheffield. Fig. 13.29 Institute of Ophthalmology, Moorfields Eye Hospital, London. Fig. 13.33 insets (malaria retinopathy) Dr Nicholas Beare, Royal Liverpool University Hospital; (blood films of P. vivax and P. falciparum) Dr Kamolrat Silamut, Mahidol Oxford Research Unit, Bangkok, Thailand. Fig. 13.41 Dr S. Sundar and Dr H.W. Murray. Fig. 13.42B Dr E.E. Zijlstra. Fig. 13.53 Dr Wendi Bailey, Liverpool School of Tropical Medicine. Fig. 13.57 insets (dimorphic fungi) Beatriz Gomez and Angela Restrepo, CIB, Medellin, Colombia.

Page 388 inset (oral hairy leucoplaikia) Audiovisual Dept, St Mary’s Hospital, London.

Fig. 15.4 Dr P. Hay, St George’s Hospital, London.

Page 463(4AB) Dr G.M. Iadorola and Dr F. Quarello, G. Bosco Hospital, Turin (from www.sin-italia.org/imago/sediment/sed.htm). Figs 17.1CE, 17.22ACDE Dr J.G. Simpson, Aberdeen Royal Infirmary. Figs 17.3AB, 17.4AB, 17.5, 17.25, 17.27, 17.32AB Dr A.P. Bayliss and Dr P. Thorpe, Aberdeen Royal Infirmary. Fig. 17.22F–H Dr R. Herriot. Fig. 17.23BC Dr J. Collar, St Mary’s Hospital, London. Fig. 17.29 Dr P. Robinson, St James’s University Hospital, Leeds.

Fig. 18.82E Dr T. Lawton. Fig. 18.83AB Dr B. Cullen.

Page 644 insets (idiopathic kyphoscoliosis) Dr I. Smith, Papworth Hospital, Cambridge; (serous, mucopurulent and purulent sputum) Dr J. Foweraker, Papworth Hospital, Cambridge. Fig. 18.83AB Dr B. Cullen.

Page 798 inset (exudative maculopathy), page 799 (background and proliferative retinopathy) Dr A.W. Patrick and Dr I.W. Campbell.

Fig. 21.4 insets (normal islet, beta-cell destruction) Dr A. Foulis, Dept of Pathology, University of Glasgow.

Fig. 22.14A Given Imaging.

Figs 23.10, 23.34, 23.37, 23.43 Dr D. Redhead, Royal Infirmary of Edinburgh.

Fig. 25.6 Dr I. Beggs; Fig. 25.7 Dr N. McKay.

Page 1138 insets (winging of scapula, 12th nerve palsy, wasting of thenar eminence) Dr R.E. Cull, Western General Hospital, Edinburgh. Figs 26.12A–C, 26.13A–C Dr D. Collie. Fig. 26.22C Dr B. Cullen. Fig. 26.27 Prof. D.A.S. Compston. Fig. 26.29 Dr J. Xuereb.

Figs 27.4AB, 27.9AB Dr A. Farrell and Prof. J. Wardlaw. Fig. 27.5A–D Dr D. Collie.
This page intentionally left blank
Medical practice 2
The doctor–patient relationship 2
Communication and other clinical skills 4
Using investigations 4
Estimating and communicating risk 7
Clinical decision-making 7
Practising medicine in low-resource settings 9
Medical ethics 9
Medical law 13

Personal and professional development 14

Complementary and alternative medicine 15
Since the time of Hippocrates, the role of the doctor has extended beyond the narrow remit of curing patients of their ailments. Good medical practice, or the art of medicine, hinges on recognising and respecting the breadth of physical, cultural, spiritual, experiential and psychosocial characteristics of each patient, and understanding their impact on the patient’s beliefs, attitudes and expectations. Doctors must deliver appropriate care which considers the technical complexities of modern treatment, and at the same time deals with the communication and interpersonal needs of the patient, at a time when he or she may feel most vulnerable. In addition to the diagnosis and treatment of illness, the scope of medicine has expanded to preventing disease through measures such as screening, vaccination and health promotion. Doctors are centrally involved in tackling lifestyle-related issues of the modern world, such as obesity, alcohol excess, cigarette smoking and sexual health.

Medical professionalism has been described in the UK by a Royal College of Physicians working party (2005) as ‘a set of values, behaviours and relationships that underpin the trust the public has in doctors’. They stated that doctors should be committed to integrity, compassion, altruism, continuous improvement, excellence, and working in partnership with members of the wider health-care team. They perceived that medical professionalism was relevant to leadership, education, career pathways, appraisal and research.

This chapter outlines how doctors must provide patients and their families with relevant but complex information, discuss management options, and reach appropriate clinical decisions, commensurate with the available resources. It also describes processes to develop, maintain and assure medical professionalism.

The contents of this book are not all based on indisputable contemporary evidence; many reflect wisdom and understanding distilled over hundreds of years and passed from generation to generation of doctors. This perceived wisdom lies at the heart of the way that doctors and patients interact; it demands respect, and if the doctor also displays compassion, sets the scene for the development of trust.

Due to the complexities of many chronic diseases and treatments, and the multifaceted impact of illness on a patient, there is an increasing role for health care to be delivered by a multidisciplinary team (Box 1.1). This model of care recognises the different skills of each allied health professional and focuses patient care beyond surgical procedures or pharmacological manipulation. The doctor usually takes the lead in determining the overall direction of care but must also:

- guide the patient through the unfamiliar landscape, language and customs of clinical care
- interpret, synthesise and convey complex information
- help patients and families to participate fully in the decision-making process.

In many clinical disciplines, doctors from several specialties form a multidisciplinary team in order to formulate a treatment plan. In oncology, for example, this ensures that various modalities of treatment (surgical, oncological and palliative) are considered.

The doctor-patient relationship is in itself therapeutic; a successful consultation with a trusted and respected practitioner will have beneficial effects irrespective of any other therapy given. The doctor-patient relationship is multilayered, dynamic and bilateral (Fig. 1.1).

Regulatory bodies, such as the UK General Medical Council, seek to define the medical side of the doctor-patient relationship in terms of the ‘Duties of a Doctor’ (Box 1.2). It is common for medical schools to require undergraduate students to sign an ethical code of conduct based on statements like this.

**Difficulties in the doctor–patient relationship**

Regardless of experience and skill, it is inevitable that, at some point in a doctor’s career, the doctor–patient relationship will break down. There can be many reasons for this; sometimes, these are beyond the control of the clinician, but often conflict arises when there is a genuine or perceived failure of the doctor to meet one or more of the duties outlined in Box 1.2. It is important to recognise a breakdown in the relationship quickly and, whenever possible, identify the reason. If patients are unhappy with an aspect of their care, they are entitled to a prompt, open, constructive and honest response that includes an explanation and, if appropriate, an apology. It is also important to reassure the patient that the issues raised will not adversely affect their future care.

Often, an acknowledgement that something is wrong and demonstration of a desire to put things right are sufficient to rectify any conflict. However, the longer
Fig. 1.1 Some aspects of the doctor–patient relationship.

1.2 The duties of a doctor registered with the UK General Medical Council

Patients must be able to trust doctors with their lives and health. To justify that trust you must show respect for human life and make sure your practice meets the standards expected of you in four domains.

Knowledge, skills and performance
- Make the care of your patient your first concern.
- Provide a good standard of practice and care.
  Keep your professional knowledge and skills up to date.
  Recognise and work within the limits of your competence.

Safety and quality
- Take prompt action if you think that patient safety, dignity or comfort is being compromised.
- Protect and promote the health of patients and the public.

Communication, partnership and teamwork
- Treat patients as individuals and respect their dignity.
  Treat patients politely and considerately.
  Respect patients’ right to confidentiality.

- Work in partnership with patients.
  Listen to, and respond to, their concerns and preferences.
  Give patients the information they want or need in a way they can understand.
  Respect patients’ right to reach decisions with you about their treatment and care.
  Support patients in caring for themselves to improve and maintain their health.
- Work with colleagues in the ways that best serve patients’ interests.

Maintaining trust
- Be honest and open and act with integrity.
- Never discriminate unfairly against patients or colleagues.
- Never abuse your patients’ trust in you or the public’s trust in the profession.

You are personally accountable for your professional practice and must always be prepared to justify your decisions and actions.
one takes to address a problem, the more difficult it becomes to resolve. The patient may continue to be dissatisfied with the doctor and it may be most appropriate for another colleague to take over their care. It is important to reflect on such incidents, to identify whether one would approach a similar challenge differently next time.

**Communication and other clinical skills**

Communication lies at the heart of good medical practice. The most technically capable clinician will fail in the duty of care if he or she is unable to communicate effectively with patients or relatives, since this is essential for accurate history-taking, information-giving and decision-making. Likewise, the delivery of holistic care requires effective communication with other doctors and members of the multidisciplinary team. Clear and appropriately detailed clinical note-keeping is essential, as are timely and accurate written communications between professionals.

Failures in communication may lead to poor health outcomes, strained working relations, dissatisfaction among patients, their families and health professionals, anger and litigation. The majority of complaints received by health-care professionals could have been avoided by effective communication. Box 1.3 lists some common barriers to good communication.

Developing communication skills to facilitate accurate history-taking and information-giving takes many years and requires frequent personal reflection on previous consultations. A detailed account of history-taking, clinical examination and communication skills is beyond the scope of this chapter but is provided in Davidson’s sister book, Macleod’s Clinical Examination. However, some communication principles are discussed below and these can be applied to most consultations.

The main aim of a medical interview is to establish a factual account of the patient’s illness. The clinician must allow the patient to describe the problems without overbearing interrogation, but should try to facilitate the process with appropriate questions (Box 1.4). Techniques such as an unhurried approach, checking prior understanding, making it clear that the interviewer is listening, the use of silence when appropriate, recappping on what has been said, and reflection of key points back to the patient are all important. A major requirement is to express complex information and concepts in language with which the patient can readily engage. Nonverbal communication is equally important. The patient’s facial expressions and body language may betray hidden fears. The clinician can help the patient to talk more freely by smiling or nodding appropriately.

Beyond the factual account of symptoms, the clinician should also explore patients’ feelings, determine how they interpret their symptoms, unearth their concerns and fears, and explore their expectations before suggesting and agreeing a plan of management. Clinicians should demonstrate understanding, sensitivity and empathy (i.e. imagine themselves in the patient’s position). Most patients have more than one concern and will be reluctant to discuss potentially important issues if they feel that the clinician is not interested or is likely to dismiss their complaints as irrational or trivial.

Specific communication scenarios, such as breaking bad news or dealing with aggression, require additional targeted strategies (see Macleod’s Clinical Examination).

While many common clinical conditions can be identified on the basis of the history from the patient, the process of physical examination remains important in most clinical scenarios. Physical examination is an important characteristic of the doctor–patient relationship, at best benefiting from and reinforcing trust, but at worst a focus of complaint when the doctor–patient relationship has not been established or has broken down. Key findings on physical examination pointing to disease in specific body systems are described in the relevant chapters of this book.

**Using investigations**

Modern medical practice has become dominated by sophisticated and often expensive investigations. It is easy to forget that the judicious use of these tools, and the interpretation of the data that they provide, are crucially dependent on good basic clinical skills. Indeed, a test should only be ordered if it is clear that the result will influence the patient’s management and the perceived value of the resulting information exceeds the
anticipated discomfort, risk and cost of the procedure. Clinicians should therefore analyse their patient’s condition carefully and draw up a provisional management plan before requesting any investigations.

The ‘normal’ (or reference) range

Although some tests provide qualitative results (present or absent, e.g. faecal occult blood testing, p. 857), most provide quantitative results (i.e. a value on a continuous numeric scale). In order to classify quantitative results as normal or abnormal, it is necessary to define a ‘normal range’. Many quantitative measurements in populations exhibit a bell-shaped, or Gaussian, frequency distribution (Fig. 1.2); this is called a ‘normal distribution’ and is characteristic of biological variables determined by a complex mixture of genetic and environmental factors (e.g. height) and of test results (e.g. plasma sodium concentration). A normal distribution can be described by the mean value (which places the centre of the bell-shaped curve on the x axis) and the standard deviation (SD, which describes the width of the bell-shaped curve). Within each SD away from the mean, there is a fixed percentage of the population. By convention, the ‘normal range’ is defined as those values which encompass 95% of the population, i.e. the values within 2.SDs above and below the mean. If this convention is used, however, 2.5% of the normal population will have values above, and 2.5% will have values below, the normal range; for this reason, it is more precise to describe ‘reference’ rather than ‘normal’ ranges.

‘Abnormal’ results, i.e. those lying beyond 2.SDs from the mean, may occur either because the person is one of the 2.5% of the normal population whose test result is outside the reference range, or because he or she has a disease characterised by a different result from the test. Test results in ‘abnormal’ populations also have a bell-shaped distribution with a different mean and SD (see Fig. 1.2). In some diseases, there is typically no overlap between results from the normal and abnormal population (e.g. elevated serum creatinine in renal failure, p. 467). In many diseases, however, there is overlap, sometimes extending into the reference range (e.g. elevated serum thyroxine in toxic multinodular goitre, p. 753). In these circumstances, the greater the difference between the test result and the limits of the reference range, the higher the chance that the person has a disease, but there is a risk that results within the reference range may be ‘false negatives’ and results outside the reference range may be ‘false positives’.

Each time a test is performed in a member of the normal population there is a 5% (1 in 20) chance that the result will be outside the reference range. If two tests are performed, the chance that one of them will be ‘abnormal’ is 10% (2 in 20), and so on; the chance of an ‘abnormal’ result increases as more tests are performed, so multiple indiscriminate testing should be avoided.

In practice, reference ranges are usually established by performing the test in a number of healthy volunteers who are assumed to be a random sample of the normal population. Not all populations are the same, however, and while it is common to have different reference ranges for men and women or children and adults, clinicians need to be aware that reference ranges defined either by test manufacturers or even within the local laboratory may have been established in small numbers of healthy young people who are not necessarily representative of their patient population.

For some tests, the clinical decision does not depend on whether or not the patient is a member of the normal population. This commonly applies to quantitative risk factors for future disease. For example, higher plasma total cholesterol levels are associated with a higher risk of future myocardial infarction (p. 583) within the normal population. Although a reference range for cholesterol can be calculated, cholesterol-lowering therapy is commonly recommended for people with values within the reference range; the ‘cutoff’ value at which therapy is recommended depends upon the presence of other risk factors for cardiovascular disease. The reference range for plasma cholesterol is therefore redundant and the phrase ‘normal plasma cholesterol level’ is unhelpful. Similar arguments apply for interpretation of values of blood pressure (p. 583), bone mineral density (p. 1065) and so on.

Some quantitative test results are not normally distributed, usually because a substantial proportion of the normal population will have an unrecordably low result (e.g. serum prostate-specific antigen, p. 518), and the distribution cannot be described by mean and SDs. Alternative statistical procedures can be used to calculate 95th centiles, but it is common in these circumstances to use information from normal and abnormal people to identify ‘cutoff’ values which are associated with a certain risk of disease, as described below.

Sensitivity and specificity

No test is completely reliable. All diagnostic tests can produce false positives (an abnormal result in the absence of disease) and false negatives (a normal result in a patient with disease). The diagnostic accuracy of a test can be expressed in terms of its sensitivity and its specificity (Box 1.5).
Sensitivity is defined as the percentage of the test population who are affected by the index condition and test positive for it. In contrast, specificity is defined as the percentage of the test population who are healthy and test negative. A very sensitive test will detect most disease but may generate abnormal findings in healthy people; a negative result will therefore reliably exclude disease but a positive test is likely to require further evaluation. On the other hand, a very specific test may miss significant pathology but is likely to establish the diagnosis, beyond doubt, when the result is positive.

In choosing how a test is used to guide decision-making, there is an inevitable trade-off between emphasising sensitivity versus specificity. For example, defining an exercise electrocardiogram (p. 534) as abnormal if there is at least 0.5 mm ST depression will ensure that very few cases of coronary artery disease are missed but will generate many false-positive tests (high sensitivity, low specificity). On the other hand, a cutoff point of at least 2.0 mm ST depression will detect most cases of important coronary disease with far fewer false-positives. This trade-off can be illustrated by the receiver operating characteristic curve of the test (Fig. 1.3).

**Predictive value**

The predictive value of a test is determined by its sensitivity and specificity, and can be expressed in several ways. The positive predictive value is the probability that a patient with a positive test has the index condition, while the negative predictive value is the probability that a patient with a negative test does not have the condition (see Box 1.5). The likelihood ratio expresses the odds that a given finding would occur in a patient with, as opposed to a patient without, the index condition (see Box 1.5); as the odds rise above 1, the probability that disease is present rises.

The interpretation and the utility of a test are critically dependent on the circumstances in which it is used. Bayes’ theorem dictates that the value of a diagnostic test is determined by the prevalence of the condition in the test population. The probability that a subject has a particular condition (the pre-test probability) can be calculated if the pre-test probability and the sensitivity and specificity of the test are known (Box 1.6). A test is most valuable when there is an intermediate pre-test probability of disease. Clinicians seldom have access to such precise information but must appreciate the importance of integrating clinical and laboratory data.

**Screening**

Many health-care systems run screening programmes to detect important (and treatable) disease in apparently healthy but at-risk individuals. These initiatives may be directed towards a single pathology (e.g. mammography for breast cancer, p. 280) or may comprise a battery
of tests for a wide range of conditions. Screening inevitably generates a number of false-positive results that require further, potentially expensive and sometimes risky, investigation. This may engender a good deal of anxiety for the patient and create dilemmas for the clinician; for example, it may be difficult to determine how to evaluate minor abnormalities of the liver function tests in an otherwise healthy person (p. 935).

Some of the criteria that must be considered before deciding if the wider costs of a screening programme can be justified are listed in Box 1.7.

### Estimating and communicating risk

Medical management decisions are usually made by weighing up the anticipated benefits of a particular procedure or treatment against the potential risks. To allow patients to contribute to the decision-making process, health professionals must be able to explain risk in an accurate and understandable way.

Providing the relevant biomedical facts is seldom sufficient to guide decision-making because a patient’s perception of risk is often coloured by emotional, and sometimes irrational, factors. Most patients will have access to information from a wide variety of sometimes conflicting sources, including the Internet, books, magazines, self-help groups, other health-care professionals, friends and family. The clinician must be aware of and sensitive to the way in which these resources influence the individual, while building trust with the patient, clarifying the problem and conveying the key facts.

Research evidence provides statistics but these can be confusing (Box 1.8). Relative risk describes the proportional increase in risk; it is a useful measure of the size of an effect. In contrast, absolute risk describes the actual chance of an event and is what matters to most patients. Terms such as ‘common’, ‘rare’, ‘probable’ and ‘unlikely’ are elastic. Whenever possible, clinicians should quote numerical information using consistent denominators (e.g. ‘90 of 100 patients who have this operation feel much better, 1 will die during the operation and 2 will suffer a stroke’). Positive framing (‘There is a 99% chance of survival’) and negative framing (‘There is a 1% chance of death’) may both be appropriate. A variety of visual aids can be used to present complex statistical information (Fig. 1.4).

Finally, it is essential to allow the patient to place his or her own weighting on the potential benefits and adverse effects of each course of action. Thus, some patients may choose to sacrifice a good chance of pain relief because they are not prepared to run even a small risk of paralysis, whilst others may opt to proceed with very high-risk spinal surgery because they find their current circumstances intolerable.

### 1.7 Factors that influence the cost-effectiveness of screening for a disease

- The prevalence of the disease in the target population
- The cost of the screening test
- The sensitivity and specificity of the screening test
- The availability and effectiveness of treatment
- The cost of not detecting and treating the disease

### 1.8 Explaining the risks and benefits of therapy

Would you take a drug once a day for a year to prevent stroke if:

- it reduced your risk of having a stroke by 47%?
- it reduced your chance of suffering a stroke from 0.26% to 0.14%?
- there was one chance in 850 that it would prevent you having a stroke?
- 849 out of 850 patients derived no benefit from the treatment?
- there was a 99.7% chance that you would not have a stroke anyway?

All these statements are derived from the same data and describe an equivalent effect.*


### Clinical decision-making

Assimilating symptoms, signs and results of investigations into a diagnosis and then planning treatment are highly complex tasks that require not only factual knowledge but also a highly developed set of skills in decision-making. Diagnostic decision-making is guided by Ockham’s razor, originally expressed by the 14th-century Englishman William of Ockham as ‘plurality should not be posited without necessity.’ In short, all things being equal, the simplest explanation is the best. In practice, clinicians formulate hypotheses about the underlying diagnosis (or shortlist of diagnoses, the ‘differential’ diagnosis) during the consultation with the patient and refine this hypothesis both by collecting selected additional information and by choosing to ignore other information which they regard as irrelevant, in order to reach the most parsimonious diagnosis.

Decision-making in health care often operates under conditions of uncertainty, where it is uncertain what is wrong with the patient or which treatment is most appropriate. This can lead to variations in how clinicians...
make decisions, and subsequently variations in the care that patients receive. Clinicians often employ a process of ‘ad hoc’ decision-making, where they use some form of global judgement about what might be the best course of action for an individual patient. These ad hoc decisions may be based on a number of factors, including what a clinician has been taught, his or her clinical experience of other patients with that particular disease, or what is common practice within a particular institution. However, such decisions may be governed by heuristics or bias, which may lead to errors. Heuristics are cognitive processes or ‘rules of thumb’ used unconsciously when making decisions (Box 1.9). Such processes may lead to mistakes, most commonly when there is a lack of evidence to inform practice. Whenever possible, clinical decision-making should be guided by evidence-based medicine.

**Evidence-based medicine**

Patient treatment should be based on the integration of best research evidence alongside clinical expertise and patient values. The discipline of evidence-based medicine (EBM) came into being in order to introduce a more systematic approach to the use of evidence in making clinical decisions. This was made possible by:

- the development of statistical methods to analyse data systematically
- recognition of the importance of analysing all data, both published and unpublished
- the development of databases of relevant information and systems by which to access such information.

The principles of EBM are based on the tenet that well-formulated questions about medical management can be answered by:

- conducting high-quality randomised controlled trials
- tracing all the available evidence
- critically appraising the evidence
- applying the evidence to the management of the individual patient.

EBM categorises different types of clinical evidence and ranks them according to their freedom from the various biases that beset medical research. It therefore places greater emphasis on evidence from a meta-analysis of randomised controlled trials than on a series of case reports or expert opinion (Box 1.10).

**Guidelines and protocols**

The terms ‘clinical guidelines’ and ‘protocols’ are often used together, yet they are inherently different.

**Guidelines**

Clinical guidelines aim to guide clinicians on how to manage specific clinical scenarios using the best available evidence. They have been in existence throughout the history of medicine, although many were based on tradition or authority. A large number of local, national and international bodies have produced guidelines, following a range of different methodologies (see [www.evidence.nhs.uk](http://www.evidence.nhs.uk)). Some are based on systematic reviews of the medical literature and others on consensus of expert opinion. When considering guidelines, it is important for clinicians to be aware of the strength of the evidence on which the recommendations are based (see Box 1.10).

---

**Box 1.9 Heuristics in clinical decision-making**

<table>
<thead>
<tr>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The probability of an event is estimated based on how easily an individual can recall a similar event, e.g. a doctor judges that a patient has a particular disease because the case reminds him or her of a similar case seen recently</td>
</tr>
<tr>
<td>• This can lead to errors, as individuals often recall recent or vivid events more easily, rather than considering the likelihood of an event in the wider population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Representativeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The probability of an event is estimated based on how similar (or representative) it is of a wider category of events, e.g. a doctor judges that a patient has a particular disease because the patient’s signs and symptoms are ‘representative’ of that disease</td>
</tr>
<tr>
<td>• This can lead to errors, such as neglecting to take into account the prevalence of the disease in a specific patient population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anchoring and adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The probability of an event is estimated by taking an initial reference point (anchor) and then adjusting this to reach a final judgement about likelihood, e.g. a doctor judges that the likelihood of a patient having a particular disease is 60%. The doctor collects information (perhaps from diagnostic tests) and re-assesses his or her estimation on the basis of these results to reach a final diagnosis</td>
</tr>
<tr>
<td>• This can lead to errors, as final estimations of likelihood are linked to the original anchor, so if this is incorrect, the final judgement is also likely to be inaccurate</td>
</tr>
</tbody>
</table>

**Box 1.10 Categories in evidence-based medicine (EBM)**

<table>
<thead>
<tr>
<th>Levels of evidence (in descending order of strength)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia Evidence obtained from meta-analysis of randomised clinical trials</td>
</tr>
<tr>
<td>Ib Evidence obtained from at least one randomised controlled trial</td>
</tr>
<tr>
<td>Ila Evidence obtained from at least one well-designed controlled study without randomisation</td>
</tr>
<tr>
<td>IIb Evidence obtained from at least one other type of well-designed quasi-experimental study</td>
</tr>
<tr>
<td>III Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case studies</td>
</tr>
<tr>
<td>IV Evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities</td>
</tr>
</tbody>
</table>

**Grades of recommendation**

- **A** Directly based on level I studies
- **B** Directly based on level II studies or extrapolations from level I studies
- **C** Directly based on level III studies or extrapolations from level I or II studies
- **D** Directly based on level IV studies or extrapolations from level I, II or III studies

*From the Scottish Intercollegiate Guidelines Network (SIGN; see www.sign.ac.uk). This scheme is widely used, although other modified schemes exist.*
Properly developed guidelines recognise that medicine is an art as well as a science and that the evidence on which the guidelines are based is, strictly speaking, only applicable to the study population in the trial(s). Clinicians must therefore use their judgement to ascertain whether the recommendations are applicable to the patient in front of them.

Some guidelines are formulated not only from evidence-based best practice but also from cost-effectiveness (see below). An example in the UK is guidance produced by the government-commissioned body, the National Institute for Health and Clinical Excellence (NICE; see www.nice.org.uk). These guidelines recognise that health services have limited resources, and that treatments should be prioritised which offer the greatest improvement in health for the largest number of people per unit of resource.

**Protocols**

Whilst guidelines recognise the individuality of the patient and help clinicians decide on which action is best, protocols are far more directive and are written to be followed exactly. Protocols usually apply in situations where the clinical decision has already been made and an intervention is then being instigated. Protocols aim to ensure that treatment will be identical, irrespective of where and by whom it is given. For example, a guideline may help a multidisciplinary team decide which modality of treatment is best for someone with lung cancer by evaluating the best evidence alongside the individual psychosocial needs of the patient. However, once a decision has been made in favour of a certain treatment, e.g. chemotherapy, the clinician will be expected to follow a strict protocol outlining dosages, routes of administration and monitoring.

**Cost-effectiveness**

The best available health care can be expensive. No country can now afford to provide unlimited state-of-the-art medicine for all its citizens. Health-care systems must therefore take account of the cost-effectiveness of the treatments they provide. This can create difficult dilemmas for clinicians, who may be asked to withhold expensive but effective therapies (e.g. implantable defibrillators) from individual patients on the basis that the money will do more good for more patients if it is spent elsewhere (e.g. offering angioplasty to all acute myocardial infarction patients). Assessing the cost-effectiveness of interventions and allocating resources accordingly follows ethical principles such as justice, which are covered in greater detail below.

**Quality-adjusted life years**

Outcomes from health care can be measured in terms of changes in the quality and quantity of life. Life expectancy is easily defined but quality of life is difficult to measure. Nevertheless, it is possible to construct a continuum between perfect health (score 1), survival with no quality of life (score 0), and states that are perceived to be worse than death (minus score). Quality and quantity of life can then be combined in a measure known as the quality-adjusted life year (QALY). For example, an intervention that results in a patient living an additional 4 years with an average quality of life rated as 0.6 on the continuum would yield 2.4 QALYs (4 × 0.6). Thus a cost per QALY can be calculated and compared with other interventions (p. 32). This approach is not perfect but offers a means of comparing the cost-effectiveness of a wide range of treatments.

Another useful measure is the disability-adjusted life year (DALY), which is used by the World Health Organization to quantify the overall burden of disease in populations; it cumulatively estimates the number of years lost due to ill health, disability and death.

### Practising medicine in low-resource settings

The challenges associated with medical care in low-resource areas cluster in four domains:

- **Prevention versus cure.** Prevention is easier, cheaper and more effective than cure for many diseases. On the other hand, curative medicine is immediate, highly visible and glamorous. This tension is most evident when a disease is common and the benefits of prevention have yet to be realised. The allocation of adequate resources for long-term prevention needs both political will and social acceptance.
- **Acute versus chronic care.** Treating chronic illness can be time-consuming and less immediately gratifying than acute emergency medicine. Facilities for chronic care are therefore accorded a low priority in many health-care systems. Unfortunately, this often results in patients who require long-term care being denied treatment altogether or being managed inappropriately in the acute sector.
- **The ideal versus the possible.** Most medical management guidelines are derived from studies that were conducted in well-resourced health-care systems. In trying to apply this knowledge to the developing world, there are tensions between best practice and what is possible. For example, anticoagulant therapy (p. 1018) may pose risks that were not evident in the studies that underpin guidelines if it is prescribed in areas where reliable laboratories are not available and medications that interact with warfarin are commonly purchased ‘over the counter’.
- **Channels of health-care provision.** In developing countries, health care may be delivered through government-run public clinics (usually free or subsidised) or non-governmental organisations (sometimes subsidised but usually privately funded). Many of the available services are too costly for the average patient. There is a need for constructive cooperation between all of the health-care sectors.

The best possible practice is that which can be delivered within the available resources in a specific setting. Compassionate care given with empathy, understanding and good communication is always within the physician’s reach, even when resources are inadequate.

### Medical ethics

Ethics has been described as the ‘science’ of morality, and defines systems of moral values. Medical ethics is
concerned both with the standards of conduct and competence expected of medical professionals, some of which are captured in legislation, and with the study of moral problems raised by the practice of medicine. Recent advances in biomedical science and their application to clinical care have thrown up many difficult ethical problems. These include human cloning, predictive genetic testing, eugenics, new reproductive technologies, antenatal screening, abortion, priority-setting, under-served populations, brain death, organ transplantation, end-of-life issues, and assisted suicide. Detailed discussion of these is beyond the scope of this chapter but a framework for the application of ethics to medical practice is described.

In general, ethical problems relate to the intentions or motives of those involved, their actions, the consequences of their actions, and the context in which their actions take place. Ethical problems can be analysed in a variety of ways, sometimes leading to different conclusions. To find the best solution, it may be necessary to apply several analytical approaches and attempt to reconcile the conclusions. In modern medical practice, there is not always time to do this systematically. However, the process of applying an ethical framework to a given situation is a key element in clinical decision-making and helps to ensure that a decision is both morally acceptable and legally defensible.

• **Virtue ethics** is concerned with the character of the persons involved and with their actions. Are my intentions (what my actions aim at) and my motives (what moves me to act) good or bad, wise or unwise, sensible or unrealistic, patient-centred or self-centred, and so on? Is the action I propose to take one which would be considered appropriate by a prudent doctor – or by a prudent patient? The focus here is on the characteristics of a virtuous person and the action they would take.

• **Deontological ethics** is concerned with whether a proposed action or course of action, in itself and regardless of its consequences, is right or wrong. Is it ever right or always wrong to kill, to tell a lie, to break a promise? Deontological (from the Greek for ‘duty’) considerations include rights as well as duties, and omissions as well as acts. An action is right if it is in accordance with an established moral rule or principle.

• **Teleological ethics** (or consequentialism) is concerned with the consequences of a proposed action. Are they likely to be good or bad, in the short term and long term, for the patient, doctor, family and society? What will promote a net balance of good over harm for the individual, as well as ‘the greatest good for the greatest number’?

An ethical problem can therefore be addressed by trying to decide what a virtuous person would do, whether an action or course of action is right or wrong in itself, or what its consequences might be. Yet the circumstances in which any decision is made will vary, and what may be right in one context may be wrong in another. **Situation ethics** recognises this, emphasising the need to consider carefully the context (or situation) in which a course of action is chosen.

Ethics is applied to the practice of medicine in three broad areas:

- **Clinical ethics** deals with the relationship between clinicians and patients, as described below.
- **Public health ethics** deals with the health issues of groups of people – the community. Examples include the banning of smoking in public places, where the autonomy of the individual may be coerced for the greater good of the community.
- **Research ethics** deals with issues related to clinical research. This is to ensure not only that research is conducted safely but also that the rights of the participants are paramount. No research can be undertaken unless it has undergone ethical scrutiny.

**Principles of clinical ethics**

In clinical ethics, four key principles are frequently used to analyse a problem, and often abbreviated to ‘autonomy, beneficence, non-maleficence and justice’.

**Respect for persons and their autonomy**

This respect is a significant aspect of the relationship between patient and doctor. The patient seeks out a doctor based on a desire to attain freedom from a disability or disease which limits his or her ability to exercise autonomy (the power or right of self-determination). Unless the patient is a child, is unconscious or is mentally incapacitated, it is the patient’s choice to seek advice. The physician must therefore respect the patient’s autonomy. This includes the patient’s right to refuse therapy. The doctor must also actively seek to empower the patient with adequate information.

**Truth-telling**

Telling the truth is essential to generating and maintaining trust between the doctor and the patient. This includes providing information about the nature of the illness, expected outcome and therapeutic alternatives, and answering questions honestly. The facts should not be given ‘brutally’ but with due sensitivity to appropriate timing and to the patient’s capacity to cope with bad news. However, the clinical uncertainties described earlier in the chapter must also be acknowledged. There are two rare situations where the truth may, at least for a time, be withheld:

- If it will cause real harm to the patient (e.g. a depressed patient likely to commit suicide who has to be told that he or she has cancer). This is sometimes called ‘therapeutic privilege’, since it should be exercised only in the patient’s interests, for serious clinical reasons.
- If the patient makes it clear that he or she does not want to hear the bad news (but always bearing in mind that this may be a stage in the patient’s adjustment to the condition).

In no case should false information be given, and the physician should always be prepared to justify any decision to withhold relevant information.

**Informed consent**

This term describes the participation of patients in decisions about their health care. In order to facilitate this, the clinician must provide the patient with an adequate explanation and details of the relevant risks, benefits and uncertainties of each possible course of action. The amount of information to provide will
vary, depending on the patient’s condition and the complexity of the treatment, and on the physician’s assessment of the patient’s understanding of the situation. Not all options need be explained, but those that a ‘prudent patient’ would consider significant should be explored – for example, by open questioning (see Box 1.4, p. 4).

From both a legal and an ethical perspective, the patient retains the right to decide what is in his or her best interests. All adults have decision-making capacity if they can understand the relevant information (which may have to be explained in simple terms), consider the implications of the relevant options, and make a communicable decision. If a patient makes choices that seem irrational or are at variance with professional advice, it does not mean that they lack capacity.

When the patient does lack decision-making capacity, the clinician should always act in the best interests of the patient. In an emergency, consent may be presumed, but only for treatment immediately necessary to preserve the patient’s life and health, and if there is no clear evidence that this would be against the previous settled wishes of the patient when competent (for example, blood transfusion in the case of an adult Jehovah’s Witness). If the patient has a legally entitled surrogate decision-maker, their consent should be sought if possible. It is also good practice to involve close relatives in decision-making but the hierarchy of surrogate decision-makers will depend on local laws and culture.

Confidentiality

Confidentiality in relation to the management of patient-specific information is important in generating and maintaining trust in the doctor–patient relationship. Health-care teams must take precautions to prevent unauthorised access to patient records, and may disclose patient-identifying information only when the patient has given consent or when required by law. When such information is shared with other health-care professionals in order to optimise patient care, this should be done on a strictly ‘need-to-know’ basis.

Beneficence

This is the principle of doing good, or acting in another person’s best interests. In clinical ethics, the term refers to the good of the individual patient. It means considering the patient’s view, as well as the medical view, of his or her own best interests. Situations may arise when there is a conflict between what is good for the individual and what is best for society, but the traditional medical approach is that stated in the Declaration of Geneva (World Medical Association): ‘The health of my patient will be my first consideration.’

Non-maleficence

This is the principle of doing no harm: in medicine, the traditional ‘primum non nocere’. In balancing beneficence and non-maleficence (benefit versus risk), the clinician must share information with the patient, who can then be helped to make an informed decision.

Justice

In the context of clinical ethics, justice relates primarily to the distribution of medical care and the allocation of resources. In order to distribute health resources justly, the concept of utility – ‘greatest good for the greatest number’ – must be considered. In the case of individual patients, however, justice is also equated with being ‘fair’ and ‘even-handed’. The concept of fair delivery of health care can be viewed from three perspectives:

- Respect for the needs of the individual. Health care is delivered first to those who need it most. This perspective is particularly relevant when need must be assessed by some kind of triage.
- Respect for the rights of a person. Everyone who needs health care is entitled to a fair share of the resources available. This perspective is particularly relevant when local or global economic, social, educational or other inequalities prevent or reduce equitable access to health care.
- Respect for merit. Health care is delivered on the basis of value judgements, according to financial, political, social or other factors relating to the value of the individual to society. For example, many national leaders have their own personal physician and medical teams. The relevance of this perspective to health care is widely disputed, not least because such value judgements are difficult to make in practice and to defend ethically.

Types of ethical problem

When faced with an ethical problem, it is often helpful to characterise it in terms of certain patterns (Fig. 1.5).

A gap or block

The ideal goal is clearly seen but there are major obstacles to achieving it. The obstacles may be economic or social, or in the belief system of the patient. The obvious answer – to bridge the gap or remove the block – may not be possible within the available time frame and resources. A young boy from a poor family in a developing country, who has Wilson’s disease and needs a liver transplant, is an example of an economic block. Some problems of this kind cannot be resolved satisfactorily in the clinical context until or unless they are resolved in the economic or political context.

Priority-setting

The right course of action is clear but prioritisation is necessary and the principles to guide that process have to be defined. A decision to allocate the last bed in intensive care to either an 80-year-old with pneumonia or a 20-year-old with advanced lymphoma is an example. While it is not possible to cover all eventualities, guidelines agreed in advance with stakeholders are helpful.

A moral dilemma

Acting in accordance with one ethical principle may conflict with another ethical principle. This can create a moral dilemma – a choice between two alternatives, neither of which is ethically satisfactory. For example, a physician may decide that a particular mode of therapy is best (principle of beneficence), while the patient makes a different choice (principle of respect for autonomy). Consider artificial feeding by a percutaneous endoscopic gastrostomy (PEG; p. 123). The doctor may be reluctant to see the patient die for lack of nutrition and believe that this is the best route for feeding. The patient may, however, refuse the procedure, based on an informed
conclusion. It helps the decision-maker to grow personally and professionally, allows communication of the process by which a decision is made, and permits the process to be constructively criticised. When, in everyday practice, time for reflection is limited, knowledge of methods of moral reasoning provides a useful background and aid for decision-making, and is often employed in ways analogous to those of ‘the novice–expert shift’ (see Box 1.14, p. 14). Some approaches that can be applied are as follows:

- **A principles approach.** This involves analysing an ethical problem in terms of the principles of respect for autonomy, beneficence, non-maleficence and justice. If all of these principles support a particular course of action, then that course of action is probably correct and there may, in fact, no longer be an ethical problem. If, however, different principles suggest different courses of action, this approach has no intrinsic mechanism for deciding which principle has priority. On the other hand, analysing the problem in terms of these principles can help to clarify the nature of the ethical problem and the issues which need to be addressed if the problem is to be resolved.

- **A casuistry (cases) approach.** This uses precedent as a guide to what to do. A case is recalled or imagined which is similar to that under discussion but where the right choice of action/behaviour was obvious. Then the features which make the present case different, if any, are analysed and considered to see if and why they lead to a different conclusion. A variation on this approach, related to virtue ethics, is to imagine what a physician who was particularly skilled or experienced in this type of situation would do, or how a previous patient might have viewed the problem.

- **A perspectives (or narrative) approach.** A perspectives approach involves considering the views of all the stakeholders: the patient, the family or carers, the healthcare team, the health service and society. The greater the degree of concordance of these views on a particular outcome, the more likely it is that the decision leading to that outcome is right. A narrative approach is similar but involves listening attentively to the different ‘stories’ told by the stakeholders about the problem and how they perceive it. Where these stories differ can provide clues to a more nuanced understanding of the problem and how it might be resolved.

- **A counter-argument approach.** A particular course of action is chosen and the best ethical arguments against it are then marshalled and evaluated. This may or may not cause the decision to be reconsidered.

- **Application of rules.** In certain common and clearly defined situations, externally imposed rules (including the law) may require, or guide towards, a specific course of action. This does not obviate the need for ethical analysis. Moreover, any such rules must be reviewed regularly.

While all of these approaches may be useful, it is important to remember that none of them removes the need, on the one hand, for the exercise of judgement and, on the other, for good communication and consensus decision-making. No less important is the assessment of their own quality of life and prospects of recovery. In theory, the dilemma can be resolved only if one of the ethical principles is given priority; ethical analysis (see below) can help to achieve resolution. True moral dilemmas are less common in practice than in theory; apparent dilemmas can often be resolved by good doctor–patient communication.

**Resolving conflict**

A conflict of opinion may arise between members of the team responsible for care of the patient. For example, doctors in a renal medicine unit providing dialysis therapy (p. 489) may have divergent views on whether this treatment is appropriate for a patient who is elderly with significant comorbidity. Differing views should normally be resolved through discussion; in this example, conflict would usually be resolved in a multidisciplinary team meeting at which discussion of patients approaching end-stage renal failure is routine. However, if this does not work, referral to a decision-making authority allocated in advance (e.g. the clinical director of the service) may be necessary. The challenge then is to ensure consistent and accurate implementation of the decision.

**Ethical analysis**

Ethical analysis (or moral reasoning) is the process of thinking through ethical problems and reaching a conclusion. It helps the decision-maker to grow personally and professionally, allows communication of the process by which a decision is made, and permits the process to be constructively criticised. When, in everyday practice, time for reflection is limited, knowledge of methods of moral reasoning provides a useful background and aid for decision-making, and is often employed in ways analogous to those of ‘the novice–expert shift’ (see Box 1.14, p. 14). Some approaches that can be applied are as follows:

- **A principles approach.** This involves analysing an ethical problem in terms of the principles of respect for autonomy, beneficence, non-maleficence and justice. If all of these principles support a particular course of action, then that course of action is probably correct and there may, in fact, no longer be an ethical problem. If, however, different principles suggest different courses of action, this approach has no intrinsic mechanism for deciding which principle has priority. On the other hand, analysing the problem in terms of these principles can help to clarify the nature of the ethical problem and the issues which need to be addressed if the problem is to be resolved.

- **A casuistry (cases) approach.** This uses precedent as a guide to what to do. A case is recalled or imagined which is similar to that under discussion but where the right choice of action/behaviour was obvious. Then the features which make the present case different, if any, are analysed and considered to see if and why they lead to a different conclusion. A variation on this approach, related to virtue ethics, is to imagine what a physician who was particularly skilled or experienced in this type of situation would do, or how a previous patient might have viewed the problem.

- **A perspectives (or narrative) approach.** A perspectives approach involves considering the views of all the stakeholders: the patient, the family or carers, the healthcare team, the health service and society. The greater the degree of concordance of these views on a particular outcome, the more likely it is that the decision leading to that outcome is right. A narrative approach is similar but involves listening attentively to the different ‘stories’ told by the stakeholders about the problem and how they perceive it. Where these stories differ can provide clues to a more nuanced understanding of the problem and how it might be resolved.

- **A counter-argument approach.** A particular course of action is chosen and the best ethical arguments against it are then marshalled and evaluated. This may or may not cause the decision to be reconsidered.

- **Application of rules.** In certain common and clearly defined situations, externally imposed rules (including the law) may require, or guide towards, a specific course of action. This does not obviate the need for ethical analysis. Moreover, any such rules must be reviewed regularly.

While all of these approaches may be useful, it is important to remember that none of them removes the need, on the one hand, for the exercise of judgement and, on the other, for good communication and consensus decision-making. No less important is the assessment of their own quality of life and prospects of recovery. In theory, the dilemma can be resolved only if one of the ethical principles is given priority; ethical analysis (see below) can help to achieve resolution. True moral dilemmas are less common in practice than in theory; apparent dilemmas can often be resolved by good doctor–patient communication.
There is no formal written ‘advance directive’ on file. The patient has not expressed a clear preference for or against mechanical ventilation. Therefore, active treatment is opposed to mechanical ventilation. Two of his children fail to confirm this and request active treatment.

The patient has been deteriorating for 3 years, with a rapid decline in cognition over the last 3 months, and he needs help to carry out activities of daily living. A neurologist has expressed differing views. In terms of the duty of a doctor to make the patient’s health the first consideration, and of the patient’s right to appropriate health care regardless of his age or mental condition, it would therefore be appropriate to institute all possible care, including ventilation on an intensive care unit.

On the one hand, considered mainly in teleological terms:

The patient is incapable of making an autonomous decision. The closest surrogate indicates that he would have preferred to forego life-sustaining therapy at this stage. (Respect for autonomy might support this.) The consequences of ventilation would probably be to prolong the process of dying (which non-maleficence could argue against) rather than increase his chances of recovery to a good quality of life. Beneficence requires that he receive general care and symptom relief immediately. An appropriate action therefore is not to ventilate the patient but to continue basic medical (fluids, oxygen and antibiotics) and nursing care in a general ward setting in order to optimise patient comfort.

On the other hand, considered in deontological as well as teleological terms:

The present illness is due to a potentially reversible infection. The patient’s real preference is uncertain and his family, who have difficulty in looking after him, have expressed differing views. In terms of the duty of a doctor to make the patient’s health the first consideration, and of the patient’s right to appropriate health care regardless of his age or mental condition, it would therefore be appropriate to institute all possible care, including ventilation on an intensive care unit.

In practice:

The physician responsible for the patient’s care should consider the different courses of action suggested, but not determined, by these ethical analyses, explain the reasons for and against each course of action to the patient’s family and, if one of them is the patient’s legal surrogate, help that person come to a decision. Where there is no legal surrogate, the physician will have to reach a judgement about what is in the patient’s best interests, recognising that, while judgement is always fallible, whatever decision is made must be defensible if challenged on ethical or legal grounds. Decisions that are reached on the basis of ethical and moral reasoning will be relatively easy to defend.

In this case, further discussion of the relevant issues with the relatives and other members of the health-care team led to concordance. The patient was treated by artificial ventilation in the intensive care unit for 3 days. He made a good recovery and appeared grateful for the care he had received.

### A clinical ethics scenario

A 70-year-old man who has chronic obstructive pulmonary disease, hypertension and diabetes mellitus is admitted to hospital with pneumonia. His memory has been deteriorating for 3 years, with a rapid decline in cognition over the last 3 months, and he needs help to carry out activities of daily living. A neurologist has excluded reversible causes of dementia. The patient deteriorates and needs mechanical ventilation. His wife states that he told her (when he was well) that he did not want to be put on ‘life support machines’ and is therefore opposed to mechanical ventilation. Two of his children fail to confirm this and request active treatment. There is no formal written ‘advance directive’ on file (pp. 171 and 291). What care should be given?

### Medical law

The law impinges on medical practice in many ways. Although a description of specific laws in different countries is beyond the scope of this book, it is important for doctors to be familiar with local legislation. Some of the ethical principles described above are captured in legislation, for example, in relation to informed consent (p. 10) and confidentiality (p. 11). Other laws enforce standard requirements for formal procedures, such as death certification. In many countries, regulatory authorities with statutory powers – for example, to license doctors to practise – also impose standards. The distinction and overlap between these domains are illustrated in Box 1.12.
A high-profile area of overlap between medicine and the law occurs in legal action (litigation) related to processes of care. The latter frequently involves the concept of negligence. In the UK, the ‘Bolam test’ is often used to define whether medical care is or is not negligent. Care is measured against what any ‘ordinarily competent’ (or sometimes ‘reasonable’) doctor would have done in the same situation. In addition to this test, it must also be established that:

- there was a duty of care between the doctor and patient (this is usually straightforward)
- there was a causal link between any breach of duty and harm to the patient
- the harm was not too remote from the episode of care.

### Personal and professional development

Good doctors never stop learning, and continue to develop their knowledge, skills and attributes throughout their working lives, to the benefit of their patients and themselves. Many also participate actively in improving medical knowledge and practice through research. These activities have become an essential component of clinical governance, which is a mechanism for ensuring high standards of clinical care (Box 1.13). Personal and professional development (PPD) requires a reflective and self-directed approach to the study and practice of medicine (Fig. 1.6), and will maximise both lifelong effectiveness and personal satisfaction. Linked to this is the concept of the novice–expert shift (Box 1.14).

PPD begins in the first days at medical school and continues through postgraduate training and subsequent professional practice. Maintaining competence and expertise requires continuous professional development (CPD). In the UK, this is formally regulated by professional bodies such as the Royal Colleges, and is linked to processes of appraisal (Box 1.15) and re-accreditation for established practitioners.

To support this process, outcomes and competences for PPD are being defined at all levels of medical training, including undergraduate and postgraduate study. These sit alongside and complement curricula that focus on discipline-based knowledge and skills. As adult learners, doctors are expected to reflect on their own practice and identify their own particular learning or developmental needs. This recognises that doctors will have different learning needs throughout their career, which will be affected by their current clinical practice, their future career plans and any areas of educational need that have become apparent through the appraisal process.

### Table: Some definitions in ethics, law and regulation of medical practice

<table>
<thead>
<tr>
<th>Definition</th>
<th>In practice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ethics</strong></td>
<td>The science of morality; a branch of philosophy concerned with human character and conduct</td>
</tr>
<tr>
<td><strong>Law</strong></td>
<td>Rules of action established by authority (normally, a community or state)</td>
</tr>
<tr>
<td><strong>Negligence</strong></td>
<td>Omission of duty and such care for the interests of others as the law may require</td>
</tr>
<tr>
<td><strong>(Practice) guidance</strong></td>
<td>Direction from another person or body</td>
</tr>
<tr>
<td><strong>(Clinical) governance</strong></td>
<td>Control, autonomy</td>
</tr>
</tbody>
</table>

### Key components in clinical governance

- Continuing education
- Clinical audit
- Clinical effectiveness
- Risk management
- Research and development
- Openness

### Novice–expert shift

- **Novices** use pre-determined methods which they learn
- **Advanced beginners** recognise that these methods are not effective in all circumstances and can adapt them
- **Competent professionals** are able to make conscious, independent choices and can manage and regulate their own practice
- **Proficient professionals** make use of intuition based on experience, and integrate multiple aspects of practice into a holistic model
- **Experts** function largely through ‘unconscious competence’ and are inseparable from the tasks they undertake

---

**Fig. 1.6** The personal and professional development of a doctor.
Each doctor has a duty to ensure that their clinical knowledge and skills are up to date and comparable with their peers. Clinical audit is one method of assessing practice in this context.

**Clinical audit**

Clinical audit is the process by which the clinical practice of a doctor or medical team and the outcomes of that practice are evaluated against an agreed standard. Where practice fails to meet the standard, changes to practice are implemented; after a period, practice can be re-evaluated to identify any improvement. The continuing evaluation, implementation of change and re-evaluation process is known as the audit loop or cycle (Fig. 1.7). The standard against which practice is measured is usually an externally agreed one, rather than a local one. It is important to know that clinical care is comparable to that delivered elsewhere. For this reason, national standards are the norm in most countries, often set alongside national guidelines which signpost the practice necessary to achieve them. Clinical audit may be conducted by the doctor or team themselves, or by an external body. Outcome measures may include success rates or complication rates of clinical procedures such as surgical operations; process variables such as waiting times for clinical care; or the perspective of patients and relatives. In the UK, all practising clinicians are now expected to participate in audit, and it is an integral part of procedures for appraisal, revalidation and relicensing of doctors.

### COMPLEMENTARY AND ALTERNATIVE MEDICINE

Complementary and alternative medicine (CAM) refers to a group of medical and health-care systems, practices and products that are not considered to be part of conventional medicine; as such, the relevant principles and skills are not included in the curricula of conventional medical education programmes. CAM covers an enormous and ever-changing range of activities, from well-established physical therapies such as osteopathy to spiritual measures such as prayer specifically for health. Proponents suggest that CAM focuses on the whole person: their lifestyle, environment, diet, and mental, emotional and spiritual health, as well as physical complaints.

‘Complementary medicine’ is the term used to describe the use of these treatments in conjunction with conventional medicine (e.g. acupuncture to reduce pain after surgery). ‘Alternative medicine’ describes their use in place of conventional medicine (e.g. reflexology instead of anti-inflammatory drugs for arthritis). Clearly, most forms of treatment can be used in either way, so the term CAM is often used generically. ‘Integrative medicine’ describes the use of conventional therapy in combination with one or more complementary therapies.

A variety of different taxonomies are used for CAM therapies. The National Center for Complementary and Alternative Medicine in the USA uses the following classification:

- **Alternative medical systems.** These have their own constructs of theory and practice, often based on ancient historical beliefs. Examples are homeopathy, naturopathy, traditional Chinese medicine and Ayurveda.
- **Mind–body interactions.** These rely on the mind’s capacity to influence physical function. Examples are meditation, biofeedback, prayer for healing, mental healing, music therapy and dance.
- **Biologically based therapies.** These involve the use or regulation of an extraneous agent or preparation. Examples include herbal medicine, dietary supplementation and nutritional medicine.
- **Manipulative and body-based methods.** These are based on manipulation or movement of parts of the body. They include osteopathy, chiropractic, reflexology and massage.
- **Energy therapies.** These involve use of energy fields. Examples include qigong, reiki and therapeutic touch.

Some forms of CAM are embedded in the cultural norms of particular social and ethnic groups, e.g. traditional Chinese medicine. In Western society, the use of CAM is extensive. For example, in 2007 in the USA, 38% of the adult population had used some form of CAM in the previous year (males 33.5%, females 42.8%); 12% of children had also used CAM. The most common medical conditions involved were back pain, neck pain, other joint pain/arthritis, anxiety, raised cholesterol, head or chest ‘cold’s’, headache, insomnia, stress and depression, and gastrointestinal symptoms.

The popularity of CAM may reflect a lack of confidence in conventional medicine, particularly a belief that...
GOOD MEDICAL PRACTICE

it will not help the condition or may cause harm. CAM is often used by cancer patients who have disease which is unresponsive to conventional medicines. In addition, it may reflect the increasing ease of access to information and therapies via the Internet. CAM is often perceived to be completely safe; patients may therefore be willing to experiment with it as a ‘no-lose’ measure. Many forms of CAM are inherently pleasurable, regardless of any therapeutic benefit.

Safety

Not all CAM therapies are safe; some are toxic in their own right (e.g. dietary supplements containing ephedrine alkaloids, now banned in the USA) and others are harmful if used in combination with conventional treatment (e.g. garlic supplements that interfere with the action of anti-HIV chemotherapy). Others have been associated with rare but serious side-effects, which can be life-threatening (e.g. bowel perforation from coffee enemas, hyponatraemia from noni juice).

There is also a potential for harm when alternative medicine is used to treat serious or life-threatening medical conditions, if the resultant delay in seeking conventional treatment compromises clinical outcome.

On balance, however, the relative safety of most CAM therapies can be regarded as a positive feature; homeopathy is an example.

Evidence

In an era where EBM is the norm, practitioners and advocates of CAM are increasingly challenged to justify these treatments through independent, well-conducted, randomised controlled clinical trials. In some cases, this may be difficult (e.g. the placebo arm of a double-blind trial of acupuncture). In addition, it can be argued that different types and standards of evidence, focusing on patient satisfaction and subjective benefit rather than measurable clinical outcomes, may be more appropriate for CAM. The literature in this area is growing rapidly but, at present, only a minority of CAM therapies are supported by any evidence that would be acceptable for conventional medicine. These are primarily the ‘big five’ CAM therapies: osteopathy, chiropractic, acupuncture, homeopathy and herbal medicine. Moreover, where such positive evidence does exist, it is often outweighed by negative studies, and limited to a small subset of the clinical conditions for which the treatment is used.

Regulation

Many CAM therapies have professional regulatory frameworks in place and others are following suit. Nevertheless, for many CAM therapies, there is still no established structure of training, certification and accreditation, and practice is effectively open to all. Set against the demanding training and life-long continuous professional development that pertain to conventional medicine, this constitutes an important barrier to integrative medicine.

Integrated health care

There is a considerable impetus behind moves to integrate CAM with conventional medicine and health care at the level of resource allocation, service design, clinical practice, education and research. Almost 50% of general practices in the UK and an increasing number of hospitals offer some form of access to CAM. In many parts of Asia in particular, this kind of medical pluralism is the norm, and patients do not necessarily make a distinction between different systems of health care. Historically, in Western societies, patients using both types of therapy have often experienced conflicting advice and value judgements, poor or absent communication between practitioners, and even hostility or ridicule. They often revert to secrecy, an inherently undesirable and potentially dangerous outcome. Integrated health care aims to understand and remove the barriers that create such dilemmas for patients. It aims to let them exercise their choice of treatment in an open environment characterised by good communication, respect, and due consideration of autonomy, efficacy and risk.

Further information and acknowledgements

Websites

www.dh.gov.uk UK Department of Health guidance and policy on confidentiality and consent.

www.evidence.nhs.uk A UK National Health Service resource providing a searchable library of clinical guidelines from all sources.

www.gmc-uk.org UK General Medical Council. Includes access to guidance on professional conduct (Duties of a Doctor, Good Medical Practice) and guidance on medical education, such as ‘Tomorrow’s Doctors’.


www.who.int World Health Organization. Includes information relevant to global health and differences in medical practice.

Figure acknowledgements

Fig. 1.4 Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. BMJ 2002; 324:827–830, reproduced with permission from the BMJ Publishing Group.
Principles of clinical pharmacology 18
Pharmacodynamics 18
Pharmacokinetics 21
Inter-individual variation in drug responses 23

Adverse outcomes of drug therapy 24
Adverse drug reactions 24
Drug interactions 28
Medication errors 29

Drug regulation and management 30
Drug development and marketing 30
Managing the use of medicines 31

Prescribing in practice 33
Decision-making in prescribing 33
Prescribing in special circumstances 36
Writing prescriptions 37
Monitoring drug therapy 39
Prescribing medicines is a major tool used by most doctors to restore or preserve the health of their patients. Medicines contain drugs (the specific chemical substances with pharmacological effects), either alone or in combination, in a formulation mixed with other ingredients. The beneficial effects of medicines must be weighed against their cost and the risks of adverse drug reactions and interactions, often caused by injudicious prescribing decisions and by prescribing errors. The modern prescriber must meet the challenges posed by an increasing number of drugs and formulations available and of indications for prescribing them, and the greater complexity of treatment regimens followed by individual patients (‘polypharmacy’, a particular challenge in the ageing population). The purpose of this chapter is to elaborate on the principles and practice that underpin good prescribing (Box 2.1).

### 2.1 Steps in good prescribing

- Make a diagnosis
- Consider factors influencing the patient’s responses to therapy (age, concomitant drug therapy, renal and liver function etc.)*
- Establish the therapeutic goal*
- Choose the therapeutic approach*
- Choose the drug and its formulation (the ‘medicine’)
- Choose the dose, route and frequency
- Choose the duration of therapy
- Write an unambiguous prescription (or ‘medication order’)
- Inform the patient about the treatment and its likely effects
- Monitor treatment effects, both beneficial and harmful
- Review/alter the prescription

*These steps in particular take the patient’s views into consideration to establish a therapeutic partnership.

### PRINCIPLES OF CLINICAL PHARMACOLOGY

Prescribers need to understand what the drug does to the body (pharmacodynamics) and what the body does to the drug (pharmacokinetics) (Fig. 2.1). Although this chapter is focused on the most common drugs, which are synthetic small molecules, the same principles apply to the increasingly numerous ‘biological’ therapies (sometimes abbreviated to ‘biologics’) now in use, which include peptides, proteins, enzymes and monoclonal antibodies (p. 74).

#### Pharmacodynamics

**Drug targets and mechanisms of action**

Modern drugs are usually discovered by screening compounds for activity either to stimulate or to block the function of a specific molecular target, which is predicted to have a beneficial effect in a particular disease (Box 2.2). Other drugs have useful but less selective chemical properties, such as chelators (e.g. for treatment of iron or copper overload), osmotic agents (used as diuretics in cerebral oedema) or general anaesthetics (that alter the biophysical properties of lipid membranes). The following characteristics of the interaction of drugs with receptors illustrate some of the important determinants of the effects of drugs:

- **Affinity** describes the propensity for a drug to bind to a receptor and is related to the ‘molecular fit’ and the strength of the chemical bond. Some drug–receptor interactions are irreversible, either because the affinity is so strong or because the drug modifies the structure of its molecular target.
- **Selectivity** describes the propensity for a drug to bind to one target rather than another. Selectivity is a relative term, not to be confused with absolute specificity. It is common for drugs targeted at a particular subtype of receptor to exhibit some effect at other subtypes. For example, β-adrenoceptors can be subtyped on the basis of their responsiveness to the endogenous agonist noradrenaline (norepinephrine): the concentration of noradrenaline required to cause bronchodilation via β2-adrenoceptors is ten times higher than that required to cause tachycardia via β1-adrenoceptors. ‘Cardioselective’ β-blockers have anti-anginal effects on the heart (β1) but may still cause bronchospasm in the lung (β2) and are contraindicated for asthmatic patients.
- **Agonists** bind to a receptor to produce a conformational change that is coupled to a biological response. As agonist concentration increases, so does the proportion of receptors occupied, and hence the biological effect. Partial agonists activate the receptor, but cannot produce a maximal signalling effect equivalent to that of a full agonist even when all available receptors are occupied.
- **Antagonists** bind to a receptor but do not produce the conformational change that initiates an intracellular signal. A competitive antagonist competes with endogenous ligands to occupy receptor binding sites, with the resulting antagonism depending on the relative affinities and concentrations of drug and ligand. **Non-competitive**
## 2.2 Examples of target molecules for drugs

<table>
<thead>
<tr>
<th>Drug target</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Receptors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel-linked receptors</td>
<td>Ligand binding controls a linked ion channel, known as ‘ligand-gated’ (in contrast to ‘voltage-gated’ channels that respond to changes in membrane potential)</td>
<td>Nicotinic acetylcholine receptor, γ-aminobutyric acid (GABA) receptor, Sulphonylurea receptor</td>
</tr>
<tr>
<td>G-protein-coupled receptors (GPCRs)</td>
<td>Ligand binding affects one of a family of ‘G-proteins’ that mediate signal transduction either by activating intracellular enzymes (such as adenylylate or guanylate cyclase, producing cyclic AMP or GMP, respectively) or by controlling ion channels</td>
<td>Muscarinic acetylcholine receptor, β-adrenoceptors, Dopamine receptors, Serotonin receptors, Opioid receptors</td>
</tr>
<tr>
<td>Kinase-linked receptors</td>
<td>Ligand binding activates an intracellular protein kinase that triggers a cascade of phosphorylation reactions</td>
<td>Insulin receptor, Cytokine receptors</td>
</tr>
<tr>
<td>Transcription factor receptors</td>
<td>Intracellular and also known as ‘nuclear receptors’; ligand binding promotes or inhibits gene transcription and hence synthesis of new proteins</td>
<td>Steroid receptors, Thyroid hormone receptors, Vitamin D receptors, Retinoid receptors, PPAR Y and α receptors</td>
</tr>
</tbody>
</table>

| Other targets                   |                                                                             |                                                                          |
|---------------------------------|                                                                             |                                                                          |
| Voltage-gated ion channels      | Mediate electrical signalling in excitable tissues (muscle and nervous system) | Na⁺ channels, Ca²⁺ channels, Cyclo-oxygenase, Angiotensin converting enzyme (ACE), Xanthine oxidase |
| Enzymes                         | Catalyse biochemical reactions. Drugs interfere with binding of substrate to the active site or of co-factors | Serotonin re-uptake transporter, Na⁺/K⁺ ATPase                          |
| Transporter proteins            | Carry ions or molecules across cell membranes                              |                                                                          |

(AMP = adenosine monophosphate; ATPase = adenosine triphosphatase; GMP = guanosine monophosphate; PPAR = peroxisome proliferator-activated receptor)

---

**Antagonists** inhibit the effect of an agonist by mechanisms other than direct competition for receptor binding with the agonist (e.g. by affecting post-receptor signalling).

### Dose–response relationships

Plotting the logarithm of drug dose against drug response typically produces a sigmoidal dose–response curve (Fig. 2.2). Progressive increases in drug dose (which for most drugs is proportional to the plasma drug concentration) produce increasing response, but only within a relatively narrow range of dose; further increases in dose beyond this range produce little extra effect. The following characteristics of the drug response are useful in comparing different drugs:

- **Efficacy** describes the extent to which a drug can produce a target-specific response when all available receptors or binding sites are occupied

![Fig. 2.2 Dose–response curve](image-url)
THERAPEUTICS AND GOOD PRESCRIBING

(i.e. $E_{\text{max}}$ on the dose–response curve). A full agonist can produce the maximum response of which the receptor is capable, while a partial agonist at the same receptor will have lower efficacy. Therapeutic efficacy describes the effect of the drug on a desired biological endpoint, and can be used to compare drugs that act via different pharmacological mechanisms (e.g. loop diuretics induce a greater diuresis than thiazide diuretics and therefore have greater therapeutic efficacy).

- Potency describes the amount of drug required for a given response. More potent drugs produce biological effects at lower doses, so they have a lower $ED_{50}$. A less potent drug can still have an equivalent efficacy if it is given in higher doses.

The dose–response relationship varies between patients because of variations in the many determinants of pharmacokinetics and pharmacodynamics. In clinical practice, the prescriber is unable to construct a dose–response curve for each individual patient. Therefore, most drugs are licensed for use within a recommended range of doses that is expected to reach close to the top of the dose–response curve for most patients. However, it is sometimes possible to achieve the desired therapeutic efficacy at doses towards the lower end of, or even below, the recommended range.

Therapeutic index

The adverse effects of drugs are often dose-related in a similar way to the beneficial effects, although the dose–response curve for these adverse effects is normally shifted to the right (see Fig. 2.2). The ratio of the $ED_{50}$ for therapeutic efficacy and for a major adverse effect is known as the ‘therapeutic index’. In reality, drugs have multiple potential adverse effects but the concept of therapeutic index is usually reserved for those requiring dose reduction or discontinuation. For most drugs, the therapeutic index is greater than 100 but there are some notable exceptions with therapeutic indices less than 10 (e.g. digoxin, warfarin, insulin, phenytoin, opioids). The doses of such drugs have to be titrated carefully for individual patients to maximise benefits but avoid adverse effects.

Desensitisation and withdrawal effects

Desensitisation refers to the common situation in which the biological response to a drug diminishes when it is given continuously or repeatedly. It may be possible to restore the response by increasing the dose of the drug but, in some cases, the tissues may become completely refractory to its effect.

- Tachyphylaxis describes desensitisation that occurs very rapidly, sometimes with the initial dose. This rapid loss of response implies depletion of chemicals that may be necessary for the pharmacological actions of the drug (e.g. a stored neurotransmitter released from a nerve terminal) or receptor phosphorylation.

- Tolerance describes a more gradual loss of response to a drug that occurs over days or weeks. This slower change implies changes in receptor numbers or the development of counter-regulatory physiological changes that offset the actions of the drug (e.g. accumulation of salt and water in response to vasodilator therapy).

- Drug resistance is a term normally reserved for describing the loss of effectiveness of an antimicrobial (p. 151) or cancer chemotherapy drug.

- In addition to these pharmacodynamic causes of desensitisation, reduced response may be the consequence of lower plasma and tissue drug concentrations as a result of altered pharmacokinetics (see below).

| 2.3 Examples of drugs associated with withdrawal effects |
|----------------|-----------------|----------------|-----------------|
| **Drug**       | **Symptoms**    | **Signs**      | **Treatment**   |
| Alcohol        | Anxiety, panic, paranoid delusions, visual and auditory hallucinations | Agitation, restlessness, confusion, tremor, tachycardia, ataxia, disorientation, seizures | Treat immediate withdrawal syndrome with benzodiazepines |
| Barbiturates, benzodiazepines | Similar to alcohol | Similar to alcohol | Transfer to long-acting benzodiazepine then gradually reduce dosage |
| Corticosteroids | Weakness, fatigue, decreased appetite, weight loss, nausea, vomiting, diarrhoea, abdominal pain | Hypotension, hypoglycaemia | Prolonged therapy suppresses the hypothalamic–pituitary–adrenal axis and causes adrenal insufficiency requiring corticosteroid replacement. Withdrawal should be gradual after prolonged therapy (p. 776) |
| Opioids        | Rhinorrhoea, sneezing, yawning, lacrimation, abdominal and leg cramping, nausea, vomiting, diarrhoea | Dilated pupils | Transfer addicts to long-acting agonist methadone |
| Selective serotonin re-uptake inhibitors (SSRIs) | Dizziness, sweating, nausea, insomnia, tremor, confusion, nightmares | Tremor | Reduce SSRIs slowly to avoid withdrawal effects |
When drugs induce chemical, hormonal and physiological changes that offset their actions, discontinuation may allow these changes to cause ‘rebound’ withdrawal effects (Box 2.3).

**Pharmacokinetics**

Understanding ‘what the body does to the drug’ (Fig. 2.3) is extremely important for prescribers because this forms the basis on which the optimal route of administration and dose regimen are chosen and explains the majority of inter-individual variation in the response to drug therapy.

**Drug absorption and routes of administration**

Absorption is the process by which drug molecules gain access to the blood stream. The rate and extent of drug absorption depend on the route of administration (see Fig. 2.3).

**Enteral administration**

These routes involve administration via the gastrointestinal tract:

- **Oral.** This is the commonest route of administration because it is simple, convenient and readily used by patients to self-administer their medicines. Absorption after an oral dose is a complex process that depends on the drug being swallowed, surviving exposure to gastric acid, avoiding unacceptable food binding, being absorbed across the small bowel mucosa into the portal venous system, and surviving metabolism by gut wall or liver enzymes (‘first-pass metabolism’). As a consequence, absorption is frequently incomplete following oral administration. The term ‘bioavailability’ describes the proportion of the dose that reaches the systemic circulation intact.

- **Buccal, intranasal and sublingual (SL).** These routes have the advantage of enabling rapid absorption into the systemic circulation without the uncertainties associated with oral administration (e.g. organic nitrates for angina pectoris, triptans for migraine, opioid analgesics).

- **Rectal (PR).** The rectal mucosa is occasionally used as a site of drug administration when the oral route is compromised because of nausea and vomiting or unconsciousness (e.g. diazepam in status epilepticus).

**Parenteral administration**

These routes avoid absorption via the gastrointestinal tract and first-pass metabolism in the liver:

- **Intravenous (IV).** The IV route enables all of a dose to enter the systemic circulation reliably, without any concerns about absorption or first-pass metabolism (i.e. the dose is 100% bioavailable), and rapidly achieve a high plasma concentration. It is ideal for very ill patients when a rapid, certain effect is critical to outcome (e.g. benzylpenicillin for meningococcal meningitis).

- **Intramuscular (IM).** IM administration is easier to achieve than the IV route (e.g. adrenaline (epinephrine) for acute anaphylaxis) but absorption is less predictable and depends on muscle blood flow.

- **Subcutaneous (SC).** The SC route is ideal for drugs that have to be administered parenterally because of low oral bioavailability, are absorbed well from subcutaneous fat, and might ideally be injected by patients themselves (e.g. insulin, heparin).

- **Transdermal.** A transdermal patch can enable a drug to be absorbed through the skin and into the circulation (e.g. oestrogens, testosterone, nicotine, nitrates).

**Other routes of administration**

- **Topical** application of a drug involves direct administration to the site of action (e.g. skin, eye, ear). This has the advantage of achieving sufficient concentration at this site while minimising systemic exposure and the risk of adverse effects elsewhere.

- **Inhaled (INH) administration** allows drugs to be delivered directly to a target in the respiratory tree, usually the small airways (e.g. salbutamol, Fig. 2.3 Pharmacokinetics summary. Most drugs are taken orally, are absorbed from the intestinal lumen and enter the portal venous system to be conveyed to the liver, where they may be subject to first-pass metabolism and/or excretion in bile. Active drugs then enter the systemic circulation, from which they may diffuse (or sometimes be actively transported) in and out of the interstitial and intracellular fluid compartments. Drug that remains in circulating plasma is subject to liver metabolism and renal excretion. Drugs excreted in bile may be reabsorbed, creating an enterohepatic circulation. First-pass metabolism in liver is avoided if drugs are administered via the buccal or rectal mucosa, or parenterally (e.g. by intravenous injection).
beclometasone). However, a significant proportion of the inhaled dose may be absorbed from the lung or is swallowed and can reach the systemic circulation. The most common mode of delivery is the metered-dose inhaler but its success depends on some degree of manual dexterity and timing (see Fig. 19.23, p. 670). Patients who find these difficult may use a ‘spacer’ device to improve drug delivery. A special mode of inhaled delivery is via a nebulised solution created by using pressurised oxygen or air to break up solutions and suspensions into small aerosol droplets that can be directly inhaled from the mouthpiece of the device.

**Drug distribution**

Distribution is the process by which drug molecules transfer into and out of the blood stream. This is influenced by the drug’s molecular size and lipid solubility, the extent to which it binds to proteins in plasma, its susceptibility to drug transporters expressed on cell surfaces, and its binding to its molecular target and to other cellular proteins (which can be irreversible). Most drugs diffuse passively across capillary walls down a concentration gradient into the interstitial fluid until the concentration of free drug molecules in the interstitial fluid is equal to that in the plasma. As drug molecules in the blood are removed by metabolism or excretion, the plasma concentration falls and drug molecules diffuse back from the tissue compartment into the blood, and eventually all will be eliminated. Note that this reverse movement of drug away from the tissues will be prevented if further drug doses are administered and absorbed into the plasma.

**Volume of distribution**

The apparent volume of distribution (V_d) is the volume into which a drug appears to have distributed following intravenous injection. It is calculated from the equation

\[ V_d = \frac{D}{C_0} \]

where D is the amount of drug given and C_0 is the initial plasma concentration (Fig. 2.4A). Drugs that are highly bound to plasma proteins may have a V_d below 10 L (e.g. warfarin, aspirin), while those that diffuse into the interstitial fluid but do not enter cells because they have low lipid solubility may have a V_d between 10 and 30 L (e.g. gentamicin, amoxicillin). It is an ‘apparent’ volume because those drugs that are lipid-soluble and highly tissue-bound may have a V_d of greater than 100 L (e.g. digoxin, amitriptyline). Drugs with a larger V_d are eliminated more slowly from the body.

**Drug elimination**

**Drug metabolism**

Metabolism is the process by which drugs are chemically altered from a lipid-soluble form suitable for absorption and distribution to a more water-soluble form that is necessary for excretion. Some drugs, known as ‘pro-drugs’, are inactive in the form in which they are administered, but are converted to an active metabolite in vivo.

Phase I metabolism involves oxidation, reduction or hydrolysis to make drug molecules suitable for phase II reactions or for excretion. Oxidation is much the commonest form of phase I reaction and chiefly involves members of the cytochrome P450 family of membrane-bound enzymes in the endoplasmic reticulum of hepatocytes.

Phase II metabolism involves combining phase I metabolites with an endogenous substrate to form an inactive conjugate that is much more water-soluble. Reactions include glucuronidation, sulphation, acetylation or methylation, and conjugation with glutathione. This is necessary to enable renal excretion because lipid-soluble metabolites will simply diffuse back into the body after glomerular filtration (p. 430).

**Drug excretion**

Excretion is the process by which drugs and their metabolites are removed from the body.

Renal excretion is the usual route of elimination for drugs or their metabolites that are of low molecular weight and sufficiently water-soluble to avoid reabsorption from the renal tubule. Drugs bound to plasma proteins are not filtered by the glomeruli. The pH of the urine is more acidic than that of plasma, so that some drugs (e.g. salicylates) become un-ionised and tend to be reabsorbed. Alkalination of the urine can hasten excretion (e.g. after a salicylate overdose). For some drugs, active secretion into the proximal tubule lumen, rather than glomerular filtration, is the predominant mechanism of excretion (e.g. methotrexate, penicillin).

Faecal excretion is the predominant route of elimination for drugs with high molecular weight, including those that are excreted in the bile after conjugation with glucuronide in the liver, and any drugs that are not absorbed after enteral administration. Molecules of drug or metabolite that are excreted in the bile enter the small intestine, where they may, if they are sufficiently lipid-soluble, be reabsorbed through the gut wall and return to the liver via the portal vein (see Fig. 2.3). This recycling between the liver, bile, gut and portal vein is known as ‘enterohepatic circulation’ and can significantly prolong the residence of drugs in the body.

**Elimination kinetics**

The net removal of drug from the circulation results from a combination of drug metabolism and excretion, and is usually described as ‘clearance’, i.e. the volume of plasma that is completely cleared of drug per unit time.

For most drugs, elimination is a high-capacity process that does not become saturated, even at high dosage. The rate of elimination is therefore directly proportional to the drug concentration because of the ‘law of mass action’, whereby higher drug concentrations will drive faster metabolic reactions and support higher renal filtration rates. This results in ‘first-order’ kinetics, when a constant fraction of the drug remaining in the circulation is eliminated in a given time and the decline in concentration over time is exponential (see Fig. 2.4A). This elimination can be described by the drug’s half-life (t_1/2), i.e. the time taken for the plasma drug concentration to halve, which remains constant throughout the period of drug elimination. The significance of this phenomenon for prescribers is that the effect of increasing doses on plasma concentration is predictable – a doubled dose leads to a doubled concentration at all time points.
After multiple drug dosing. serious toxicity. titration, the drug will accumulate progressively, leading to if the rate of administration exceeds the maximum rate of elimination. Its significance for prescribers is that, if the rate of absorption (the product of concentration and clearance) is equal to the rate of drug absorption (the product of rate of administration and bioavailability). The long half-life in this example means that it takes 6 days for steady state to be achieved and, for most of the first 3 days of treatment, plasma drug concentrations are below the therapeutic range (yellow-shaded area). This problem can be overcome if a larger loading dose (red line) is used to achieve steady state drug concentrations more rapidly.

For a few drugs in common use (e.g. phenytoin, alcohol), elimination capacity is exceeded (saturated) within the usual dose range. This is called ‘zero-order’ kinetics. Its significance for prescribers is that, if the rate of administration exceeds the maximum rate of elimination, the drug will accumulate progressively, leading to serious toxicity.

Repeated dose regimens
The goal of therapy is usually to maintain drug concentrations within the therapeutic range (see Fig. 2.2) over several days (e.g. antibiotics) or even for months or years (e.g. antihypertensives, lipid-lowering drugs, thyroid hormone replacement therapy). This goal is rarely achieved with single doses, so prescribers have to plan a regimen of repeated doses. This involves choosing the size of each individual dose and the frequency of dose administration.

As illustrated in Figure 2.4B, the time taken to reach drug concentrations within the therapeutic range depends on the half-life of the drug. Typically, with doses administered regularly, it takes approximately 5 half-lives to reach a ‘steady state’ in which the rate of drug elimination is equal to the rate of drug administration. This applies when starting new drugs and when adjusting doses of current drugs. With appropriate dose selection, steady state drug concentrations will be maintained within the therapeutic range. This is important for prescribers because it means that the effects of a new prescription, or dose titration, for a drug with a long half-life (e.g. digoxin – 36 hours) may not be known for a few days. In contrast, drugs with a very short half-life (e.g. dobutamine – 2 minutes) have to be given continuously by infusion but reach a new steady state within minutes.

For drugs with a long half-life, if it is unacceptable to wait for 5 half-lives until concentrations within the therapeutic range are maintained, then an initial ‘loading dose’ can be given that is much larger than the maintenance dose, and equivalent to the amount of drug required in the body at steady state. This achieves a peak plasma concentration close to the plateau concentration, which can then be maintained by successive maintenance doses.

‘Steady state’ actually involves fluctuations in drug concentrations, with peaks just after administration followed by troughs just prior to the next administration. The manufacturers of medicines recommend dosing regimens that predict that, for most patients, these oscillations result in troughs within the therapeutic range and peaks that are not high enough to cause adverse effects. The optimal dose interval is a compromise between convenience for the patient and a constant level of drug exposure. More frequent administration (e.g. 25 mg 4 times daily) achieves a smoother plasma concentration profile than 100 mg once daily but is much more difficult for patients to sustain. A solution to this need for compromise in dosing frequency for drugs with half-lives of less than 24 hours is the use of ‘modified-release’ formulations. These allow drugs to be absorbed more slowly from the gastrointestinal tract and reduce the oscillation in plasma drug concentration profile, which is especially important for drugs with a low therapeutic index (e.g. levodopa).

Inter-individual variation in drug responses
Prescribers have numerous sources of guidance about how to use drugs appropriately (e.g. dose, route, frequency, duration) for many conditions. However, this advice is based on average dose–response data derived from observations in many individuals. When applying this information to an individual patient, prescribers
must take account of inter-individual variability in response. Some of this variability is predictable and good prescribers are able to anticipate it and adjust their prescriptions accordingly to maximise the chances of benefit and minimise harm. Inter-individual variation in responses also mandates that effects of treatment should be monitored (p. 39).

Some inter-individual variation in drug response is accounted for by differences in pharmacodynamics. For example, the beneficial natriuresis produced by the loop diuretic furosemide is often significantly reduced at a given dose in patients with renal impairment, while confusion caused by opioid analgesics is more likely in the elderly. However, differences in pharmacokinetics more commonly account for different drug responses. Examples of factors influencing the absorption, metabolism and excretion of drugs are shown in Box 2.4.

It is hoped that a significant proportion of the inter-individual variation in drug responses can be explained by studying genetic differences in single genes (‘pharmacogenetics’) (Box 2.5) or the effects of multiple gene variants (‘pharmacogenomics’). The aim is to identify those patients most likely to benefit from specific treatments and those most susceptible to adverse effects. In this way, it may be possible to select drugs and dose regimens for individual patients to maximise the benefit-hazard ratio (‘personalised medicine’).

### Adverse drug reactions

Some important definitions for the adverse effects of drugs are:

- **Adverse event.** A harmful event that occurs while a patient is taking a drug, irrespective of whether the drug is suspected of being the cause.

- **Adverse drug reaction (ADR).** An unwanted or harmful reaction that is experienced following the administration of a drug or combination of drugs under normal conditions of use and is suspected to be related to the drug. An ADR will usually require the drug to be discontinued or the dose reduced.

- **Side-effect.** Any effect caused by a drug other than the intended therapeutic effect, whether beneficial, neutral or harmful. The term ‘side-effect’ is often used interchangeably with ‘ADR’, although the former usually implies an effect that is less harmful, is predictable and may not require discontinuation of therapy (e.g. ankle oedema with vasodilators).

- **Drug toxicity.** Adverse effects of a drug that occur because the dose or plasma concentration has risen above the therapeutic range, either unintentionally or intentionally (drug overdose, see Fig. 2.2, p. 19).

- **Drug abuse.** The misuse of recreational or therapeutic drugs that may lead to addiction or dependence, serious physiological injury (such as liver damage), psychological harm (abnormal behaviour patterns, hallucinations, memory loss) or death (p. 240).

### Prevalence of ADRs

ADRs are a common cause of illness, accounting in the United Kingdom (UK) for approximately 3% of consultations in primary care and 7% of emergency consulting department (p. 23). For example, the beneficial natriuresis produced by the loop diuretic furosemide is often significantly reduced at a given dose in patients with renal impairment, while confusion caused by opioid analgesics is more likely in the elderly. However, differences in pharmacokinetics more commonly account for different drug responses. Examples of factors influencing the absorption, metabolism and excretion of drugs are shown in Box 2.4.

It is hoped that a significant proportion of the inter-individual variation in drug responses can be explained by studying genetic differences in single genes (‘pharmacogenetics’) (Box 2.5) or the effects of multiple gene variants (‘pharmacogenomics’). The aim is to identify those patients most likely to benefit from specific treatments and those most susceptible to adverse effects. In this way, it may be possible to select drugs and dose regimens for individual patients to maximise the benefit-hazard ratio (‘personalised medicine’).
Adverse outcomes of drug therapy

2.5 Examples of pharmacogenetic variations that influence drug response

<table>
<thead>
<tr>
<th>Genetic variant</th>
<th>Drug affected</th>
<th>Clinical outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pharmacokinetic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aldehyde dehydrogenase-2</td>
<td>Ethanol</td>
<td>Elevated blood acetaldehyde causes facial flushing and increased heart rate in ~50% of Japanese, Chinese and other Asian populations</td>
</tr>
<tr>
<td>deficiency</td>
<td></td>
<td>Increased responses in slow acetylator, up to 50% of some populations</td>
</tr>
<tr>
<td>Acetylation</td>
<td>Isoniazid, hydralazine,</td>
<td>Increased risk of toxicity in poor metabolisers</td>
</tr>
<tr>
<td></td>
<td>procainamide</td>
<td></td>
</tr>
<tr>
<td>Oxidation (CYP2D6)</td>
<td>Nortriptyline, Codeine</td>
<td>Reduced responses with slower conversion of codeine to more active</td>
</tr>
<tr>
<td></td>
<td></td>
<td>morphine in poor metabolisers, 10% of European populations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increased risk of toxicity in ultra-fast metabolisers, 3% of Europeans but</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40% of North Africans</td>
</tr>
<tr>
<td>Oxidation (CYP2C18)</td>
<td>Proguanil</td>
<td>Reduced efficacy with slower conversion to active cycloguanil in poor metabolisers</td>
</tr>
<tr>
<td>Oxidation (CYP2C9)</td>
<td>Warfarin</td>
<td>Polymorphisms known to influence dosages</td>
</tr>
<tr>
<td>Oxidation (CYP2C19)</td>
<td>Clopidogrel</td>
<td>Reduced enzymatic activation results in reduced antiplatelet effect</td>
</tr>
<tr>
<td>Sulphoxidation</td>
<td>Penicillamine</td>
<td>Increased risk of toxicity in poor metabolisers</td>
</tr>
<tr>
<td>HLA-B*1502</td>
<td>Carbamazepine</td>
<td>Increased risk of serious dermatological reactions (e.g. Stevens–Johnson syndrome) for 1 in 2000 in Caucasian populations (much higher in some Asian countries)</td>
</tr>
<tr>
<td>Pseudocholinesterase deficiency</td>
<td>Suxamethonium (succinylcholine)</td>
<td>Decreased drug inactivation leads to prolonged paralysis and sometimes persistent apnoea requiring mechanical ventilation until the drug can be eliminated by alternate pathways in 1 in 1500 people</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pharmacodynamic</th>
<th>Oxidant drugs including antimalarials (e.g. chloroquine, primaquine)</th>
<th>Risk of haemolysis in G6PD deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose-6-phosphate dehydrogenase (G6PD) deficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute intermittent porphyria</td>
<td>Enzyme-inducing drugs</td>
<td>Increased risk of an acute attack</td>
</tr>
<tr>
<td>SLC01B1 polymorphism</td>
<td>Statins</td>
<td>Increased risk of rhabdomyolysis</td>
</tr>
<tr>
<td>HLA-B*5701 polymorphism</td>
<td>Abacavir</td>
<td>Increased risk of skin hypersensitivity reactions</td>
</tr>
<tr>
<td>HLA-B*5801 polymorphism</td>
<td>Allopurinol</td>
<td>Increased risk of rashes in Han Chinese</td>
</tr>
<tr>
<td>HLA-B*1502 polymorphism</td>
<td>Carbamazepine</td>
<td>Increased risk of skin hypersensitivity reactions in Han Chinese</td>
</tr>
</tbody>
</table>

(HLA = human leucocyte antigen)

Adverse outcomes of drug therapy, and affecting around 15% of hospital inpatients. Many ‘disease’ presentations are eventually attributed to ADRs, emphasising the importance of always taking a careful drug history (Box 2.6). Factors accounting for the rising prevalence of ADRs are the increasing age of patients, polypharmacy (higher risk of drug interactions), increasing availability of over-the-counter medicines, increase in use of herbal or traditional medicines, and increase in medicines available via the Internet. Risk factors for ADRs are shown in Box 2.7.

ADRs are important because they reduce quality of life for patients, reduce adherence to and therefore efficacy of beneficial treatments, cause diagnostic confusion, undermine the confidence of patients in their health-care professional(s), and consume health-care resources.

Retrospective analysis of ADRs has shown that more than half could have been avoided if the prescriber had taken more care in anticipating the potential hazards of drug therapy. Each year in the UK, non-steroidal anti-inflammatory drug (NSAID) use alone accounts for 65000 emergency admissions, 12000 gastrointestinal bleeding episodes and 2000 deaths. In many cases, the patients were at increased risk due to their age, interacting drugs (e.g. aspirin, warfarin) or a past history of peptic ulcer disease. Drugs that commonly cause ADRs are listed in Box 2.8.

Prescribers and their patients ideally want to know the frequency with which ADRs occur for a specific drug. Although this may be well characterised for more common ADRs observed in clinical trials, it is less clear for rarely reported ADRs when the total numbers of reactions and patients exposed are not known. The words used to describe frequency can be misinterpreted by patients, but widely accepted meanings include: very common (10% or more), common (1–10%), uncommon (0.1–1%), rare (0.01–0.1%) and very rare (0.01% or less).

**Classification of ADRs**

ADRs have traditionally been classified into two major groups:

- **Type A (‘augmented’) ADRs.** These are predictable from the known pharmacodynamic effects of the drug, and are dose-dependent, common (detected early in drug development) and usually mild.
2.6 How to take a drug history

Information from the patient (or carer)

Use language that patients will understand (e.g. ‘medicines’ rather than ‘drugs’, which may be mistaken for drugs of abuse) while gathering the following information:

- Current prescribed drugs, including formulations (e.g. modified-release tablets), doses, routes of administration, frequency and timing, duration of treatment
- Other medications that are often forgotten (e.g. over-the-counter drugs, herbal remedies, vitamins)
- Drugs that have been taken in the recent past and reasons for stopping them
- Previous drug hypersensitivity reactions, their nature and time course (e.g. rash, anaphylaxis)
- Previous ADRs, their nature and time course (e.g. muscle aches with simvastatin)
- Adherence to therapy (e.g. ‘are you taking your medication regularly?’)

Information from the general practitioner (GP) medical records and/or pharmacist

- Up-to-date list of medications
- Previous ADRs
- Last order dates for each medication

Inspection of medicines

- Drugs and their containers (e.g. blister packs, bottles, vials) should be inspected for name, dosage, and the number of dosage forms taken since dispensed

(ADR = adverse drug reaction)

2.7 Risk factors for ADRs

Patient factors

- Elderly age (e.g. low physiological reserve)
- Gender (e.g. ACE inhibitor-induced cough in women)
- Polypharmacy (e.g. drug interactions)
- Genetic predisposition (see Box 2.5)
- Hypersensitivity/allergy (e.g. β-lactam antibiotics)
- Diseases altering pharmacokinetics (e.g. hepatic or renal impairment) or pharmacodynamic responses (e.g. bladder instability)
- Adherence problems (e.g. cognitive impairment)

Drug factors

- Steep dose–response curve (e.g. insulin)
- Low therapeutic index (e.g. digoxin, cytotoxic drugs)

Prescriber factors

- Inadequate understanding of principles of clinical pharmacology
- Inadequate knowledge of the prescribed drug
- Inadequate instructions and warnings provided to patients
- Inadequate monitoring arrangements planned

2.8 Drugs that are common causes of ADRs

<table>
<thead>
<tr>
<th>Drug or drug class</th>
<th>Common ADRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitors</td>
<td>(e.g. lisinopril) Renal impairment</td>
</tr>
<tr>
<td></td>
<td>Hyperkalaemia</td>
</tr>
<tr>
<td>Antibiotics</td>
<td>(e.g. amoxicillin) Nausea</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea</td>
</tr>
<tr>
<td>Anticoagulants</td>
<td>(e.g. warfarin, heparin)</td>
</tr>
<tr>
<td></td>
<td>Bleeding</td>
</tr>
<tr>
<td>Antipsychotics</td>
<td>(e.g. haloperidol) Falls</td>
</tr>
<tr>
<td></td>
<td>Sedation</td>
</tr>
<tr>
<td></td>
<td>Confusion</td>
</tr>
<tr>
<td>Aspirin</td>
<td>Gastrotoxicity (dyspepsia, gastrointestinal bleeding)</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>(e.g. diazepam) Drowsiness</td>
</tr>
<tr>
<td></td>
<td>Falls</td>
</tr>
<tr>
<td>β-blockers</td>
<td>(e.g. atenolol) Cold peripheries</td>
</tr>
<tr>
<td></td>
<td>Bradycardia</td>
</tr>
<tr>
<td>Calcium channel blockers</td>
<td>(e.g. amlodipine) Ankle oedema</td>
</tr>
<tr>
<td>Diuretics</td>
<td>(e.g. furosemide, bendroflumethiazide) Dehydration</td>
</tr>
<tr>
<td></td>
<td>Electrolyte disturbance</td>
</tr>
<tr>
<td></td>
<td>(hypokalaemia, hyponatraemia)</td>
</tr>
<tr>
<td></td>
<td>Hypotension</td>
</tr>
<tr>
<td></td>
<td>Renal impairment</td>
</tr>
<tr>
<td>Insulin</td>
<td>Hypoglycaemia</td>
</tr>
<tr>
<td>NSAIDs</td>
<td>(e.g. ibuprofen) Gastrotoxicity (dyspepsia, gastrointestinal bleeding)</td>
</tr>
<tr>
<td></td>
<td>Renal impairment</td>
</tr>
<tr>
<td>Opioid analgesics</td>
<td>(e.g. morphine) Nausea and vomiting</td>
</tr>
<tr>
<td></td>
<td>Confusion</td>
</tr>
<tr>
<td></td>
<td>Constipation</td>
</tr>
</tbody>
</table>

(ACE = angiotensin-converting enzyme; NSAID = non-steroidal anti-inflammatory drug)

Examples include constipation caused by opioids, hypotension caused by antihypertensives and dehydration caused by diuretics.

‘hyper-susceptible’ because of unpredictable immunological or genetic factors (e.g. anaphylaxis caused by penicillin, peripheral neuropathy caused by isoniazid in poor acetylators).

This simple classification has shortcomings and a more detailed classification based on dose (see Fig. 2.2, p. 19), timing and susceptibility (DoTS) is now used by those analysing ADRs in greater depth (Box 2.9). The AB classification can be extended as a reminder of some other types of ADR:

- Type C (‘chronic/continuous’) ADRs. These occur only after prolonged continuous exposure to a drug. Examples include osteoporosis caused by corticosteroids, retinopathy caused by chloroquine, and tardive dyskinesia caused by phenothiazines.
- Type D (‘delayed’) ADRs. These are delayed until long after drug exposure, making diagnosis difficult. Examples include malignancies that may emerge after immunosuppressive treatment post transplantation (e.g. azathioprine, tacrolimus) and vaginal cancer occurring many years after exposure to diethylstilboestrol.
Anaphylaxis

Hepatotoxicity

What

Does

Clear

See

Benzodiazepine

Have

Nausea

What

exploration of drug exposures (including self-medication should lead to the suspicion of an ADR and a careful drugs. Congenital defects in a live infant or aborted fetus teratogenicity and led to mandatory testing of all new mide disaster in the early 1960s highlighted the risk of the development of the fetus in the first 10 weeks of intrauterine life (e.g. phenytoin, warfarin). The thalido -the collection of population statistics. Many health-care systems routinely collect patient identifiable data on prescriptions (a surrogate marker of exposure to a drug), health-care events (e.g. hospitalisation, operations, new clinical diagnoses) and other clinical data (e.g. haematology, biochemistry). As these records are linked, with appropriate safeguards for confidentiality and data protection, they are providing a much more powerful mechanism for assessing both the harms and benefits of drugs.

A teratogen is a drug with the potential to affect the development of the fetus in the first 10 weeks of intrauterine life (e.g. phenytoin, warfarin). The thalidomide disaster in the early 1960s highlighted the risk of teratogenicity and led to mandatory testing of all new drugs. Congenital defects in a live infant or aborted fetus should lead to the suspicion of an ADR and a careful exploration of drug exposures (including self-medication and herbal remedies).

Detecting ADRs
– pharmacovigilance

Type A ADRs become apparent early in the development of a new drug. However, by the time a new drug is licensed and launched on to a possible worldwide market, hundreds rather than thousands of patients may have been exposed to it, so that rarer but potentially serious type B ADRs may remain undiscovered. Pharmacovigilance is the process of detecting (‘signal generation’) and evaluating ADRs in order to help prescribers and patients to be better informed about the risks of drug therapy. Drug regulatory agencies may respond to this information by placing restrictions on the licensed indications, reducing the recommended dose range, adding special warnings and precautions for prescribers in the product literature, writing to all health-care professionals, or withdrawing the product from the market.

Voluntary reporting systems allow health-care professionals and patients to report suspected ADRs to the regulatory authorities. A good example is the ‘Yellow Card’ scheme that was set up in the UK in response to the thalidomide tragedy. Reports are analysed to assess the likelihood that they represent a true ADR (Box 2.10).

Although voluntary reporting is a continuously operating and effective early warning system for previously unrecognised rare ADRs, its weaknesses include low reporting rates (only 3% of all ADRs and 10% of serious ADRs are ever reported), an inability to quantify risk (because the ratio of ADRs to prescriptions is unknown), and the influence of prescriber awareness on likelihood of reporting (reporting rates rise rapidly following publicity about potential ADRs).

More systematic approaches to collecting information on ADRs include ‘prescription event monitoring’, in which a sample of prescribers of a particular drug are issued with questionnaires concerning the clinical outcome for their patients, and the collection of population statistics. Many health-care systems routinely collect patient identifiable data on prescriptions (a surrogate marker of exposure to a drug), health-care events (e.g. hospitalisation, operations, new clinical diagnoses) and other clinical data (e.g. haematology, biochemistry). As these records are linked, with appropriate safeguards for confidentiality and data protection, they are providing a much more powerful mechanism for assessing both the harms and benefits of drugs.

All prescribers will inevitably see patients experiencing ADRs caused by prescriptions written by themselves or their colleagues. It is important that these are recognised early. In addition to the features in Box 2.10, features that should raise suspicion of an ADR and the need to respond (by drug withdrawal, dosage reduction or reporting to the regulatory authorities) include:

• concern expressed by a patient that a drug has harmed him/her
• abnormal clinical measurements (e.g. blood pressure, temperature, pulse, blood glucose and

DoT5 classification of ADRs

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dose</td>
<td>Anaphylaxis with penicillin</td>
</tr>
<tr>
<td>Below therapeutic dose</td>
<td>Nausea with morphine</td>
</tr>
<tr>
<td>In the therapeutic dose</td>
<td>Hepatotoxicity with paracetamol</td>
</tr>
<tr>
<td>range</td>
<td></td>
</tr>
<tr>
<td>At high doses</td>
<td></td>
</tr>
<tr>
<td>Timing</td>
<td>Anaphylaxis with penicillin</td>
</tr>
<tr>
<td>With the first dose</td>
<td>Hyponatraemia with diuretics</td>
</tr>
<tr>
<td>Early stages of treatment</td>
<td>Benzoazepine withdrawal syndrome</td>
</tr>
<tr>
<td>On stopping treatment</td>
<td>Clear cell cancer with diethylstilboestrol</td>
</tr>
<tr>
<td>Significantly delayed</td>
<td></td>
</tr>
<tr>
<td>Susceptibility</td>
<td>See patient factors in Box 2.7</td>
</tr>
</tbody>
</table>

(INR = international normalised ratio)

• Type E (‘end of treatment’) ADRs. These occur after abrupt drug withdrawal (see Box 2.3, p. 20).

TREND analysis of suspected ADRs

<table>
<thead>
<tr>
<th>Factor</th>
<th>Key question</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal relationship</td>
<td>What is the time interval between the start of drug therapy and the reaction?</td>
<td>Most ADRs occur soon after starting treatment and within hours in the case of anaphylactic reactions</td>
</tr>
<tr>
<td>Re-challenge</td>
<td>What happens when the patient is re-challenged with the drug?</td>
<td>Re-challenge is rarely possible because of the need to avoid exposing patients to unnecessary risk</td>
</tr>
<tr>
<td>Exclusion</td>
<td>Have concomitant drugs and other non-drug causes been excluded?</td>
<td>ADR is a diagnosis of exclusion following clinical assessment and relevant investigations for non-drug causes</td>
</tr>
<tr>
<td>Novelty</td>
<td>Has the reaction been reported before?</td>
<td>The suspected ADR may already be recognised and mentioned in the SPC approved by the regulatory authorities</td>
</tr>
<tr>
<td>De-challenge</td>
<td>Does the reaction improve when the drug is withdrawn or the dose is reduced?</td>
<td>Most, but not all, ADRs improve on drug withdrawal, although recovery may be slow</td>
</tr>
</tbody>
</table>

(SP5 = Summary of Product Characteristics)
weight) or laboratory results (e.g. abnormal liver or renal function, low haemoglobin white cell count) while on drug therapy

• new therapy started which could be in response to an ADR (e.g. omeprazole, allopurinol, naloxone)
• the presence of risk factors for ADRs (see Box 2.7).

### Drug interactions

A drug interaction has occurred when the administration of one drug increases or decreases the beneficial or adverse responses to another drug. Although the number of potential interacting drug combinations is very large, only a small number are common in clinical practice. Important drug interactions are most likely to occur when the affected drug has a low therapeutic index, steep dose-response curve, high first-pass or saturable metabolism, or single mechanism of elimination.

#### Mechanisms of drug interactions

Pharmacodynamic interactions occur when two drugs produce additive, synergistic or antagonistic effects at the same drug target (e.g. receptor, enzyme) or physiological system (e.g. electrolyte excretion, heart rate). These are the most common interactions in clinical practice and some important examples are given in Box 2.11.

Pharmacokinetic interactions occur when the administration of a second drug alters the concentration of the first at its site of action. There are numerous potential mechanisms:

- **Absorption interactions.** Drugs that either delay (e.g. anticholinergic drugs) or enhance (e.g. prokinetic drugs) gastric emptying influence the rate of rise in plasma concentration of other drugs but not the total amount of drug absorbed. Drugs that bind to form insoluble complexes or chelates (e.g. aluminium-containing antacids binding with ciprofloxacin) can reduce drug absorption.
- **Distribution interactions.** Co-administration of drugs that compete for protein binding in plasma (e.g. phenytoin and diazepam) can increase the unbound drug concentration, but the effect is usually short-lived due to increased elimination and hence restoration of the pre-interaction equilibrium.
- **Metabolism interactions.** Many drugs rely on metabolism by different isoenzymes of cytochrome P450 (CYP) in the liver. CYP enzyme inducers (e.g. phenytoin, rifampicin) generally reduce plasma concentrations of other drugs, although they may enhance activation of prodrugs. CYP enzyme inhibitors (e.g. clarithromycin, cimetidine, grapefruit

<table>
<thead>
<tr>
<th><strong>Mechanism</strong></th>
<th><strong>Object drug</strong></th>
<th><strong>Precipitant drug</strong></th>
<th><strong>Result</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pharmaceutical</strong>*</td>
<td>Sodium bicarbonate</td>
<td>Calcium gluconate</td>
<td>Precipitation of insoluble calcium carbonate</td>
</tr>
<tr>
<td><strong>Pharmacokinetic</strong></td>
<td>Tetracyclines</td>
<td>Calcium, aluminium, and magnesium salts</td>
<td>Reduced tetracycline absorption</td>
</tr>
<tr>
<td>Reduced protein binding</td>
<td>Phenytoin</td>
<td>Aspirin</td>
<td>Increased unbound and reduced total phenytoin plasma concentration</td>
</tr>
<tr>
<td>Reduced metabolism</td>
<td>Terfenadine</td>
<td>Grapefruit juice</td>
<td>Cardiac arrhythmias because of prolonged QT interval (p. 570)</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>Warfarin</td>
<td>Clarithromycin</td>
<td>Enhanced anticoagulation</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>Phenytoin</td>
<td>Omeprazole</td>
<td>Phenytoin toxicity</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>Clozapine</td>
<td>Paroxetine</td>
<td>Clozapine toxicity</td>
</tr>
<tr>
<td>Xanthine oxidase</td>
<td>Azathioprine</td>
<td>Allopurinol</td>
<td>Azathioprine toxicity</td>
</tr>
<tr>
<td>Monoamine oxidase inhibitors</td>
<td>Catecholamines</td>
<td></td>
<td>Hypertensive crisis due to monoamine toxicity</td>
</tr>
<tr>
<td>Increased metabolism (enzyme induction)</td>
<td>Ciclosporin</td>
<td>St John's wort</td>
<td>Loss of immunosuppression</td>
</tr>
<tr>
<td>Reduced renal elimination</td>
<td>Lithium</td>
<td>Diuretics</td>
<td>Lithium toxicity</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>NSAIDs</td>
<td>Methotrexate toxicity</td>
<td></td>
</tr>
<tr>
<td><strong>Pharmacodynamic</strong></td>
<td>Opiates</td>
<td>Naloxone</td>
<td>Reversal of opiate effects used therapeutically</td>
</tr>
<tr>
<td>Direct antagonism at same receptor</td>
<td>Salbutamol</td>
<td>Atenolol</td>
<td>Inhibits bronchodilator effect</td>
</tr>
<tr>
<td>Direct potentiation in same organ system</td>
<td>Benzodiazepines</td>
<td>Alcohol</td>
<td>Increased sedation</td>
</tr>
<tr>
<td>ACE inhibitors</td>
<td>NSAIDs</td>
<td>Increased risk of renal impairment</td>
<td></td>
</tr>
<tr>
<td>Indirect potentiation</td>
<td>Digoxin</td>
<td>Diuretics</td>
<td>Digoxin toxicity enhanced because of hypokalaemia</td>
</tr>
<tr>
<td>Warfarin</td>
<td>Aspirin, NSAIDs</td>
<td>Increased risk of bleeding because of gastrotoxicity and antiplatelet effects</td>
<td></td>
</tr>
</tbody>
</table>

*Pharmaceutical interactions are related to the formulation of the drugs and occur before drug absorption.*
juice) have the opposite effect. Enzyme induction effects usually take a few days to manifest because of the need to synthesise new CYP enzyme, in contrast with the rapid effects of enzyme inhibition.

- **Excretion interactions.** These primarily affect renal excretion. For example, drug-induced reduction in glomerular filtration rate (e.g. diuretic-induced dehydration, ACE inhibitors, NSAIDs) can reduce the clearance and increase the plasma concentration of many drugs, including some with a low therapeutic index (e.g. digoxin, lithium, aminoglycoside antibiotics). Less commonly, interactions may be due to competition for a common tubular organic anion transporter (e.g. methotrexate excretion may be inhibited by competition with NSAIDs).

### Avoiding drug interactions

Drug interactions are increasing as patients are prescribed more medicines (polypharmacy). Prescribers can avoid the adverse consequences of drug–drug interactions by taking a careful drug history (see Box 2.6) before prescribing additional drugs, only prescribing for clear indications, and taking special care when prescribing drugs with a narrow therapeutic index (e.g. warfarin). When prescribing an interacting drug is unavoidable, good prescribers will seek further information and anticipate the potential risk. This will allow them to provide special warnings for the patient and arrange for monitoring, either of the clinical effects (e.g. coagulation tests for warfarin) or of plasma concentration (e.g. digoxin).

### Medication errors

A medication error is any *preventable* event that may lead to inappropriate medication use or patient harm while the medication is in the control of the health-care professional or patient. Errors may occur in prescribing, dispensing, preparing solutions, administration or monitoring. Many ADRs are considered in retrospect to have been ‘avoidable’ with more care or forethought; in other words, an adverse event considered by one prescriber to be an unfortunate one prescriber to be an unfortunate event that may preventable event that may preventable event that may preventable event that may have occurred because of the need to synthesise new CYP enzyme, in contrast with the rapid effects of enzyme inhibition.

### Responding when an error is discovered

All prescribers will make errors. When they do, their first duty is to protect the patient’s safety. This will involve a clinical review and taking any steps that will reduce harm (e.g. remedial treatment, monitoring, recording the event in the notes, informing colleagues). Patients should be informed if they have been exposed to potential harm. For errors that do not reach the patient, it is the prescriber’s duty to report them, so that calculating doses adjusted for body weight, or planning appropriate monitoring after drug administration, and by health-care systems providing clinical pharmacist support (e.g. for checking the patient’s previous medications and current prescriptions) and electronic prescribing (which avoids errors due to illegibility or serious dosing mistakes, and may be combined with a clinical decision support system to take account of patient characteristics and drug history and provide warnings of potential contraindications and drug interactions).
2.13 Causes of prescribing errors

**Systems factors**
- Working hours of prescribers (and others)
- Patient throughput
- Professional support and supervision by colleagues
- Availability of information (medical records)
- Design of prescription forms
- Distractions
- Decision support available
- Checking routines (e.g. clinical pharmacy)
- Reporting and reviewing of incidents

**Prescriber factors**

Knowledge
- Clinical pharmacology principles
- Drugs in common use
- Therapeutic problems commonly encountered
- Knowledge of workplace systems

Skills
- Taking a good drug history
- Obtaining information to support prescribing
- Communicating with patients
- Numeracy and calculations
- Prescription writing

Attitudes
- Coping with risk and uncertainty
- Monitoring of prescribing

---

Fig. 2.5 Human error theory. Unintended errors may occur because the prescriber fails to complete the prescription correctly (a slip) (e.g. writes the dose in ‘mg’ not ‘micrograms’) or forgets part of the action that is important for success (a lapse) (e.g. forgets to co-prescribe folic acid with methotrexate); prevention requires the system to provide appropriate checking routines. Intended errors occur when the prescriber acts incorrectly due to lack of knowledge (a mistake) (e.g. prescribes atenolol for a patient with known severe asthma because of ignorance about the contraindication); prevention must focus on training the prescriber.

others can learn from the experience, and take the opportunity to reflect on how a similar incident might be avoided in the future.

---

2.14 Clinical development of new drugs

**Phase I**
- Healthy volunteers (20–80). These involve initial single-dose, ‘first-into-man’ studies, followed by repeated-dose studies. They aim to establish the basic pharmacokinetic and pharmacodynamic properties, and short-term safety.
  - Duration: 6–12 months

**Phase II**
- Patients (100–200). These investigate clinical effectiveness (‘proof of concept’), safety and dose–response relationship, often with a surrogate clinical endpoint, in the target patient group to determine the optimal dosing regimen for larger confirmatory studies.
  - Duration: 1–2 years

**Phase III**
- Patients (100s–1000s). These are large, expensive clinical trials that confirm safety and efficacy in the target patient population, using relevant clinical endpoints. They may be placebo-controlled studies or comparisons with other active compounds.
  - Duration: 1–2 years

**Phase IV**
- Patients (100s–1000s). These are undertaken after the medicine has been marketed for its first indication to evaluate new indications, new doses or formulations, long-term safety or cost-effectiveness.

---

DRUG REGULATION AND MANAGEMENT

Given the powerful beneficial and potentially adverse effects of drugs, the production and use of medicines are strictly regulated (e.g. by the Food and Drug Administration in the United States (US), Medicines and Healthcare Products Regulatory Agency in the UK, and Central Drugs Standard Control Organisation in India). Regulators are responsible for licensing medicines, monitoring their safety (pharmacovigilance, p. 27), approving clinical trials, and inspecting and maintaining standards of drug development and manufacture.

In addition, because of the high costs of drugs and of their adverse effects, health-care services must prioritise their use in light of the evidence of their benefit and harm, a process referred to as ‘medicines management’.

Drug development and marketing

Naturally occurring products have been used to treat illnesses for thousands of years and some remain in common use today. Examples include morphine from the opium poppy (*Papaver somniferum*), digitalis from the foxglove (*Digitalis purpurea*), curare from the bark of a South American tree, and quinine from another bark (*Cinchona* species). Although plants and animals remain a source of discovery, the majority of new drugs come from drug discovery programmes that aim to identify...
small-molecule compounds with specific interactions with a molecular target that will induce a predicted biological effect.

The usual pathway for development of these small molecules includes: identifying a plausible molecular target by investigating pathways in disease; screening a large library of compounds for those that interact with the molecular target in vitro; conducting extensive medicinal chemistry to optimise the properties of lead compounds; testing efficacy and toxicity of these compounds in vitro and in animals; and undertaking a clinical development programme (Box 2.14). This process typically takes longer than 10 years and may cost up to $1 billion. Manufacturers have a defined period of exclusive marketing of the drug while it remains protected by an original patent, typically 10–15 years, during which time they must recoup the costs of developing the drug. Meanwhile, competitor companies will often produce similar ‘me too’ drugs of the same class. Once the drug’s patent has expired, ‘generic’ drug manufacturers may step in to produce cheaper formulations.

The number of new drugs produced by the pharmaceutical industry has declined in recent years. The traditional approach of targeting membrane-bound receptors and enzymes with small molecules (see Box 2.2) is now giving way to new targets such as complex second messenger systems, cytokines, nucleic acids and cellular networks. These require novel therapeutic agents which present new challenges for ‘translational medicine’, the discipline of converting scientific discoveries into a useful medicine with a well-defined benefit-risk profile (Box 2.15).

**Licensing new medicines**

New drugs are given a ‘market authorisation’ based on the evidence of quality, safety and efficacy presented by the manufacturer. The regulator will not only approve the drug but will also take great care to ensure that the accompanying information reflects the evidence that has been presented. The summary of product characteristics (SPC), or ‘label’, provides detailed information about indications, dosage, adverse effects, warnings, monitoring etc. If approved, drugs can be made available with different levels of restriction:

- **Controlled drug (CD).** These drugs are subject to strict legal controls on supply and possession, usually due to their abuse potential (e.g. opioid analgesics).
- **Prescription-only medicine (PoM).** These are only available from a pharmacist and can only be supplied if prescribed by an appropriate practitioner.
- **Pharmacy (P).** These are only available from a pharmacist but can be supplied without a prescription.
- **General sales list (GSL).** These medicines may be bought ‘over the counter’ (OTC) from any shop and without a prescription.

**Drug marketing**

The marketing activities of the pharmaceutical industry are well resourced and are important in the process of recouping the massive costs of drug development. In some countries, such as the US, it is possible to promote a new drug by direct-to-consumer advertising, although this is illegal in the countries of the European Union. A major focus is on promotion to prescribers via educational events, sponsorship of meetings, adverts in journals, involvement with opinion leaders, and direct contact by company representatives. Such largesse has the potential to cause significant conflicts of interest, and might tempt prescribers to favour one drug over another, even in the face of evidence on effectiveness or cost-effectiveness.

**Managing the use of medicines**

Many medicines meet the three key regulatory requirements of quality, safety and efficacy. Although prescribers are legally entitled to prescribe any of them, it is desirable to limit the choice so that treatments for specific diseases can be focused on the most effective and cost-effective options, prescribers (and patients) gain familiarity with a smaller number of medicines, and pharmacies can concentrate stocks on them.

The process of ensuring optimal use of available medicines is known as ‘medicines management’ or ‘quality use of medicines’. It involves careful evaluation of the evidence of benefit and harm from using the medicine, an assessment of cost-effectiveness, and support for processes to implement the resulting recommendations.
These activities usually involve both national and local organisations.

**Evaluating evidence**

The principles of evidence-based medicine are described on page 8. Drugs are often evaluated in high-quality randomised controlled trials, the results of which can be considered by systematic review (Fig. 2.6). Ideally, data are available not only for comparison with placebo but also in ‘head-to-head’ comparison with alternative therapies. However, trials are conducted in selected patient populations and are not representative of every clinical scenario, so extrapolation to individual patients is not always straightforward.

Clinical trials typically report the percentage reduction in risk of a primary outcome, such as a stroke. However, if the absolute risk of stroke for a given patient is low, then even an apparently substantial percentage reduction in risk may not be worthwhile. As an aid to interpreting the results of clinical trials, it is often helpful to consider the number needed to treat (NNT).

For example, in a systematic review of warfarin therapy compared with placebo, it was shown that the absolute risk of stroke were halved, the relative risk reduction was 0.40, i.e. a 60% reduction. It equates to an NNT of 18, i.e. 18 patients would need to be treated for 1 year to prevent 1 stroke; NNT is calculated as the inverse of the difference in absolute rate of stroke = 1/[(133/1450) − (53/1450)]. However, if the absolute risk of stroke were halved, the relative risk reduction with warfarin would still be 60%, but the NNT would increase to 1/[(67/1450) − (27/1450)] = 36, i.e. 36 patients would need to be treated for 1 year to prevent 1 stroke, with a consequent increase in cost and risk of adverse effects compared with benefit. NNT can usefully be applied to assess both benefit (NNTB) and harm (NNTH).

**Evaluating cost-effectiveness**

New drugs often represent an incremental improvement over the current standard of care but are usually more expensive. The principles for evaluating and comparing cost-effectiveness of treatment are described on page 9. A major challenge is to compare the value of interventions for different clinical outcomes. One method is to calculate the quality-adjusted life years (QALYs) gained if the new drug is used rather than standard treatment. This analysis involves estimating the ‘utility’ of various health states between 1 (perfect health) and 0 (dead). If the additional costs and any savings are known, then it is possible to derive the incremental cost-effectiveness ratio (ICER) in terms of cost/QALY. These principles are exemplified in Box 2.16. However, there are inherent weaknesses in this kind of analysis: it usually depends on modelling future outcomes well beyond the duration of the clinical trial data that are available; it assumes that QALYs gained at all ages are of equivalent value; and the appropriate standard care against which the new drug should be compared is often uncertain.

These pharmacoeconomic assessments are challenging and resource-intensive, and are undertaken at national level in most countries: for example in the UK by the National Institute for Health and Clinical Excellence (NICE).

**Implementing recommendations**

Many recommendations about drug therapy are included in clinical guidelines written by an expert group after systematic review of the evidence. As described on page 8, these provide recommendations rather than obligations for prescribers, and are helpful in promoting more consistent and higher-quality care.
prescribing. However, guidelines are often written without concern for cost-effectiveness, and may be limited by the quality of available evidence. Guidelines cannot anticipate the extent of the variation between individual patients who may, for example, have unexpected contraindications to recommended drugs or choose different priorities for treatment. However, when deviating from respected national guidance, prescribers should be able to justify their practice.

Additional recommendations for prescribing are often implemented locally or imposed by bodies responsible for paying for health care. Most health-care units have a drug and therapies committee (or equivalent) comprised of senior and junior medical staff, pharmacists and nurses, as well as managers (because of the implications of the committee’s work for governance and resources). This group typically develops local prescribing policy and guidelines, maintains a local drug formulary and evaluates requests to use new drugs. The local formulary contains a more limited list than any national formulary (e.g. British National Formulary) because the latter lists all licensed medicines that can be prescribed legally, while the former contains only those which the health-care organisation has approved for local use. The local committee may also be involved, with local specialists, in providing explicit protocols for management of clinical scenarios (p. 9).

**PRESCRIBING IN PRACTICE**

**Decision-making in prescribing**

Prescribing should be based on a rational approach to a series of challenges (see Box 2.1, p. 18).

**Making a diagnosis**

Ideally, prescribing should be based on a confirmed diagnosis but, in reality, many prescriptions are based on the balance of probability, taking into account the differential diagnosis (e.g. proton pump inhibitors for post-prandial retrosternal discomfort).

**Establishing the therapeutic goal**

The goals of treatment are usually clear, particularly when relieving symptoms (e.g. pain, nausea, constipation). Sometimes the goal is less obvious to the patient, especially when aiming to prevent future events (e.g. ACE inhibitors to prevent hospitalisation and extend life in chronic heart failure). Prescribers should be clear about the therapeutic goal against which they will judge success or failure of treatment. It is also important to establish that the value placed on this goal by the prescriber is shared by the patient.

**Choosing the therapeutic approach**

For many clinical problems, drug therapy is not absolutely mandated. Having taken the potential benefits and harms into account, prescribers must consider whether drug therapy is in the patient’s interest and is preferred to no treatment or one of a range of alternatives (e.g. physiotherapy, psychotherapy, surgery). Assessing the balance of benefit and harm is often complicated and depends on various features associated with the patient, disease and drug (Box 2.17).

---

### 2.17 Factors to consider when balancing benefits and harms of drug therapy

- Seriousness of the disease or symptom
- Efficacy of the drug
- Seriousness of potential adverse effects
- Likelihood of adverse effects
- Efficacy of alternative drugs or non-drug therapies
- Safety of alternative drugs or non-drug therapies

**Choosing a drug**

For most common clinical indications (e.g. type 2 diabetes, depression), more than one drug is available, often from more than one drug class. Although prescribers often have guidance about which represents the rational choice for the average patient, they still need to consider whether this is the optimal choice for the individual patient. Certain factors may influence the choice of drug: Absorption. Patients may find some formulations easier to swallow than others or may be vomiting and require a drug available for parenteral administration. Distribution. Distribution of a drug to a particular tissue sometimes dictates choice (e.g. tetracyclines and rifampicin are concentrated in the bile, and lincomycin and clindamycin in bones). Metabolism. Drugs that are extensively metabolised should be avoided in severe liver disease (e.g. opioid analgesics). Excretion. Drugs that depend on renal excretion for elimination (e.g. digoxin, aminoglycoside antibiotics) should be avoided in patients with impaired renal function if suitable alternatives exist. Efficacy. Prescribers normally choose drugs with the greatest efficacy in achieving the goals of therapy (e.g. proton pump inhibitors rather than histamine, receptor antagonists). However, it may be appropriate to compromise on efficacy if other drugs are more convenient, safer to use or less expensive. Avoiding adverse effects. Prescribers should be wary of choosing drugs that are more likely to cause adverse effects (e.g. cephalosporins rather than alternatives for patients allergic to penicillin) or worsen coexisting conditions (e.g. β-blockers as treatment for angina in patients with asthma). Features of the disease. This is most obvious when choosing antibiotic therapy, which should be based on the known or suspected sensitivity of the infective organism (p. 149). Severity of disease. The choice of drug should be appropriate to disease severity (e.g. paracetamol for mild pain, morphine for severe pain). Coexisting diseases may be either an indication or a contraindication to therapy. Hypertensive patients might be prescribed a β-blocker if they also have left ventricular impairment but not if they have asthma. Avoiding adverse drug interactions. Prescribers should avoid giving combinations of drugs that might interact, either directly or indirectly (see Box 2.11). Patient adherence to therapy. Prescribers should choose drugs with a simple dosing schedule or easier administration (e.g. the ACE inhibitor enalapril once daily rather than captopril 3 times daily for hypertension). Cost. Prescribers should choose the cheaper drug if two drugs are of equal efficacy and safety. Even if cost is not
a concern for the individual patient, it is important to remember that unnecessary expenditure will ultimately limit choices for other prescribers and patients. Sometimes a more costly drug may be appropriate (e.g. if it yields improved adherence).

Genetic factors. There are already a small number of examples where genotype influences the choice of drug therapy (see Box 2.5).

Choosing a dosage regimen

Prescribers have to choose a dose, route and frequency of administration (dosage regimen) to achieve a steady-state drug concentration that provides sufficient exposure of the target tissue without producing toxic effects. Manufacturers draw up dosage recommendations based on average observations in many patients but the optimal regimen that will maximise the benefit/harm ratio for an individual patient is never certain. Rational prescribing involves treating each prescription as an experiment and gathering sufficient information to amend it if necessary. There are some general principles that should be followed:

Dose titration. Prescribers should generally start with a low dose and titrate this slowly upwards as necessary. This cautious approach is particularly important if the patient is likely to be more sensitive to adverse pharmacodynamic effects (e.g. confusion or postural hypotension in the elderly) or have altered pharmacokinetic handling (e.g. renal or hepatic impairment), and when using drugs with a low therapeutic index (e.g. benzodiazepines, lithium, digoxin). However, there are some exceptions. Some drugs must achieve therapeutic concentration quickly because of the clinical circumstance (e.g. antibiotics, glucocorticoids, carbimazole). When early effect is important but there may be a delay in achieving steady state because of a drug’s long half-life (e.g. digoxin, warfarin, amiodarone), an initial loading dose is given prior to establishing the appropriate maintenance dose (see Fig. 2.4, p. 23).

If adverse effects occur, the dose should be reduced or an alternative drug prescribed; in some cases, a lower dose may suffice if it can be combined with another synergistic drug (e.g. the immunosuppressant azathioprine reduces glucocorticoid requirements in patients with inflammatory disease). It is important to remember that the shape of the dose–response curve (see Fig. 2.2, p. 19) means that higher doses may produce little added therapeutic effect and might increase the chances of toxicity.

Route. There are many reasons for choosing a particular route of administration (Box 2.18).

Frequency. Frequency of doses is usually dictated by a manufacturer’s recommendation. Less frequent doses are more convenient for patients but result in greater fluctuation between peaks and troughs in drug concentration (see Fig. 2.4, p. 23). This is relevant if the peaks are associated with adverse effects (e.g. dizziness with antihypertensives) or the troughs are associated with troublesome loss of effect (e.g. anti-Parkinsonian drugs). These problems can be tackled either by splitting the dose or by employing a modified-release formulation, if available.

Timing. For many drugs the time of administration is unimportant. However, there are occasionally pharmacokinetic or therapeutic reasons for giving drugs at particular times (Box 2.19).

Formulation. For some drugs there is a choice of formulation, some for use by different routes. Some are easier to ingest, particularly by children (e.g. elixirs). The formulation is important when writing repeat prescriptions for drugs with a low therapeutic index that come in different formulations (e.g. lithium, phenytoin, theophylline). Even if the prescribed dose remains constant, another formulation may differ in its absorption and bioavailability, and hence plasma drug concentration. These are examples of the small number of drugs that should be prescribed by specific brand name rather than ‘generic’ International Non-proprietary Name (INN).

Duration. Some drugs require a single dose (e.g. thrombolysis post myocardial infarction), while for others the duration of the course of treatment is certain at the outset (e.g. antibiotics). For most, the duration will be largely at the prescriber’s discretion and will depend on response and disease progression (e.g. analgesics, antidepressants). For many, the treatment will be long-term (e.g. insulin, antihypertensives, levothyroxine).

Involving the patient

Patients should, whenever possible, be engaged in making choices about drug therapy. Their beliefs and
expectations affect the goals of therapy and help in judging the acceptable benefit/harm balance when selecting treatments. Very often, patients may wish to defer to the professional expertise of the prescriber. Nevertheless, they play key roles in adherence to therapy and in monitoring treatment, not least by providing early warning of adverse events. It is important that they are provided with the necessary information to understand the choice that has been made, what to expect from the treatment, and any measurements that must be undertaken (Box 2.20).

A major drive to include patients has been the recognition that up to half of the drug doses for chronic preventative therapy are not taken. This is often termed ‘non-compliance’ but is more appropriately called ‘non-adherence’, to reflect a less paternalistic view of the doctor–patient relationship; it may or may not be intentional. Non-adherence to the dose regimen reduces the likelihood of benefits to the patient and can be costly in terms of wasted medicines and unnecessary healthcare episodes. An important reason may be a failure of concordance with the prescriber about the goals of treatment. A more open and shared decision-making process might resolve any misunderstandings at the outset and foster improved adherence as well as improved satisfaction with health-care services and confidence in prescribers.

**Writing the prescription**

The culmination of the planning described above is writing an accurate and legible prescription so that the drug will be dispensed and administered as planned (see ‘Writing prescriptions’ below).

---

**2.19 Factors influencing the timing of drug therapy**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Recommended timing</th>
<th>Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diuretics (e.g. furosemide)</td>
<td>Once in the morning</td>
<td>Night-time diuresis undesirable</td>
</tr>
<tr>
<td>Statins (e.g. simvastatin)</td>
<td>Once at night</td>
<td>HMG CoA reductase activity is greater at night</td>
</tr>
<tr>
<td>Antidepressants (e.g. amitriptyline)</td>
<td>Once at night</td>
<td>Allows adverse effects to occur during sleep</td>
</tr>
<tr>
<td>Salbutamol</td>
<td>Before exercise</td>
<td>Reduces symptoms in exercise-induced asthma</td>
</tr>
<tr>
<td>Glyceryl trinitrate</td>
<td>When required</td>
<td>Relief of acute symptoms only</td>
</tr>
<tr>
<td>Paracetamol</td>
<td>Eccentric dosing regimen (e.g. twice daily at 8 a.m. and 2 p.m.)</td>
<td>Reduces development of nitrate tolerance by allowing drug-free period each night</td>
</tr>
<tr>
<td>Aspirin</td>
<td>With food</td>
<td>Minimises gastrotoxic effects</td>
</tr>
<tr>
<td>Alendronate</td>
<td>Once in the morning before breakfast, sitting upright</td>
<td>Minimises risk of oesophageal irritation</td>
</tr>
<tr>
<td>Tetracyclines</td>
<td>2 hours before or after food or antacids</td>
<td>Divalent and trivalent cations chelate tetracyclines, preventing absorption</td>
</tr>
<tr>
<td>Hypnotics (e.g. temazepam)</td>
<td>Once at night</td>
<td>Maximises therapeutic effect and minimises daytime sedation</td>
</tr>
<tr>
<td>Antihypertensive drugs (e.g. amlodipine)</td>
<td>Once in the morning</td>
<td>Blood pressure is higher during the daytime</td>
</tr>
</tbody>
</table>

(HMG CoA = 3-hydroxy-3-methylglutaryl-coenzyme A)

---

**2.20 What patients need to know about their medicines**

<table>
<thead>
<tr>
<th>Knowledge</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The reason for taking the medicine</td>
<td>Reinforces the goals of therapy</td>
</tr>
<tr>
<td>How the medicine works</td>
<td>May be important for the effectiveness (e.g. inhaled salbutamol in asthma) and safety (e.g. alendronate for osteoporosis) of treatment</td>
</tr>
<tr>
<td>How to take the medicine</td>
<td>May help to support adherence or prompt review because of treatment failure</td>
</tr>
<tr>
<td>What benefits to expect</td>
<td>Discuss common and mild effects that may be transient and might not require discontinuation</td>
</tr>
<tr>
<td>What adverse effects might occur</td>
<td>Mention rare but serious effects that might influence the patient’s consent</td>
</tr>
<tr>
<td>Precautions that improve safety</td>
<td>Explain symptoms to report that might allow serious adverse effects to be averted, monitoring that will be required and potentially important drug–drug interactions</td>
</tr>
<tr>
<td>When to return for review</td>
<td>This will be important to enable monitoring</td>
</tr>
</tbody>
</table>

*Many medicines are provided with patient information leaflets, which the patient should be encouraged to read.*
**Monitoring treatment effects**

Rational prescribing involves monitoring for the beneficial and adverse effects of treatment so that the balance remains in favour of a positive outcome (see ‘Monitoring drug therapy’ below).

**Stopping drug therapy**

It is also important to review long-term treatment at regular intervals to assess whether continued treatment is required. Elderly patients are keen to reduce their medication burden and are often prepared to compromise on the original goals of long-term preventative therapy to achieve this.

### Prescribing in special circumstances

### Prescribing for patients with renal disease

Patients with renal impairment are readily identified by a low estimated glomerular filtration rate (eGFR < 60 mL/min) based on their serum creatinine, age, sex and ethnic group (p. 466). This group includes a large proportion of elderly patients. If a drug (or its active metabolites) is eliminated predominantly by the kidneys, it will tend to accumulate and so the maintenance dose must be reduced. For some drugs, renal impairment makes patients more sensitive to their adverse pharmacodynamic effects. Examples of drugs that require extra caution in patients with renal disease are listed in Box 2.21.

### Prescribing for patients with hepatic disease

The liver has a large capacity for drug metabolism and hepatic insufficiency has to be advanced before drug dosages need to be modified. Patients who may have impaired metabolism include those with jaundice, ascites, hypoalbuminaemia, malnutrition or encephalopathy. Hepatic drug clearance may also be reduced in acute hepatitis, in hepatic congestion due to cardiac failure, and if there is intrahepatic arteriovenous shunting (for example, in hepatic cirrhosis). There are no good tests of hepatic drug-metabolising capacity or of biliary excretion, so dosage should be guided by the therapeutic response and careful monitoring for adverse effects. The presence of liver disease also increases the susceptibility to adverse pharmacological effects of drugs. Some drugs that require extra caution in patients with hepatic disease are listed in Box 2.21.

### Prescribing for elderly patients

See Box 2.22.

#### 2.21 Some drugs that require extra caution in patients with renal or hepatic disease

<table>
<thead>
<tr>
<th>Kidney disease</th>
<th>Liver disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitors and ARBs (renal impairment, hyperkalaemia)</td>
<td>Warfarin (increased anticoagulation because of reduced clotting factor)</td>
</tr>
<tr>
<td>Metformin (lactic acidosis)</td>
<td>Metformin (lactic acidosis)</td>
</tr>
<tr>
<td>Spironolactone (hyperkalaemia)</td>
<td>NSAIDs (gastrointestinal bleeding, fluid retention)</td>
</tr>
<tr>
<td>NSAIDs (impaired renal function)</td>
<td>NSAIDs (gastrointestinal bleeding, fluid retention)</td>
</tr>
<tr>
<td>Sulphonylureas (hypoglycaemia)</td>
<td>Sulphonylureas (hypoglycaemia)</td>
</tr>
<tr>
<td>Insulin (hypoglycaemia)</td>
<td>Benzodiazepines (coma)</td>
</tr>
</tbody>
</table>

#### Pharmacodynamic effects enhanced

- Digoxin: increased sensitivity to Na+/K+ pump; hypokalaemia due to diuretics; renal impairment favours accumulation → increased risk of toxicity
- Antihypertensive drugs: reduced baroreceptor function → increased risk of postural hypotension
- Antidepressants, hypnotics, sedatives, tranquillisers: increased sensitivity of the brain; reduced metabolism → increased risk of toxicity
- Warfarin: increased tendency to falls and injury and to bleeding from intra- and extracranial sites; increased sensitivity to inhibition of clotting factor synthesis → increased risk of bleeding
- Clomethiazole, lidocaine, nifedipine, phenobarbital, propranolol, theophylline: metabolism reduced → increased risk of toxicity
- NSAIDs: poor renal function → increased risk of renal impairment; susceptibility to gastrotoxicity → increased risk of upper gastrointestinal bleeding.

### Prescribing for women who are pregnant or breastfeeding

Prescribing in pregnancy should be avoided if possible to minimise the risk of adverse effects in the fetus.
While drug therapy in pregnancy may be required either for a pre-existing problem (e.g. epilepsy, asthma, hypothyroidism) or for problems that arise during pregnancy (e.g. morning sickness, anaemia, prevention of neural tube defects, gestational diabetes, hypertension). About 35% of women take drug therapy at least once during pregnancy and 6% take drug therapy during the first trimester (excluding iron, folic acid and vitamins). The most commonly used drugs are simple analgesics, antibacterial drugs and antacids. Some considerations when prescribing in pregnancy are listed in Box 2.23.

Drugs that are excreted in breast milk may cause adverse effects in the baby. Prescribers should always consult the summary of product characteristics for each drug or a reliable formulary when treating a pregnant woman or breastfeeding mother.

Writing prescriptions

A prescription is a means by which a prescriber communicates the intended plan of treatment to the pharmacist who dispenses a medicine and to a nurse or patient who administers it. It should be precise, accurate, clear and legible. The two main kinds of prescription are those written, dispensed and administered in hospital and those written in primary care (in the UK by a GP), dispensed at a community pharmacy and self-administered by the patient. The information supplied must include:

- the date
- identification of the patient
- the name of the drug
- the formulation
- the dose
- the frequency of administration
- the route and method of administration
- the amount to be supplied (primary care only)
- instructions for labelling (primary care only)
- the prescriber’s signature.

Prescribing in hospital

Although GP prescribing is increasingly electronic, most hospital prescribing continues to be based around the prescription and administration record (the ‘drug chart’). A variety of charts is in use and prescribers must familiarise themselves with the local version. Most contain the following sections:

- Basic patient information. Will usually include name, age, date of birth, hospital number and address. These details are often ‘filled in’ using a sticky addressograph label, but this increases the risk of serious error.
- Previous adverse reactions/allergies. Communicates important patient safety information based on a careful drug history and/or the medical record.
- Other medicines charts. Notes any other hospital prescription documents that contain current prescriptions being received by the patient (e.g. anticoagulants, insulin, oxygen, fluids).
- Once-only medications. For prescribing medicines to be used infrequently, such as single-dose prophylactic antibiotics and other pre-operative medications.
- Regular medications. For prescribing medicines to be taken for a number of days or continuously, such as a course of antibiotics, antihypertensive drugs etc.
- ‘As required’ medications. For prescribing for symptomatic relief, usually to be administered at the discretion of the nursing staff (e.g. antiemetics, analgesics).

Prescribers should be aware of the risks of prescription error (Box 2.24 and Box 2.13, p. 30), ensure they have considered the rational basis for their prescribing decision described above, and then follow the rules in Box 2.25 in order to write the prescription. It is a basic principle that a prescription will be followed by a judgement as to its success or failure and appropriate changes made, often by discontinuing one prescription and writing another.

Hospital discharge (‘to take out’) medicines

Most patients will be prescribed a short course of their medicines at discharge. This prescription is particularly important because it usually informs future therapy at the point of transfer of prescribing responsibility to primary care. Great care is required to ensure that this list is accurate and that any hospital medicines to be stopped are not included or are identified as of specified short duration. It is also important that any significant ADRs experienced in hospital are recorded and that any specific monitoring or review is identified.

Prescribing in primary care

Most of the considerations above are equally applicable to primary care (GP) prescriptions. In many health-care systems, community prescribing is electronic, making issues of legibility irrelevant and often providing basic decision support to limit the range of doses that can be written and highlight potential drug interactions.
2.24 High-risk prescribing moments

- Trying to amend an active prescription (e.g. altering the dose/timing) — *always avoid and start again*
- Writing up drugs in the immediate presence of more than one prescription chart or set of notes — *avoid*
- Allowing one's attention to be diverted in the middle of completing a prescription — *avoid*
- Prescribing ‘high risk’ drugs (e.g. anticoagulants, opioids, insulin, sedatives) — *ask for help if necessary*
- Prescribing parenteral drugs — *take care*
- Rushing prescribing (e.g. in the midst of a busy ward round) — *avoid*
- Prescribing unfamiliar drugs — *consult the formulary and ask for help if necessary*
- Transcribing multiple prescriptions from an expired chart to a new one — *take care*
- Writing prescriptions based on information from another source such as a referral letter (the list may contain errors and some of the medicines may be the cause of the patient’s illness) — *review the justification for each as if it is a new prescription*
- Writing up ‘to take out’ drugs (because these will become the patient’s regular medication for the immediate future) — *take care and seek advice if necessary*
- Calculating drug doses — *ask a colleague to perform an independent calculation*
- Prescribing sound-alike or look-alike drugs (e.g. chlorphenamine and chlorpromazine) — *take care*

2.25 How to write a drug prescription

- Write in block capitals, legibly, with black ballpoint pen. Do not amend what is already written: if a mistake is made, then start again
- Ensure there is clear and unambiguous labelling to *identify the patient*. Write the patient’s name, hospital number and date of birth (with age if under 12 years) on every sheet. The patient’s weight and height may be required to calculate safe doses for many drugs with narrow therapeutic indices
- Check the **drug sensitivities/allergies** box and obtain further details of the drug history if there are any doubts
- Use the generic **International Non-proprietary Name (INN)** rather than brand name (e.g. write ‘SIMVASTATIN’, not ‘ZOCOR®’). The only exceptions are when variation occurs in the properties of alternative branded formulations (e.g. modified-release preparations of drugs such as lithium, theophylline, phenytoin and nifedipine) or when the drug is a combination product with no generic name (e.g. Kilovance®). Do not use abbreviations, e.g. write ‘ISOSORBIDE MONONITRATE’ not ‘ISMN’
- Write the **drug dose**. The only acceptable abbreviations are ‘g’ and ‘mg’. ‘Units’ (e.g. of insulin or heparin) and ‘micrograms’ must always be written in full, never as ‘U’ or ‘µg’ (nor ‘mcg’, nor ‘ug’). Avoid decimal points (i.e. 500 mg not 0.5 g) or, if unavoidable, put a ‘0’ in front of it (e.g. ‘0.5 micrograms’ not ‘.5 micrograms’). Do not use a decimal point for round numbers (e.g. ‘7 mg’ not ‘7.0 mg’). For liquid preparations write the dose in mg; ‘mL’ can only be written for a combination product (e.g. Gaviscon® liquid) or if the strength is not expressed in weight (e.g. adrenaline (epinephrine) 1 in 1000). Use numbers/figures (e.g. 1 or ‘one’) to denote use of a sachet/enema but avoid prescribing numbers of tablets without specifying their strength. Always include the dose of inhaled drugs in addition to stating numbers of ‘puffs’, as strengths can vary. For some drugs a maximum dose may need to be stated (e.g. colchicine in gout)
- Write the **route and method of administration**. Widely accepted abbreviations are: intravenous – ‘IV’; intramuscular – ‘IM’; subcutaneous – ‘SC’; sublingual – ‘SL’; per rectum – ‘PR’; per vaginam – ‘PV’; nasogastric – ‘NG’, inhaled – ‘INH’; and topical – ‘TOP’. ‘ORAL’ is preferred to per oram – ‘PO’. Never abbreviate ‘INTRATHecal’. Care should be taken in specifying ‘RIGHT’ or ‘LEFT’ for eye and ear drops. It may be necessary to specify the method of giving a medicine intravenously (e.g. as a single undiluted bolus injection, or as an infusion in a specified volume of saline over a specified time)  
- Indicate the **frequency and timing of administration** clearly. For example: furosemide 40 mg once daily; amoxicillin 250 mg 3 times daily. Widely accepted Latin abbreviations for dose frequency are: once daily – ‘OD’; twice daily – ‘BD’; 3 times daily – ‘TDS’; 4 times daily – ‘QDS’, as required – ‘PRN’; in the morning – ‘OM’ (omni mane); at night – ‘ON’ (omni nocte); and immediately – ‘stat’. Alternatives are, for example, 6-hourly and 8-hourly, but these are less precise. The hospital chart usually requires specific times to be identified for regular medicines that coincide with nursing drug rounds. If treatment is for a known time period, cross off subsequent days when the medicine is not required. Similarly, if a drug is not to be given every day, cross off the days when it is not required. For ‘as required’ medicines describe the indication, frequency, minimal time interval between doses, and maximum dose in any 24-hour period
- Use the space provided for **added information**, e.g. whether a medicine should be taken with food, type of inhaler device used, and anything else that the drug dispenser should know. State here the times for peak/trough plasma levels for drugs requiring therapeutic monitoring
- **Sign and print your name** clearly so that you can be identified by colleagues. The prescription should be dated
- **Discontinue** a prescription by drawing a vertical line at the point of discontinuation, horizontal lines through the remaining days on the chart, and diagonal lines through the drug details and administration boxes. Sign and date this action and consider writing a supplementary note to explain it (e.g. describing any adverse effect)
Monitoring drug therapy

Prescribers should measure the effects of the drug, both beneficial and harmful, to inform decisions about dose titration (up or down), discontinuation or substitution of treatment. Monitoring can be achieved subjectively by asking the patient about symptoms or more objectively by measuring a clinical effect. Alternatively, if the pharmacodynamic effects of the drug are difficult to assess, then the plasma drug concentration may be measured on the basis that it will be closely related to the effect of the drug (see Fig. 2.2, p. 19).

**Clinical and surrogate endpoints**

Ideally, clinical endpoints are measured directly and the drug dosage titrated to achieve the therapeutic goal and avoid toxicity (e.g. control of ventricular rate in a patient with atrial fibrillation). Sometimes this is impractical because the clinical endpoint is a future event (e.g. prevention of myocardial infarction by statins or resolution of a chest infection with antibiotics); in these circumstances it may be possible to select a ‘surrogate’ endpoint that will predict success or failure. This may be an intermediate step in the pathophysiological process (e.g. serum cholesterol as a surrogate for risk of myocardial infarction) or a measurement which follows the pathophysiology even if it is not a key factor in its progression (e.g. serum C-reactive protein as a surrogate for resolution of inflammation in chest infection). Such surrogates are sometimes termed ‘biomarkers’.

**Plasma drug concentration**

The following criteria must be met to justify routine monitoring by plasma drug concentration:

- Clinical endpoints and other pharmacodynamic (surrogate) effects are difficult to monitor.
- The relationship between plasma concentration and clinical effects is predictable.
- The therapeutic index is low. For drugs with a high therapeutic index any variability in plasma concentrations is likely to be irrelevant clinically.

Some examples of drugs that fulfil these criteria are listed in Box 2.26.

Measurement of plasma concentration may be useful in planning adjustments of drug dose and frequency of administration; to explain an inadequate therapeutic

<table>
<thead>
<tr>
<th>Drug</th>
<th>Half-life (hrs)*</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digoxin</td>
<td>36</td>
<td>Steady state takes several days to achieve. Samples should be taken 6 hrs post dose. Measurement is useful to confirm the clinical impression of toxicity or non-adherence but clinical effectiveness is better assessed by ventricular heart rate. Risk of toxicity increases progressively at concentrations &gt; 1.5 µg/L, and is likely at concentrations &gt; 3.0 µg/L (toxicity is more likely in the presence of hypokalaemia)</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>2</td>
<td>Measure pre-dose trough concentration (should be &lt; 1 µg/mL) to ensure that accumulation and the risk of nephrotoxicity and ototoxicity is avoided; see p. 156</td>
</tr>
<tr>
<td>Levothyroxine</td>
<td>&gt; 120</td>
<td>Steady state may take up to 6 wks to achieve (p. 743)</td>
</tr>
<tr>
<td>Lithium</td>
<td>24</td>
<td>Steady state takes several days to achieve. Samples should be taken 12 hrs post dose. Target range 0.4–1 mmol/L</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>24</td>
<td>Measure pre-dose trough concentration (should be 10–20 mg/L) to ensure that accumulation is avoided. Good correlation between concentration and toxicity. Concentration may be misleading in the presence of hypoalbuminaemia</td>
</tr>
<tr>
<td>Theophylline (oral)</td>
<td>6</td>
<td>Steady state takes 2–3 days to achieve. Samples should be taken 6 hrs post dose. Target concentration is 10–20 mg/L but its relationship with bronchodilator effect and adverse effects is variable</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>6</td>
<td>Measure pre-dose trough concentration (should be 10–15 mg/L) to ensure clinical efficacy and that accumulation and the risk of nephrotoxicity is avoided</td>
</tr>
</tbody>
</table>

*Half-lives vary considerably with different formulations and between patients.
response (by identifying subtherapeutic concentration or incomplete adherence); to establish whether a suspected ADR is likely to be caused by the drug; and to assess and avoid potential drug interactions.

**Timing of samples in relation to doses**

The concentration of drug rises and falls during the dosage interval (see Fig. 2.4B, p. 23). Measurements made during the initial absorption and distribution phases are unpredictable because of the rapidly changing concentration, so samples are usually taken at the end of the dosage interval (a ‘trough’ or ‘pre-dose’ concentration). This measurement is normally made in steady state, which usually takes five half-lives to achieve after the drug is introduced or the dose changed (unless a loading dose has been given).

**Interpreting the result**

A target range is provided for many drugs, based on average thresholds for therapeutic benefit and toxicity. Inter-individual variability means that these can only be used as a guide. For instance, a patient who describes symptoms that could be consistent with toxicity but has a drug concentration in the top half of the target range should still be suspected of suffering toxic effects. Another important consideration is that some drugs are heavily protein-bound (e.g. phenytoin) but only the unbound drug is pharmacologically active. Therefore, patients with hypoalbuminaemia may have a therapeutic or even toxic concentration of unbound drug, despite a low ‘total’ concentration.

---

**Further information**

**Websites**

- [www.bnf.org](http://www.bnf.org) The British National Formulary (BNF) is a key reference resource for UK NHS prescribers, with a list of licensed drugs, chapters on prescribing in renal failure, liver disease, pregnancy and during breastfeeding, and appendices on drug interactions.
- [www.cochrane.org](http://www.cochrane.org) The Cochrane Collaboration is a leading international collaboration to provide evidence-based reviews (over 4000 so far).
- [www.evidence.nhs.uk](http://www.evidence.nhs.uk) NHS Evidence provides a wide range of health information relevant to delivering quality patient care.
- [www.icp.org.nz](http://www.icp.org.nz) The Interactive Clinical Pharmacology site is designed to increase understanding of principles in clinical pharmacology.
- [www.medicines.org/emc/](http://www.medicines.org/emc/) The electronic Medicines Compendium (eMC) contains up-to-date, easily accessible information about medicines licensed by the UK Medicines and Healthcare Products Regulatory Agency (MHRA) and the European Medicines Agency (EMA).
Molecular and genetic factors in disease

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional anatomy and physiology</td>
<td>42</td>
</tr>
<tr>
<td>Genetic disease and inheritance</td>
<td>50</td>
</tr>
<tr>
<td>- Meiosis</td>
<td>50</td>
</tr>
<tr>
<td>- Patterns of disease inheritance</td>
<td>51</td>
</tr>
<tr>
<td>- Classes of genetic variant</td>
<td>53</td>
</tr>
<tr>
<td>- Consequences of genetic variation</td>
<td>57</td>
</tr>
<tr>
<td>- Constitutional genetic disease</td>
<td>58</td>
</tr>
<tr>
<td>- Somatic genetic disease</td>
<td>59</td>
</tr>
<tr>
<td>Investigation of genetic disease</td>
<td>60</td>
</tr>
<tr>
<td>- General principles of diagnosis</td>
<td>60</td>
</tr>
<tr>
<td>- Genetic testing in pregnancy and pre-implantation genetic testing</td>
<td>62</td>
</tr>
<tr>
<td>- Genetic testing in children</td>
<td>63</td>
</tr>
<tr>
<td>- Identifying a disease gene in families</td>
<td>63</td>
</tr>
<tr>
<td>- Genetic investigation in populations</td>
<td>64</td>
</tr>
<tr>
<td>- Predictive genetic testing</td>
<td>64</td>
</tr>
<tr>
<td>Presenting problems in genetic disease</td>
<td>64</td>
</tr>
<tr>
<td>Research frontiers in molecular medicine</td>
<td>69</td>
</tr>
<tr>
<td>- Gene therapy</td>
<td>69</td>
</tr>
<tr>
<td>- Induced pluripotent stem cells and regenerative medicine</td>
<td>69</td>
</tr>
<tr>
<td>- Pathway medicine</td>
<td>70</td>
</tr>
<tr>
<td>Major categories of genetic disease</td>
<td>64</td>
</tr>
<tr>
<td>- Inborn errors of metabolism</td>
<td>64</td>
</tr>
<tr>
<td>- Neurological disorders</td>
<td>65</td>
</tr>
<tr>
<td>- Connective tissue disorders</td>
<td>65</td>
</tr>
<tr>
<td>- Learning disability, dysmorphism and malformations</td>
<td>66</td>
</tr>
<tr>
<td>- Familial cancer syndromes</td>
<td>66</td>
</tr>
<tr>
<td>Genetics of common diseases</td>
<td>68</td>
</tr>
<tr>
<td>- Measuring the genetic contribution to complex disease</td>
<td>68</td>
</tr>
<tr>
<td>- Genetic testing in complex disease</td>
<td>69</td>
</tr>
<tr>
<td>- Pharmacogenomics</td>
<td>69</td>
</tr>
</tbody>
</table>
Almost all diseases have a genetic component. In children and young adults in particular, many of the disorders causing long-term morbidity and mortality are genetically determined. The molecular basis of most Mendelian (or ‘single-gene’) diseases has now been determined, and our understanding of the abnormalities in cell function responsible for the clinical presentation is improving. It has also become clear that variants in many genes contribute to the pathogenesis of several common diseases such as asthma, rheumatoid arthritis and osteoporosis. In this chapter, we review key principles of cell biology, cellular signalling and molecular genetics, with emphasis on the diagnosis and assessment of patients with genetic diseases.

**FUNCTIONAL ANATOMY AND PHYSIOLOGY**

**Cell and molecular biology**

All human cell types are derived from a single totipotent stem cell, the zygote (the fertilised ovum). During development, organs and tissues are formed by the integration of four closely regulated cellular processes: cell division, migration, differentiation and programmed cell death. In many adult tissues such as skin, liver and the intestine, these processes continue throughout life, mediated by populations of stem cells that are responsible for tissue maintenance and repair. Cell biology is the study of these processes and of intracellular compartments, called organelles, which maintain cellular homeostasis. Dysfunction of any of these processes may lead to disease.

**DNA, chromosomes and chromatin**

The nucleus is a membrane-bound compartment found in all cells except erythrocytes and platelets. The human genome contains 46 chromosomes, each a single linear molecule of deoxyribonucleic acid (DNA) complexed with proteins to form chromatin. The basic protein unit of chromatin is the nucleosome, comprising 147 base pairs (bp) of DNA wound round a core of four different histone proteins. The vast majority of chromosomal DNA is double-stranded, with the exception of the ends of chromosomes, where ‘knotted’ domains of single-stranded DNA, called telomeres, are found. Telomeres prevent degradation and accidental fusion of chromosomal DNA.

The genome comprises approximately 3.1 billion bp of DNA. Humans are diploid organisms, meaning that each nucleus contains two copies of the genome, visible microscopically as 22 identical chromosomal pairs – the autosomes – named 1 to 22 in descending size order (see Fig. 3.11, p. 57), and two sex chromosomes (XX in females and XY in males). Each DNA strand consists of a linear sequence of four bases – guanine (G), cytosine (C), adenine (A) and thymine (T) – covalently linked by phosphate bonds. The sequence of one strand of double-stranded DNA determines the sequence of the opposite strand because the helix is held together by hydrogen bonds between adenine and thymine or guanine and cytosine nucleotides.

**Genes and transcription**

Genes are functional elements on the chromosome that are capable of transmitting information from the DNA template via the production of messenger ribonucleic acid (mRNA) to the production of proteins. The human genome contains an estimated 21,500 genes, although many of these are inactive or silenced in different cell types. For example, although the gene for parathyroid hormone (PTH) is present in every cell, activation of gene expression and production of PTH mRNA is virtually restricted to the parathyroid glands. Genes that are active in different cells undergo transcription, which requires binding of an enzyme called RNA polymerase II to a segment of DNA at the start of the gene termed the promoter. Once bound, RNA polymerase II moves along one strand of DNA, producing an RNA molecule that is complementary to the DNA template. A DNA sequence close to the end of the gene, called the polyadenylation signal, acts as a signal for termination of the RNA transcript (Fig. 3.1). The activity of RNA polymerase II is regulated by transcription factors. These proteins bind to specific DNA sequences at the promoter, or to enhancer elements that may be many thousands of base pairs away from the promoter. A loop in the chromosomal DNA brings the enhancer close to the promoter, enabling the bound proteins to interact.

The human genome encodes approximately 1200 different transcription factors, and mutations in many of these can cause genetic diseases (Fig. 3.2). Mutation of the transcription factor binding sites within promoters or enhancers also causes genetic disease. For example, the blood disorder alpha-thalassaemia can result from loss of an enhancer located more than 100,000 bp from the alpha-globin gene promoter, leading to greatly reduced transcription. Similarly, variation in the promoter of the gene encoding intestinal lactase determines whether or not this is ‘shut off’ in adulthood, producing lactose intolerance.

The accessibility of promoters to RNA polymerase II depends on the structural configuration of chromatin. Transcriptionally active regions have decondensed (or ‘open’) chromatin (euchromatin). Conversely, transcriptionally silent regions are associated with densely packed chromatin called heterochromatin. Chemical modification of both the DNA and core histone proteins allows heterochromatic regions to be distinguished from open chromatin. DNA can be modified by addition of a methyl group to cytosine molecules (methylation). In promoter regions, this silences transcription, since methyl cytosines are usually not available for transcription factor binding or RNA transcription. The core histones can also be modified via methylation, phosphorylation, acetylation or sumoylation at specific amino acid residues in a pattern that reflects the functional state of the chromatin; this is called the histone code – reflecting an emerging understanding of the ‘rules’ by which specific modifications mark transcriptionally activating (trimethylation of lysine 4 on histone H3; acetylation of many histone residues) or silencing (methylation of lysine 9 on histone H4; deacetylation of many histone residues) effects. Such DNA and protein modifications are termed epigenetic, as they do not alter the primary sequence of the DNA code but have biological significance in chromosomal function. Abnormal epigenetic changes are increasingly recognised as
**Functional anatomy and physiology**

**Fig. 3.1 RNA synthesis and its translation into protein.** Gene transcription involves binding of RNA polymerase II to the promoter of genes being transcribed with other proteins (transcription factors) that regulate the transcription rate. The primary RNA transcript is a copy of the whole gene and includes both introns and exons, but the introns are removed within the nucleus by splicing and the exons are joined to form the messenger RNA (mRNA). Prior to export from the nucleus, a methylated guanosine nucleotide is added to the 5′ end of the RNA (‘cap’) and a string of adenine nucleotides is added to the 3′ (‘poly A tail’). This protects the RNA from degradation and facilitates transport into the cytoplasm. In the cytoplasm, the mRNA binds to ribosomes and forms a template for protein production.

Methylated to cytoplasm. Protects and methylated form are the regulators of genes transcription.

**RNA splicing, editing and degradation**

Transcription produces an RNA molecule that is a copy of the whole gene, termed the primary or nascent transcript. RNA differs from DNA in three main ways:

- RNA is single-stranded.
- The sugar residue within the nucleotide is ribose, rather than deoxyribose.
- Uracil (U) is used in place of thymine (T).

The nascent RNA molecule then undergoes splicing, to generate the shorter, ‘mature’ mRNA molecule, which provides the template for protein production. Splicing removes the regions of the nascent RNA molecule that are not required to make protein (intronic regions), and retains and rejoins those segments that are necessary for protein production (exonic regions). Splicing is a highly regulated process that is carried out by a multimeric protein complex called the spliceosome. Following splicing, the mRNA molecule is exported from the nucleus and used as a template for protein synthesis. It should be noted that many genes produce more than one form of mRNA (and thus protein) by a process termed alternative splicing. Different proteins from the same gene can have entirely distinct functions. For example, in thyroid C cells the calcitonin gene produces mRNA encoding the osteoclast inhibitor calcitonin (p. 738), but in neurons the same gene produces an mRNA with a different complement of exons via alternative splicing, which encodes the neurotransmitter calcitonin-gene-related peptide.

The portion of the mRNA molecule that directs synthesis of a protein product is called the open reading frame (ORF). This comprises a contiguous series of three sequential bases (codons), which specify that a particular amino acid should be incorporated into the protein. There are 64 different codons; 61 of these specify incorporation of one of the 20 amino acids, whereas the remaining three codons – UAA, UAG and UGA (stop codons) – cause termination of the growing polypeptide chain. In humans, most ORF start with the amino acid methionine, which is specified by the codon AUG. All mRNA molecules have domains before and after the ORF called the 5′ untranslated region (5′UTR) and 3′UTR, respectively. The start of the 5′UTR contains a cap structure that protects mRNA from enzymatic degradation, and other elements within the 5′UTR are required for efficient translation. The 3′UTR also contains elements that regulate efficiency of translation and mRNA stability, including a stretch of adenine bases known as a polyA tail.

However, there are approximately 4500 genes in humans in which the transcribed RNA molecules do not code for proteins. There are various categories of important events in the progression of cancer, allowing expression of genes which are normally silenced during development to support cancer cell de-differentiation (see Box 3.3, p. 54). They also afford therapeutic targets. For instance, the histone deacetylase inhibitor vorinostat has been successfully used to treat cutaneous T-cell lymphoma, due to the re-expression of genes that had previously been silenced in the tumour. These genes encode transcription factors which promote T-cell cell differentiation as opposed to proliferation, thereby causing tumour regression.
non-coding RNA (ncRNA), including transfer RNA (tRNA), ribosomal RNA (rRNA), ribozymes and microRNA (miRNA). There are more than 1000 miRNAs that bind to various target mRNAs, typically in the 3'UTR, to affect mRNA stability. This usually results in enhanced degradation of the target mRNA, leading to translational gene silencing. Together, miRNAs affect over half of all human genes and have important roles in normal development, cancer and common degenerative disorders. This is the subject of considerable research interest at present.

**Translation and protein production**

Following splicing and export from the nucleus, mRNAs associate with ribosomes, which are the sites of protein production (see Fig. 3.1). Each ribosome consists of two subunits (40S and 60S), which comprise non-coding rRNA molecules complexed with proteins. During translation, tRNA binds to the ribosome. The tRNAs deliver amino acids to the ribosome so that the newly synthesised protein can be assembled in a stepwise fashion. Individual tRNA molecules bind a specific amino acid and ‘read’ the mRNA ORF via an ‘anti-codon’ of three nucleotides that is complementary to the codon in mRNA. A proportion of ribosomes are bound to the membrane of the endoplasmic reticulum (ER), a complex tubular structure that surrounds the nucleus. Proteins synthesised on these ribosomes are translocated into the lumen of the ER, where they undergo folding and processing. From here the protein may be transferred to the Golgi apparatus, where it undergoes post-translational modifications, such as glycosylation (covalent attachment of sugar moieties), to form the mature protein that can be exported into the cytoplasm or packaged into vesicles for secretion. The clinical importance of post-translational modification of proteins is shown by the severe developmental, neurological, haemostatic and soft-tissue abnormalities that occur in patients with mutations of the enzymes that catalyse the addition of chains of sugar moieties to proteins. An example is phosphomannose isomerase deficiency, in

**Fig. 3.2** Examples of genetic diseases caused by mutations in genes encoding either transcription factors or receptors.
which there is a defect in the conversion of fructose-6-phosphate to mannose-6-phosphate. This results in a defect in supply of D-mannose derivatives for glycosylation of a variety of proteins, resulting in a multi-system disorder characterised by protein-losing enteropathy, hepatic fibrosis, coagulopathy and hypoglycaemia. Post-translational modifications can also be disrupted by the synthesis of proteins with abnormal amino acid sequences. For example, the most common mutation in cystic fibrosis (ΔF508) results in an abnormal protein that cannot be exported from the ER and Golgi.

Mitochondria and energy production

The mitochondrion is the main site of energy production within the cell. Mitochondria arose during evolution via the symbiotic association with an intracellular bacterium. They have a distinctive structure with functionally distinct inner and outer membranes. Mitochondria produce energy in the form of adenosine triphosphate (ATP). ATP is mostly derived from the metabolism of glucose and fat (Fig. 3.3). Glucose cannot

---

**Fig. 3.3** Mitochondria. [A] Mitochondrial structure. There is a smooth outer membrane surrounding a convoluted inner membrane, which has inward projections called cristae. The membranes create two compartments: the inter-membrane compartment, which plays a crucial role in the electron transport chain, and the inner compartment (or matrix), which contains mitochondrial DNA and the enzymes responsible for the citric acid (Krebs) cycle and the fatty acid β-oxidation cycle. [B] Mitochondrial DNA. The mitochondrion contains several copies of a circular double-stranded DNA molecule, which has a non-coding region, and a coding region which encodes the genes responsible for energy production, mitochondrial rRNA molecules and mitochondrial tRNA molecules. ATP = adenosine triphosphate; NADH = nicotinamide adenine dinucleotide. [B] Mitochondrial energy production. Fatty acids enter the mitochondrion conjugated to carnitine by carnitine-palmitoyl transferase type I (CPT I) and, once inside the matrix, are unconjugated by CPT II to release free fatty acids (FFA). These are broken down by the β-oxidation cycle to produce acetyl-CoA. Pyruvate can enter the mitochondrion directly and is metabolised by pyruvate dehydrogenase (PDH) to produce acetyl-CoA. The acetyl-CoA enters the Krebs cycle, leading to the production of NADH and flavine adenine dinucleotide (reduced form) (FADH2), which are used by proteins in the electron transport chain to generate a hydrogen ion gradient across the inter-membrane compartment. Reduction of NADH and FADH2 by proteins I and II respectively releases electrons (e), and the energy released is used to pump protons into the inter-membrane compartment. As these electrons are exchanged between proteins in the chain, more protons are pumped across the membrane, until the electrons reach complex IV (cytochrome oxidase), which uses the energy to reduce oxygen to water. The hydrogen ion gradient is used to produce ATP by the enzyme ATP synthase, which consists of a proton channel and catalytic sites for the synthesis of ATP from ADP. When the channel opens, hydrogen ions enter the matrix down the concentration gradient, and energy is released that is used to make ATP.
enter mitochondria directly but is first metabolised to pyruvate via glycolysis. Pyruvate is then imported into the mitochondrion and metabolised to acetyl-CoA. Fatty acids are transported into the mitochondria following conjugation with carnitine and are sequentially catabolised by a process called β-oxidation to produce acetyl-CoA. The acetyl-CoA from both pyruvate and fatty acid oxidation is used in the citric acid (Krebs) cycle – a series of enzymatic reactions that produces CO₂, NADH and FADH₂. Both NADH and FADH₂ then donate electrons to the respiratory chain. Here these electrons are transferred via a complex series of reactions resulting in the formation of a proton gradient across the inner mitochondrial membrane. The gradient is used by an inner mitochondrial membrane protein, ATP synthase, to produce ATP, which is then transported to other parts of the cell. Dephosphorylation of ATP is used to produce the energy required for many cellular processes.

Each mitochondrion contains 2-10 copies of a 16 kilobase (kB) double-stranded circular DNA molecule (mtRNA). mtDNA contains 13 protein-coding genes, all involved in the respiratory chain, and the ncRNA genes required for protein synthesis within the mitochondria (see Fig. 3.3). The mutational rate of mtDNA is relatively high due to the lack of protection by chromatin. Several mtDNA diseases characterised by defects in ATP production have been described. mtDNA diseases are inherited exclusively via the maternal line (see Fig. 3.7, p. 51). This unusual inheritance pattern exists because all mtDNA in an individual is derived from that person’s mother via the egg cell, as sperm contribute no mtDNA. Inheritance has been described. mtDNA diseases characterised by defects in ATP production have been described. mtDNA diseases are inherited exclusively via the maternal line (see Fig. 3.7, p. 51). This unusual inheritance pattern exists because all mtDNA in an individual is derived from that person’s mother via the egg cell, as sperm contribute no mtDNA. Inheritance has been described. mtDNA diseases characterised by defects in ATP production have been described. mtDNA diseases are inherited exclusively via the maternal line (see Fig. 3.7, p. 51). This unusual inheritance pattern exists because all mtDNA in an individual is derived from that person’s mother via the egg cell, as sperm contribute no mtDNA. Inheritance has been described. mtDNA diseases characterised by defects in ATP production have been described. mtDNA diseases are inherited exclusively via the maternal line (see Fig. 3.7, p. 51). This unusual inheritance pattern exists because all mtDNA in an individual is derived from that person’s mother via the egg cell, as sperm contribute no mtDNA. Inheritance has been described.

### Protein degradation

The cell uses several different systems to degrade proteins and other molecules that are damaged, are potentially toxic or have simply served their purpose. The proteasome is the main site of protein degradation within the cell. The first step in proteasomal degradation is ubiquitination – the covalent attachment of a protein called ubiquitin as a side chain to the target protein. Ubiquitination is carried out by a large group of enzymes called E3 ligases, whose function is to recognise specific proteins that should be targeted for degradation by the proteasome. The E3 ligases ubiquitinate their target protein, which is then transported to a large multipeptide complex called the 26S proteasome, where it is degraded. There is mounting evidence that defects in the proteasome contribute to the pathogenesis of many diseases, particularly degenerative diseases of the nervous system like Parkinson’s disease and some types of dementia that are characterised by formation of abnormal protein aggregates (inclusion bodies) within neurons. At least one inherited disease, termed Angelman’s syndrome, is due to a mutation affecting the UBE3 E3 ligase.

Proteins with complex post-translational modifications are degraded in membrane-bound structures called lysosomes, which have an acidic pH and contain proteolytic enzymes that degrade proteins. There are many inherited defects in lysosomal enzymes that result in failure to degrade intracellular toxic substances. For instance, in Gaucher’s disease, mutations of the gene encoding lysosomal (acid) β-glucosidase lead to undigested lipid accumulating in macrophages, producing hepatosplenomegaly and, if severe, deposition in the brain and mental retardation.

Lysosomes are also crucial for the process of autophagy, a process of self-cannibalisation that allows the cell to adapt to periods of starvation by recycling cellular components. Autophagy is triggered by metabolic stress and begins with the formation of a membrane-bound vesicle called the autophagosome, which contains targeted cellular components such as long-lived proteins.

### 3.1 The structure of the respiratory chain complexes and the diseases associated with their dysfunction

<table>
<thead>
<tr>
<th>Complex</th>
<th>Enzyme</th>
<th>nDNA subunits</th>
<th>mtDNA subunits</th>
<th>Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>NADH dehydrogenase 38</td>
<td>7</td>
<td>MELAS, bilateral striatal necrosis, LHON, myopathy and exercise intolerance, Parkinsonism, Leigh’s disease, exercise myoglobinuria, leucodystrophy/myoclonic epilepsy</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Succinate dehydrogenase 4</td>
<td>0</td>
<td>Phaeochromocytoma</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Cytochrome bc complex 10</td>
<td>1</td>
<td>Parkinsonism/MELAS, cardiomyopathy, myopathy, exercise myoglobinuria</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Cytochrome c oxidase 10</td>
<td>3</td>
<td>Sideroblastic anaemia, myoclonic ataxia, deafness, myopathy, MELAS, mitochondrial encephalomyopathy, motor neuron disease-like, exercise myoglobinuria</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>ATP synthase 14</td>
<td>2</td>
<td>Leigh’s disease, NARP, bilateral striatal necrosis</td>
<td></td>
</tr>
</tbody>
</table>

1^nDNA subunits
2^mtDNA subunits = number of different protein subunits in each complex that are encoded in the nDNA and mtDNA respectively.

(LHON = Leber hereditary optic neuropathy; MELAS = myopathy, encephalopathy, lactic acidosis and stroke-like episodes; MERRF = myoclonic epilepsy and ragged red fibres; mtDNA = mitochondrial DNA; NARP = neuropathy, ataxia and retinitis pigmentosa; nDNA = nuclear DNA)
and organelles. The autophagosome then fuses with the lysosome to start the degradation and recycling process. Mutations in proteins that are crucial for formation of the autophagosome lead to neurodegenerative diseases in humans, such as juvenile neuronal ceroid lipofuscinosis (Batten’s disease), caused by autosomal recessive mutations in CLN3.

Peroxisomes are small, single membrane-bound cytoplasmic organelles containing many different oxidative enzymes such as catalase. Peroxisomes degrade hydrogen peroxide, bile acids and amino acids. However, the β-oxidation of very long-chain fatty acids appears to be their most important function, since mutations in the peroxisomal β-oxidation enzymes (or the proteins that import these enzymes into the peroxisome) result in the same severe congenital disorder as mutations that cause complete failure of peroxisomal biogenesis. This group of disorders is called Zellweger’s syndrome (cerebrohepatorenal syndrome) and is characterised by severe developmental delay, seizures, hepatomegaly and renal cysts; the biochemical diagnosis is made on the basis of elevated plasma levels of very long-chain fatty acids.

**The cell membrane and cytoskeleton**

The cell membrane is a phospholipid bilayer, with hydrophilic surfaces and a hydrophobic core (Fig. 3.4). The cell membrane is, however, much more than a simple wall. Cholesterol-rich ‘rafts’ float within the membrane, and proteins are anchored to them via the post-translational addition of complex lipid moieties. The membrane also hosts a series of transmembrane proteins that function as receptors, pores, ion channels, pumps and associated energy suppliers. These proteins allow the cell to monitor the extracellular milieu, import crucial molecules for function, and exclude or exchange unwanted substances. Many protein–protein interactions within the cell membrane are highly dynamic, and individual peptides will associate and dissociate to effect specific roles.

**Fig. 3.4** An archetypal human cell. The basic cell components required for function within a tissue: (1) cell-to-cell communication taking place via gap junctions and the various types of receptor that receive signals from the extracellular environment and transduce these into intracellular messengers; (2) the nucleus containing the chromosomal DNA; (3) intracellular organelles, including the mechanisms for proteins and lipid catabolism; (4) the cellular mechanisms for import and export of molecules across the cell membrane. (ABC = ATP-binding cassette transporters; ATP = adenosine triphosphate; cAMP = cyclic adenosine monophosphate; CTR = cystic fibrosis transmembrane regulator; CREB = cAMP response element-binding protein; GDP/GTP = guanine diphosphate/triphosphate; LDL = low-density lipoproteins; LH/FSH = luteinising hormone/follicle-stimulating hormone; PTH = parathyroid hormone; TSH = thyroid-stimulating hormone)
Molecular and genetic factors in disease

The cell membrane is permeable to hydrophobic substances, such as anaesthetic gases. Water is able to pass through the membrane via a pore formed by aquaporin proteins; mutations of an aquaporin gene cause congenital nephrogenic diabetes insipidus (p. 794). Most other molecules must be actively transported using either channels or pumps. Channels are responsible for the transport of ions and other small molecules across the cell membrane. They open and close in a highly regulated manner. The cystic fibrosis transmembrane conductance regulator (CFTR) is an example of an ion channel that is responsible for transport of chloride ions across epithelial cell membranes. Mutation of the CFTR chloride channel, highly expressed in the lung and gut, leads to defective chloride transport, producing cystic fibrosis. Pumps are highly specific for their substrate and often use energy (ATP) to drive transport against a concentration gradient.

Endocytosis is a cellular process that allows internalisation of larger complexes and molecules by invagination of plasma membrane to create intracellular vesicles. This process is typically mediated by specific binding of the particle to surface receptors. An important example is the binding of low-density lipoprotein (LDL) cholesterol-rich particles to the LDL receptor (LDLR) in a specialised region of the membrane called a clathrin pit. In some cases of familial hypercholesterolaemia (p. 453), LDLR mutations cause failure of this binding and thus reduce cellular uptake of LDL. Other LDLR mutations change a specific tyrosine in the intracellular tail of the receptor, preventing LDLR from concentrating in clathrin-coated pits and hence impairing uptake of LDL, even though LDLR bound to LDL is present elsewhere in the cell membrane.

The shape and structure of the cell are maintained by the cytoskeleton, which consists of a series of proteins which form microfilaments (actin), microtubules (tubulins) and intermediate filaments (keratins, desmin, vimentin, laminins) that facilitate cellular movement and provide pathways for intracellular transport. Dysfunction of the cytoskeleton may result in a variety of human disorders. For instance, some keratin genes encode intermediate filaments in epithelia. In epidermolysis bullosa simplex (p. 1292), mutations in keratin genes (KRT5, KRT14) lead to cell fragility, producing the characteristic blistering on mild trauma.

Receptors, cellular communication and intracellular signalling

Several mechanisms exist that allow cells to communicate with one another. Direct communication between adjacent cells occurs through gap junctions. These are pores formed by the interaction of ‘hemichannels’ in the membrane of adjacent cells. Many diseases are due to mutations in gap junction proteins, including the most common form of autosomal recessive hearing loss (GJB2) and the X-linked form of Charcot–Marie–Tooth disease (GJB1).

Communication between cells that are not directly in contact with each other occurs through hormones, cytokines and growth factors, which bind to and activate receptors on the target cell. Receptors then bind to various other proteins within the cell termed signalling molecules, which directly or indirectly activate gene expression to produce a cellular response.

There are many different signalling pathways; for example, in nuclear steroid hormone signalling, the ligands (steroid hormones or thyroid hormone) bind to their cognate receptor in the cytoplasm of target cells and the receptor/ligand complex then enters the nucleus, where it acts as a transcription factor to regulate the expression of target genes (Box 3.2). However, the most diverse and abundant types of receptor are located at the cell surface, and these activate gene expression and cellular responses indirectly. Activation of a cell surface receptor by its ligand results in a series of intracellular events, involving a cascade of phosphorylation of specific residues in target proteins by an important group of enzymes called kinases. This cascade typically culminates in phosphorylation and activation of transcription factors, which bind DNA and modulate gene expression.

Figure 3.5 depicts some of the signalling molecules downstream of the tumour necrosis factor (TNF) receptor. On activation of the receptor by the ligand (in this case, TNF), other molecules, including TNF-receptor-associated proteins (TRAFs), are recruited to the intracellular domain of the receptor. These regulate the activity of a kinase termed IKKγ, which in turn regulates activity of two further kinases termed IKKα and IKKβ. These regulate degradation of an inhibitory protein called IκB, which normally binds to the effector protein NFXB, holding it in the cytoplasm. On receptor activation, a signal is transmitted through TRAFs and the IKK proteins to cause phosphorylation and
degradation of 1kB, allowing NFkB to translocate to the nucleus and activate gene expression. The system also has negative regulators, including the ciliopathies (CYLD) enzyme, which regulates the activity of TRAFs by de-ubiquitination. Other transmembrane receptors can be grouped into:

- ion channel-linked receptors (glutamate and the nicotinic acetylcholine receptor)
- G protein-coupled receptors (GnRH, rhodopsin, olfactory receptors, parathyroid hormone receptor)
- receptors with kinase activity (insulin receptor, erythropoietin receptor, growth factor receptors)
- receptors which have no kinase activity, but interact with kinases via their intracellular domain when activated by ligand (TNF receptor) (see Figure 3.5 and Box 3.2).

Many receptors can signal only when they assemble as a multimeric complex. Mutations which interfere with assembly of the functional receptor multimer can result in disease. For example, mutations of the insulin receptor that inhibit dimerisation lead to childhood insulin resistance and growth failure. Conversely, some fibroblast growth factor receptor 2 (FGFR2) gene mutations cause dimerisation in the absence of ligand binding, leading to bone overgrowth and an autosomal dominant form of craniosynostosis called Crouzon’s syndrome.

It is becoming clear that specialised projections on the cell surface known as cilia are essential for normal signalling in many tissues. Cilia can be motile or non-motile. Motile cilia are crucial for normal respiratory tract function, with primary ciliary dyskinesia (PCD) resulting in early-onset bronchiectasis due to failure to clear lung secretions. PCD is commonly associated with situs inversus (left-right laterality reversal) as a result of failure of a specific signalling process in very early embryogenesis. Mutations in proteins that are essential for non-motile cilia formation or function are responsible for a large number of autosomal recessive disorders known collectively as ciliopathies, which are commonly associated with intellectual disability, renal cystic dysplasia and retinal degeneration. For example, in the Bardet–Biedl syndrome, mutations in a series of genes associated with intellectual disability, renal cystic dysplasia and retinal degeneration. For example, in the Bardet–Biedl syndrome, mutations in a series of genes associated with intellectual disability, renal cystic dysplasia and retinal degeneration. For example, in the Bardet–Biedl syndrome, mutations in a series of genes associated with intellectual disability, renal cystic dysplasia and retinal degeneration. For example, in the Bardet–Biedl syndrome, mutations in a series of genes associated with intellectual disability, renal cystic dysplasia and retinal degeneration.

### 3.2 Examples of molecules involved in specific signalling cascades

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Receptor type</th>
<th>Ligands</th>
<th>Signal transduction</th>
<th>Clinical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFR1</td>
<td>TNF receptor superfamily</td>
<td>TNF</td>
<td>TRAF2/5, TRADD, IKK, 1kB, NFkB, CYLD, RANK</td>
<td>Mediator of inflammatory diseases and immune responses</td>
</tr>
<tr>
<td>RANK</td>
<td>TNF receptor superfamily</td>
<td>RANKL</td>
<td>TRAF6, IKK, 1kB, NFkB</td>
<td>Regulates bone resorption</td>
</tr>
<tr>
<td>Insulin receptor</td>
<td>Receptor tyrosine kinase</td>
<td>Insulin</td>
<td>IRS1, PI3K, PIK3, PKB, PDK1, mTORC2, GSK3</td>
<td>Regulation of energy homeostasis and glucose metabolism</td>
</tr>
<tr>
<td>Erythropoietin receptor</td>
<td>Receptor tyrosine kinase</td>
<td>Erythropoietin</td>
<td>JAK2, STAT5, c-Jun, c-Fos, Src PI3K, PI3K, PDK1, PKB</td>
<td>Regulates erythropoiesis</td>
</tr>
<tr>
<td>THRx and THrβ</td>
<td>Nuclear receptor superfamily</td>
<td>T3</td>
<td>Ligand/receptor complex</td>
<td>Regulates differentiation and function of many cells and tissues</td>
</tr>
<tr>
<td>ERα and ERβ</td>
<td>Nuclear receptor superfamily</td>
<td>Oestrogen</td>
<td>Ligand/receptor complex</td>
<td>Important for fertility, reproduction and bone health</td>
</tr>
<tr>
<td>GnRHR</td>
<td>GPCR</td>
<td>GnRH</td>
<td>G/J/G11, PLCbeta1, PLA(2), PLD, PKC, MAPK</td>
<td>Regulates fertility</td>
</tr>
<tr>
<td>PTHR1</td>
<td>GPCR</td>
<td>PTH, PTHL</td>
<td>Gs, adenyly cyclase, cAMP, PKA, CREB, G/J/G11, PLC, DAG, IP3, PKC, Ca++</td>
<td>Regulates calcium homeostasis and bone turnover</td>
</tr>
</tbody>
</table>

(cAMP = cyclic adenosine monophosphate; CREB = Ca++ intracellular calcium; CYLD = ciliopathy; DAG = diacylglycerol; ER = (o)estrogen receptor; GnRH = gonadotrophin releasing hormone receptor; GPCR = G protein-coupled receptor; G/J/G11/Gs = guanine nucleotide binding proteins; GSK3 = glycogen synthetase kinase 3; 1kB = inhibitor of kappa B; IKK = I kappa B kinase; IP3 = D-myo-inositol-1,4,5-trisphosphate; IRS1 = insulin receptor substrate 1; JAK2 = Janus activated kinase 2; MAPK = mitogen-activated kinase; mTOR = mammalian target of rapamycin; NFkB = nuclear factor kappa B; PDK1 = phosphoinositol-dependent kinase 1; PI3K = phosphoinositol 3 kinase; PKA/PKB/PKC = protein kinase A/B/C; PLA/PLC/PLD = phospholipase A/C/D; PTHR1 = parathyroid hormone receptor 1; PTHL = parathyroid hormone–like protein; RANK = receptor activator of nuclear factor kappa B; STAT = signal transducer and activator of transcription; TNF = tumour necrosis factor; TNFR1 = TNF receptor 1; TRAF = TNF receptor-associated factors; TRADD = tumour necrosis factor receptor type 1–associated death domain protein; TRH = thyrotrophin releasing hormone)

**Cell division, differentiation and migration**

In normal tissues, molecules such as hormones, growth factors and cytokines provide the signal to activate the cell cycle, a controlled programme of biochemical events that culminates in cell division. During the first phase, G1, synthesis of the cellular components necessary to complete cell division occurs. In S phase, the cell produces an identical copy of each chromosome – which carries the cell’s genetic information – via a process called DNA replication. The cell then enters G2, when any errors in the replicated DNA are repaired before proceeding to mitosis, in which identical copies of all chromosomes are segregated to the daughter cells. The progression from one phase to the next is tightly controlled by cell cycle checkpoints. For example, the checkpoint between G2 and mitosis ensures that all damaged DNA is repaired prior to segregation of the chromosomes. Failure of these control processes is a crucial driver in the pathogenesis of cancer, as discussed in Chapter 11 (p. 262).
Molecular and genetic factors in disease

During development, cells must become progressively less like a stem cell and acquire the morphological and biochemical configuration of the tissue to which they will contribute. This process is called differentiation and it is achieved by activation of tissue-specific genes and inactivation or silencing of genes that maintain the cell in a progenitor state. This epigenetic process enables cells containing the same genetic material to have very different structures and functions. The programme of differentiation is often deranged or partially reversed in cancer cells. A similar mechanism allows adult stem cells to maintain and repair tissues. Cell migration is a process that is also necessary for development and wound healing. Migration also requires the activation of a specific set of genes, such as the transcription factor TWIST, that give the cell polarity and enable the leading edge of the cell to interact with the extracellular environment to control the speed and direction of travel. Again, this process can be reactivated in cancer cells and is thought to facilitate tumour metastasis.

Cell death, apoptosis and senescence

With the exception of stem cells, human cells have only a limited capacity for cell division. The Hayflick limit is the number of divisions a cell population can go through in culture before division stops and the cell enters a state known as senescence. This ‘biological clock’ is of great interest in the study of the normal ageing process. Rare human diseases associated with premature ageing, called progeric syndromes, have been very helpful in identifying the importance of DNA repair mechanisms in senescence (p. 168). For example, in Werner syndrome, a DNA helicase (an enzyme that separates the two DNA strands) is mutated, leading to failure of DNA repair and premature ageing. A distinct mechanism of cell death is seen in apoptosis, or programmed cell death.

Apoptosis is an active process that occurs in normal tissues and plays an important role in development, tissue remodelling and the immune response. The signal that triggers apoptosis is specific to each tissue or cell type. This signal activates enzymes, called caspases, which actively destroy cellular components, including chromosomal DNA. This degradation results in cell death, but the cellular corpse contains characteristic vesicles called apoptotic bodies. The corpse is then recognised and removed by phagocytic cells of the immune system, such as macrophages, in a manner that does not provoke an inflammatory response.

A third mechanism of cell death is necrosis. This is a pathological process in which the cellular environment loses one or more of the components necessary for cell viability. Hypoxia is probably the most common cause of necrosis.

Genetic disease and inheritance

Meiosis is a special form of cell division that only occurs in the post-pubertal testis and the fetal and adult ovary (Fig. 3.6). Meiosis differs from mitosis in two main ways; there are two separate cell divisions and before the first of these there is extensive swapping of genetic material between homologous chromosomes, a process known as recombination. The result of recombination is that each chromosome that a parent passes to his or her offspring is a mix of the chromosomes that the parent inherited from his or her own mother and father. The
end products of meiosis are sperm and egg cells (gametes), which contain only 23 chromosomes: one of each homologous pair of autosomes and a sex chromosome. When a sperm cell fertilises the egg, the resulting zygote will thus return to a diploid chromosome complement of 46 chromosomes. The sperm determines the sex of the offspring, since 50% of sperm will carry an X chromosome and 50% a Y chromosome, while each egg cell carries an X chromosome.

The individual steps in meiotic cell division are similar in males and females. However, the timing of the cell divisions is very different (see Fig. 3.6). In females, meiosis begins in fetal life but does not complete until after ovulation. A single meiotic cell division can thus take more than 40 years to complete. In males, meiotic division does not begin until puberty and continues throughout life. In the testes, both meiotic divisions are completed in a matter of days.

### Patterns of disease inheritance

Five modes of genetic disease inheritance are discussed below and illustrated in Figures 3.7 and 3.8.

**Autosomal dominant inheritance**

Autosomal dominant disorders result from a genetic abnormality in one of the two copies (alleles) of a single gene. The risk of an affected individual transmitting an autosomal disease to his or her offspring is 50% for each pregnancy, since half the affected individual gametes (sperm or egg cells) will contain the affected chromosome and half will contain the normal chromosome. However, even within a family, individuals with the same mutation rarely have identical patterns of disease due to variable penetrance and/or expressivity. Penetration is defined as the proportion of individuals bearing a mutated allele who develop the disease phenotype. The mutation is said to be fully penetrant if all individuals who inherit a mutation develop the disease. Expressivity describes the level of severity of each aspect of the disease phenotype. Neurofibromatosis type 1 (NF1, neurofibromin, 17q11.2) is an example of a disease that is fully (100%) penetrant but which shows extremely variable expressivity. The environmental factors and/or variation in other genes that act as modifiers of the mutated gene’s function are mostly unknown. A good example of an environmental influence that can profoundly influence expression of autosomal dominant

![Fig. 3.7 Drawing a pedigree and patterns of inheritance.](image-url)

A The main symbols used to represent pedigrees in diagrammatic form. B The main modes of disease inheritance (see text for details).
Molecular and genetic factors in disease

Fig. 3.8 Genomic imprinting and associated diseases. Several regions of the genome exhibit the phenomenon of imprinting, whereby expression of one or a group of genes is influenced by whether the chromosome is derived from the mother or the father; one such region lies on chromosome 15.

**A** Normal imprinting. Under normal circumstances, expression of several genes is suppressed (silenced) on the maternal chromosome (red), whereas these genes are expressed normally by the paternal chromosome (blue). However, two genes in the paternal chromosome (UBE3 and ATP10A) are silenced.

**B** In sporadic Prader–Willi syndrome, there is a non-disjunction defect on chromosome 15, and both copies of the chromosomal region are derived from the mother (maternal uniparental disomy). In this case, Prader–Willi syndrome occurs because there is loss of function of several paternally expressed genes, including MKRN3, MAGEK2, NDN, PWRN2, C15orf2, SNURF-SNRNP.

**C** In sporadic Angelman’s syndrome, both chromosomal regions are derived from the father (paternal uniparental disomy) due to non-disjunction during paternal meiosis. As a result, both copies of the UBE3 gene are silenced and this causes Angelman’s syndrome. Note that the syndrome can also be caused by deletion of this region on the maternal chromosome or a loss-of-function mutation on the maternal copy of UBE3, causing an inherited form of Angelman’s, as illustrated in panel D.

**D** Pedigree of a family with inherited Angelman’s syndrome due to a loss-of-function mutation in UBE3. Inheriting this mutation from a father causes no disease (because the gene is normally silenced in the paternal chromosome) (see individuals I-1, II-1, II-3, III-6), but the same mutation inherited from the mother causes the syndrome (individuals III-3, III-4, IV-4), as this is the only copy expressed and the UBE3 gene is mutated.
Homozygosity is increased. Genetic risk calculation of inheriting the same mutant allele from both parents frequency of autosomal recessive disorders increases with who have had a previous child affected by an autosomal straightforward. Each subsequent pregnancy of a couple for a fully penetrant autosomal recessive disorder is serious autosomal recessive disorder in the UK is cystic fibrosis, which has a birth incidence of 1:2000. The frequency of autosomal recessive disorders increases with the degree of inbreeding of a population because the risk of inheriting the same mutant allele from both parents (homozygosity) is increased. Genetic risk calculation for a fully penetrant autosomal recessive disorder is straightforward. Each subsequent pregnancy of a couple who have had a previous child affected by an autosomal recessive disorder will have a 25% (1:4) risk of being affected; a healthy individual who has a sibling with an autosomal recessive disorder will have a 25% (1:4) risk of being affected. The risk of an affected individual having children with the same condition is usually low but is dependent on the carrier rate of the mutant allele in the population.

**Autosomal recessive inheritance**

In autosomal recessive disorders, both alleles of a gene must be mutated before the disease is manifest in an individual, and an affected individual must inherit one mutant allele from each parent. What distinguishes autosomal dominant and recessive diseases is that carrying one mutant allele does not produce a disease phenotype. Autosomal recessive disorders are rare in most populations. For example, the most common serious autosomal recessive disorder in the UK is cystic fibrosis, which has a birth incidence of 1:2000. The frequency of autosomal recessive disorders increases with the degree of inbreeding of a population because the risk of inheriting the same mutant allele from both parents (homozygosity) is increased. Genetic risk calculation for a fully penetrant autosomal recessive disorder is straightforward. Each subsequent pregnancy of a couple who have had a previous child affected by an autosomal recessive disorder will have a 25% (1:4) risk of being affected; a healthy individual who has a sibling with an autosomal recessive disorder will have 2/3 chance of being a carrier. The risk of an affected individual having children with the same condition is usually low but is dependent on the carrier rate of the mutant allele in the population.

**X-linked inheritance**

Genetic diseases caused by mutations on the X chromosome have specific characteristics. X-linked diseases are mostly recessive and restricted to males who carry the mutant allele. This is because males have only one X chromosome, whereas females have two. Thus females who carry a single mutant allele are generally unaffected. Occasionally, female carriers may exhibit signs of an X-linked disease due to a phenomenon called skewed X-inactivation. All female embryos, at about 100 cells in size, stably inactivate one of their two X chromosomes in each cell. This process is random in each cell but if, by chance, there is a disproportionate inactivation of normal X chromosomes carrying the normal allele, then an affected female carrier will be more likely, an extreme example being the rare cases of carrier females affected with Duchenne muscular dystrophy. X-linked recessive disorders have a recognisable pattern of inheritance, with transmission of the disease from carrier females to affected males and absence of father-to-son transmission. The risk of a female carrier having an affected child is 25% (1:4; half of her male offspring). If the carrier status of a woman is unclear, then the risk may be altered by conditional information, as discussed in the autosomal dominant disease section above. Bayes’ theorem is commonly used to calculate such modified risks and this is discussed in more detail later in this chapter (p. 68).

**Mitochondrial inheritance**

The inheritance of mtDNA disorders is characterised by transmission from females, but males and females are generally affected equally. Unlike the other inheritance patterns mentioned above, mitochondrial inheritance has nothing to do with meiosis but reflects the fact that mitochondrial DNA is transmitted by oocytes. Mitochondrial disorders tend to be very variable in penetrance and expressivity within families, and this is mostly accounted for by the fact that only a proportion of multiple mtDNA molecules within mitochondria contain the causal mutation (the degree of mtDNA heteroplasmy).

**Epigenetic inheritance and imprinting**

Several chromosomal regions (loci) have been identified where gene repression is inherited in a parent-of-origin-specific manner; these are called imprinted loci. Within these loci the paternal alleles of a gene may be active while the maternal one may be silenced, or vice versa (see Fig. 3.8). Mutations within imprinted loci lead to a very unusual pattern of inheritance in which the phenotype is only manifest if inherited from the parent who contributes the transcriptionally active allele (see Fig. 3.8). Examples of these disorders are given in Box 3.3.

### Classes of genetic variant

There are many different classes of variation in the human genome (Figs 3.9 and 3.10). Rare genetic variations that result in a disease are generally referred to as mutations, whereas common variations and those that do not cause disease are referred to as polymorphisms. These different types of variation are further categorised by the size of the DNA segment involved and/or by the mechanism giving rise to the variation.

**Nucleotide substitutions**

The substitution of one nucleotide for another is the most common type of variation in the human genome. Depending on their frequency and functional consequences, these changes are known as a point mutation or a single nucleotide polymorphism (SNP). They occur by misincorporation of a nucleotide during DNA synthesis or by the action of a chemical mutagen. When these substitutions occur within ORFs of a protein-coding gene, they are further classified into:

- **Synonymous** – resulting in a change in the codon but no change in the amino acid and thus no phenotype
- **Missense** – altering a codon, resulting in an amino acid change in the protein
- **Nonsense** – introducing a premature stop codon, resulting in truncation of the protein
- **Splicing** – occurring at the junction of an intron and an exon, thereby adversely affecting splicing.

Examples of these types of variation are shown in Figures 3.9 and 3.10.

---

**Figure 3.9**

Examples of these types of variation are shown in Figures 3.9 and 3.10.

**Figure 3.10**

Examples of these types of variation are shown in Figures 3.9 and 3.10.
**3.3 Epigenetic disease**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Genes</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Imprinting disorders</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beckwith–Wiedemann syndrome</td>
<td>11p15</td>
<td>p57KIP2, HASH2, INS2, IGF2, H19</td>
<td>General 'over-growth', advanced bone age and increased childhood tumours. Some cases due to mutations in p57KIP2</td>
</tr>
<tr>
<td>Prader–Willi syndrome</td>
<td>15q11–q13</td>
<td>SNRPN, Necdin and others</td>
<td>Obesity, hypogonadism and learning disability. Lack of paternal contribution (due to deletion of paternal 15q11–q13, or inheritance of both chromosome 15q11–q13 regions from the mother)</td>
</tr>
<tr>
<td>Angelman’s syndrome (AS)</td>
<td>15q11–q13</td>
<td>UBE3A</td>
<td>Severe mental retardation, ataxia, epilepsy and inappropriate laughing bouts. Due to loss-of-function mutations in the maternal UBE3A gene. The neurological phenotype results because most tissues express both maternal and paternal alleles of UBE3A, whereas the brain expresses predominantly the maternal allele</td>
</tr>
<tr>
<td>Pseudohypoparathyroidism</td>
<td>20q13</td>
<td>GNAS1</td>
<td>Inheritance of the mutation from the mother results in hypocalcaemia, hyperphosphataemia, raised parathyroid hormone (PTH) levels, ectopic calcification, obesity, delayed puberty, shortened 4th and 5th metacarpals and ectopic calcification. When the mutation is inherited from the father, PTH, calcium and phosphate levels are normal but the other features are present. These differences are due to the fact that, in the kidney (the main target organ through which PTH regulates serum calcium and phosphate), the paternal allele is silenced and the maternal allele is expressed, whereas both alleles are expressed in other tissues</td>
</tr>
<tr>
<td><strong>X-inactivation disorders</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duchenne muscular dystrophy (DMD) in females</td>
<td>Xp22</td>
<td>DMD</td>
<td>If, by chance, a sufficient number of the X chromosomes containing the normal dystrophin gene are inactivated in muscle, heterozygous females may rarely develop full-blown DMD. Conversely, if a higher proportion of the disease gene-carrying chromosome is inactivated, a carrier female may test negative on biochemical screening for elevated creatine kinase levels</td>
</tr>
<tr>
<td><strong>Epigenetic silencing (oncogenesis)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon cancer</td>
<td>3p21</td>
<td>MLH1</td>
<td>Hypermethylation of the promoter results in silencing of MLH1, which encodes a DNA repair gene</td>
</tr>
</tbody>
</table>

![Fig. 3.9 Different types of mutation affecting coding exons.](image)

**Fig. 3.9 Different types of mutation affecting coding exons.**

[A] Normal sequence.

[B] A synonymous nucleotide substitution changing the third base of a codon; the resulting amino acid sequence is unchanged.

[C] A missense mutation in which the nucleotide substitution results in a change in a single amino acid from the normal sequence (AAG) encoding lysine to glutamine (CAG).

[D] Insertion of a G residue (boxed) causes a frameshift mutation, completely altering the amino acid sequence downstream. This usually results in a loss-of-function mutation.

[E] A nonsense mutation resulting in a single nucleotide change from a lysine codon (AAG) to a premature stop codon (TAG).
Insertions and deletions

One or more nucleotides may be inserted or lost in a DNA sequence, resulting in an insertion/deletion (indel) polymorphism or mutation (see Fig. 3.9). If an indel change affects one or two nucleotides within the ORF of a protein-coding gene, this can have serious consequences because the triple nucleotide sequence of the codons is disrupted, resulting in a frameshift mutation. The effect upon the gene is typically severe because the amino acid sequence is totally disrupted.

Simple tandem repeat mutation

Variations in the length of simple tandem repeats of DNA are thought to arise as the result of slippage of DNA during meiosis and are termed microsatellite (small) or minisatellite (larger) repeats. These repeats are unstable and can expand or contract in different generations. This instability is proportional to the size of the original repeat, in that longer repeats tend to be more unstable. Many microsatellites and minisatellites occur in introns or in chromosomal regions between genes and have no obvious adverse effects. However, some genetic diseases, including Huntington’s disease and myotonic dystrophy, are caused by microsatellite repeats, which result in duplication of amino acids within the affected gene product or affect gene expression (Box 3.4).

Copy number variations

Variation in the number of copies of an individual segment of the genome from the usual diploid (two copies) content can be categorised by the size of the segment involved. Rarely, individuals may gain or lose a whole chromosome. Such numerical chromosome anomalies most commonly occur by a process known as meiotic non-dysjunction (Box 3.5). This is the most common cause of Down’s syndrome, which results from trisomy (three copies) of chromosome 21.

Large insertions or deletions of chromosomal DNA also occur and are usually associated with learning disability and/or malformations. Such structural chromosomal anomalies arise as the result of two different processes:

- non-homologous end-joining
- non-allelic homologous recombination.

Random double-stranded breaks in DNA are a necessary process in meiotic recombination and also occur during mitosis at a predictable rate. The rate of these breaks is dramatically increased by exposure to ionising radiation. When such breaks occur, they are usually repaired accurately by DNA repair mechanisms within the cell. However, a proportion of breaks undergoes non-homologous end-joining, which results in the joining of two segments of DNA that are not normally contiguous. If the joined fragments are from different chromosomes, this results in a translocation. If they are from the same chromosome, this will result in inversion, duplication or deletion of a chromosomal fragment (Fig. 3.11). Large insertions and deletions may be cytogenetically visible as chromosomal deletions or duplications. If the anomalies are too small to be detected by microscopy, they are termed microdeletions and microduplications. Many microdeletion syndromes have been described and most stem from non-allelic homologous recombination between repeats of highly similar DNA sequences, which results in identical chromosome
anomalies – and clinical syndromes – occurring in unrelated individuals (see Fig. 3.11 and Box 3.5).

Polymorphic copy number variants

In addition to the disease-causing structural chromosomal anomalies mentioned above, there are also a considerable number of polymorphic CNVs that exist as common genetic polymorphisms in humans. These involve duplication of large segments of the genome, often containing multiple genes and regulatory elements. These duplications usually result from non-allelic homologous recombination via misalignment of tandem
repeated DNA elements in the chromosome during recombination (see Fig. 3.11). The consequences of CNV for genetic disease have not been fully explored, although recent studies have shown a strong association between an increased copy number of the gene FCGR3B and the risk of systemic lupus erythematosus.

Consequences of genetic variation

Genetic variants can generally be classed into three groups:

- those associated with no detectable change in gene function (neutral variants)
- those which cause a loss of function of the gene product
- those which cause a gain of function of the gene product.

The consequence of an individual mutation depends on many factors, including the mutation type, the nature of the gene product and the position of the variant in the protein. Mutations can have profound effects or subtle effects on gene and cell function (Box 3.6). Variations

Fig. 3.11 Chromosomal analysis and structural chromosomal disorders. A How chromosome analysis is carried out. Starting with a blood sample, the white cells are stimulated to divide by adding the mitogen phytohaemagglutinin (PHA), and colchicine is used to trap the cells in metaphase, which allows the chromosomes to be seen using light microscopy following staining with Giemsa, resulting in a banding pattern. B How structural chromosomal anomalies are described. Human chromosomes can be classed as metacentric if the centromere is near the middle, or acrocentric if the centromere is at the end. The bands of each chromosome are given a number, starting at the centromere and working out along the short (p) arm and long (q) arm. Translocations and inversions are balanced structural chromosome anomalies where no genetic material is missing but it is in the wrong order. Translocations can be divided into reciprocal (direct swap of chromosomal material between non-homologous chromosomes) and Robertsonian (fusion of acrocentric chromosomes). Deletions and duplications can also occur due to non-allelic homologous recombination (illustrated in panel C). Deletions are classified as interstitial if they lie within a chromosome, and terminal if the terminal region of the chromosome is affected. Duplications can either be in tandem (where the duplicated fragment is orientated in the correct direction) or inverted (where the duplicated fragment is in the wrong direction). N = normal; A = abnormal. C A common error of meiotic recombination, known as non-allelic homologous recombination, can occur (right panel), resulting in a deletion on one chromosome and a duplication in the homologous chromosome. The error is induced by tandem repeats in the DNA sequences (green), which can misalign and bind to each other, thereby ‘threading’ the DNA into thinking the pairing prior to recombination is correct.
that have profound effects are responsible for ‘classical’ genetic diseases, whereas those with subtle effects may contribute to the pathogenesis of complex diseases with a genetic component.

**Loss-of-function mutations**

These mutations cause the normal function of a protein to be reduced or lost. Deletion of the whole gene is the most extreme example but the same phenotype can be seen with a nonsense or frameshift mutation early in the ORF. Missense mutations that alter a critical domain within the protein can also result in loss of function. In autosomal recessive diseases, mutations that result in no protein function whatsoever are known as null mutations. If loss-of-function mutations result in an autosomal dominant disease, the genetic mechanism is known as haploinsufficiency and indicates that both functional copies of the gene are required for normal cellular function. Mutations in *PKD1* or *PKD2* that cause autosomal dominant adult polycystic kidney disease are mostly loss of function.

**Gain-of-function and dominant negative mutations**

Gain-of-function and dominant-negative effect mutations are most commonly the result of missense mutation or in-frame deletions but may also be caused by triplet repeat expansion mutations. Gain of function results where a mutation alters the protein structure, causing activation of its normal function, causing it to interact with a novel substrate or causing it to change its normal function. Constitutive activation of fibroblast growth factor receptors by missense mutation, which leads to many disorders such as achondroplasia, is an example of a gain-of-function mutation. Dominant-negative mutations are heterozygous changes that have a more deleterious effect on the protein function than a heterozygous ‘null’ mutation. For example, heterozygous mutations in *FBN1* cause Marfan’s syndrome by the production of a protein with an abnormal amino acid sequence that disrupts the normal assembly of microfibrils. In comparison, complete loss of function of one allele of *FBN1* is usually completely benign.

**Polymorphisms**

A polymorphism is defined as a change in the nucleotide sequence that exists with a population frequency of more than 1%. Most common polymorphisms are neutral (see below), but some cause subtle changes in gene expression or in protein structure and function (see Box 3.15, p. 69). It is thought that these polymorphisms lead to variations in phenotype within the general population, including variations in susceptibility to common diseases. An example is polymorphism in the gene *SLC2A9* that not only explains a significant proportion of the normal population variation in serum urate concentration but also predisposes ‘high-risk’ allele carriers to the development of gout. Other examples are listed in Box 3.6.

**Neutral variants**

The vast majority of variations within the human genome have no discernible effect on the cell or organism. This may be because the variation is non-coding, occurring outside the gene but within an intron, or is within the coding regions of a gene but does not change the amino acid because of a synonymous substitution at the third base of a codon (see Fig. 3.9). Some variations that do change the amino acid may be completely tolerated with regard to protein function.

**Evolutionary selection**

Genetic variants play an important role in evolutionary selection; some are advantageous to an organism, resulting in positive selection through evolution via improved reproductive fitness. However, variations that decrease reproductive fitness become less common and are excluded through evolution. Given this simple paradigm, it would be tempting to assume that common mutations are all advantageous and all rare mutations are pathogenic. Unfortunately, it is often difficult to classify any common mutation as either advantageous or deleterious – or, indeed, neutral. Mutations that are advantageous in early life and thus enhance reproductive fitness may be deleterious in later life. There may be mutations that are advantageous for survival in particular conditions (for example, famine or pandemic), which may be disadvantageous in more benign circumstances by resulting in a predisposition to obesity or autoimmune disorders. This complexity of balancing selection through evolution is likely to be an important feature of the genetics of common disease.

**Constitutional genetic disease**

All familial genetic disease is caused by constitutional mutations, which are inherited through the germ line.
However, different mutations in the same gene can have different consequences, depending on the genetic mechanism underlying that disease. About 1% of the population carries constitutional mutations that cause disease.

**Allelic heterogeneity**

Allelic heterogeneity is where several different mutations cause the same phenotype. This is seen in almost all genetic disease. In familial adenomatous polyposis coli, whole-gene deletions, nonsense mutations, frameshift mutations, and some missense mutations result in exactly the same phenotype because they all cause loss of function in the **FAP** gene on chromosome 5q. Many other Mendelian disorders show this phenomenon with loss-of-function mutations, including adult polycystic kidney disease (PKD1, 16p13; PKD2, 4q21). Allelic heterogeneity can also be seen in gain-of-function and dominant-negative mutations. In connective tissue disorders, dominant-negative mutations are almost always missense mutations or in-frame deletions or insertions, since the aberrant protein has to be made for the disease to manifest. In most diseases caused by gain-of-function mutations, allelic heterogeneity is severely restricted. A good example of this is achondroplasia, in which the mutations in **FGFR3** are restricted to a few specific codons that cause constitutive activation of the receptor that is required to cause the disease.

**Locus heterogeneity**

Locus heterogeneity is where a similar phenotype results from mutations in several different genes. One of the best examples is retinitis pigmentosa, which can occur as the result of mutations in more than 75 genes, each of which has a different chromosomal location.

**De novo mutations**

Although the vast majority of constitutional mutations are inherited, each gamete will contain mutations that have occurred as a result of meiosis; these are called de novo mutations. Each individual has approximately 70 de novo mutations scattered throughout their genome. This occurs in each generation and is presumably required for evolution to occur. Most are neutral but such mutations may also cause human disease. De novo mutations cause severe congenital disorders such as thanatophoric dysplasia (**FGFR3** gain-of-function mutation), bilateral anophthalma (**SOX2** haplinsufficiency), campomelic dysplasia (**SOX9** loss of function) (Fig. 3.2) and the severe form of osteogenesis imperfecta (dominant-negative mutations in **COL1A1** or **COL1A2**).

**Somatic genetic disease**

Somatic mutations are not inherited but instead occur during post-zygotic mitotic cell divisions at any point from embryonic development to late adult life. An example of this phenomenon is polyostotic fibrous dysplasia (McCune–Albright syndrome), in which a somatic mutation in the **G** α<sub>11**</sub> alpha gene causes constitutive activation of receptor signalling downstream of many G protein-coupled receptors, resulting in focal lesions in the skeleton and endocrine dysfunction (p. 770).

The most important example of human disease caused by somatic mutations is cancer. Here, ‘driver’ mutations occur within genes that are involved in regulating cell division or apoptosis, resulting in abnormal cell growth and tumour formation. The two general categories of cancer-causing mutation are gain-of-function mutations in growth-promoting genes (oncogenes) and loss-of-function mutations in growth-suppressing genes (tumour suppressor genes). Whichever mechanism is acting, most tumours require an initiating mutation in a single cell that can then escape from normal growth controls. This cell replicates more frequently or fails to undergo programmed death, resulting in clonal expansion. As the size of the clone increases, one or more cells may acquire additional mutations that confer further growth advantage, leading to proliferation of these subclones, which may ultimately lead to aggressive metastatic cancer. The cell’s complex self-regulating machinery means that more than one mutation is usually required to produce a malignant tumour (see Fig. 11.3, p. 264). For example, if a mutation results in activation of a growth factor gene or receptor, then that cell will replicate more frequently as a result of autocrine stimulation. However, this mutant cell will still be subject to normal cell cycle checkpoints to promote DNA integrity in its progeny. But if additional mutations in the same cell result in defective cell cycle checkpoints, it will rapidly accumulate further mutations, which may allow completely unregulated growth and/or separation from its matrix and cellular attachments and/or resistance to apoptosis. As cell growth becomes increasingly dysregulated, cells de-differentiate, lose their response to normal tissue environment and cease to appropriate mitotic chromosomal segregation. These processes combine to generate the classical malignant characteristics of disorganised growth, variable levels of differentiation, and numerical and structural chromosome abnormalities. An increase in somatic mutation rate can occur on exposure to external mutagens, such as ultraviolet light or cigarette smoke, or if the cell has defects in DNA repair systems. Cancer therefore affects the fundamental processes of molecular and cell biology.

In many familial cancer syndromes, somatic mutations act together with an inherited mutation to cause cancer. Familial cancer syndromes may be due to loss-of-function mutations in tumour suppressor genes or genes encoding DNA repair enzymes. In DNA repair diseases, the inherited mutations increase the somatic mutation rate. Autosomal dominant mutations in genes encoding components of specific DNA repair systems are relatively common causes of familial colon cancer and breast cancer (e.g. **BRCA1**). Autosomal recessive DNA repair disorders are rare and are associated with almost complete loss of DNA repair enzymes. This is usually associated with a severe multifaceted degenerative disorder with cancer susceptibility as a significant component (e.g. xeroderma pigmentosum, p. 267).

Cancer syndromes are also caused by loss-of-function mutations in tumour suppressor genes. At the cellular level, loss of one functional copy of a tumour suppressor gene does not have any functional consequences, as the cell is protected by the remaining normal copy. However, a somatic mutation affecting the normal allele is likely to occur in one cell at some point during life, resulting in complete loss of tumour suppressor activity and a tumour developing by clonal expansion of that cell. This
two-hit mechanism (one inherited, one somatic) for cancer development is known as the Knudsen hypothesis. It explains why tumours may not develop for many years (or ever) in some members of these cancer-prone families. Yet another group of cancer syndromes are the result of gain-of-function mutations in tumour promoter genes (proto-oncogenes).

**INVESTIGATION OF GENETIC DISEASE**

**General principles of diagnosis**

Many genetic diseases can be diagnosed by a careful clinical history and examination together with an awareness and knowledge of rare diseases. Although DNA-based diagnostic tools are now widely used, not all diagnostic genetic tests involve analysis of DNA. For example, an electrocardiogram (ECG) can establish the diagnosis in long QT syndrome or a renal ultrasound can detect adult polycystic kidney disease. By definition, all genetic testing (whether DNA-based or not) has implications for both the patient and other members of the family. These issues should be considered before genetic testing is undertaken and a plan should be in place to deliver medical information and support to family members and to organise any relevant downstream investigations.

**Constructing a family tree**

The family tree – or pedigree – is fundamental to the diagnosis of genetic diseases. The basic symbols and nomenclature used in drawing a pedigree are shown in Figure 3.7 (p. 51). A three-generation family history taken in a routine medical clerking may reveal important genetic information of relevance to the presenting complaint, particularly relating to cancer.

A pedigree should include details from both sides of the family, any history of pregnancy loss or infant death, consanguinity, and details of all medical conditions in family members, including dates of birth and age at death.

It is important to be aware that a diagnosis given by a family member, or even obtained from a death certificate, may be wrong. This is often true in cases of cancer, where ‘stomach’ may mean any part of the bowel, and ‘brain’ may refer to secondary deposits or be used where the primary site has not been identified.

**Polymerase chain reaction and DNA sequencing**

The polymerase chain reaction (PCR) is a fundamental laboratory technique that amplifies targeted sections of the human genome for DNA diagnostic analysis. Almost any tissue can be used to extract DNA for PCR analysis, but most commonly, a sample of peripheral blood is used. PCR is very often used in association with DNA sequencing to determine the exact nucleotide sequence of a specific region of a gene or chromosome. The principles of PCR are shown in Figure 3.12. The technique of DNA sequencing is used for DNA diagnostic analysis in clinical practice. Until recently, most diagnostic DNA laboratories used Sanger sequencing for diagnosis (Fig. 3.13A), but
**Fig. 3.12 (see opposite)** The polymerase chain reaction. The polymerase chain reaction (PCR) involves adding a tiny amount of the patient’s DNA to a reaction containing primers (short oligonucleotides 18–21 bp in length, which bind to the DNA flanking the region of interest) and deoxynucleotide phosphates (dATP, dCTP, dGTP, dTTP), which are used to synthesise new DNA and a heat-stable polymerase. The reaction mix is first heated to 95°C, which causes the double-stranded DNA molecules to separate. The reaction is then cooled to 50–60°C, which allows the primers to bind to the target DNA. The reaction is then heated to 72°C, at which point the polymerase starts making new DNA strands. These cycles are repeated 20–30 times, resulting in exponential amplification of the DNA fragment between the primer sites. The resulting PCR products can then be used for further analysis – most commonly DNA sequencing (see Fig. 3.13).

**Fig. 3.13** DNA sequencing. **A** Sanger sequencing of DNA, which is very widely used in DNA diagnostics. This is performed using PCR-amplified fragments of DNA corresponding to the gene of interest. The sequencing reaction is performed using a combination of dNTP and fluorescently labelled dideoxy dNTP (ddATP, ddCTP, ddGTP, and ddTTP), which become incorporated into the newly synthesised DNA, causing termination of the chain at that point. The reaction products are then subject to capillary electrophoresis and the different-sized fragments are detected by a laser, producing a sequence chromatogram that corresponds to the target DNA sequence. **B** Next-generation sequencing. Samples of patient DNA are fragmented and adapters ligated to each end of the fragment. The sequencing reaction is then performed with primers specific for the adapters, much as described for Sanger sequencing. In next-generation sequencing, however, the reaction products are detected without the need for electrophoresis, and assembled by computer to produce the final sequence read. The absence of electrophoresis allows next-generation sequencing to generate data 100–1000 times faster than Sanger sequencing.
Assessing DNA copy number

For decades, metaphase chromosome analysis by light microscopy has been the mainstay of clinical cytogenetic analysis to detect gain or loss of whole chromosomes or large chromosomal segments (>4 million bp); such anomalies are collectively known as aneuploidy. More recently, whole-genome microarrays have replaced chromosome analysis, allowing rapid and precise detection of segmental gain or loss of DNA throughout the genome (see Box 3.5, p. 56). Microarrays consist of dense grids of short sequences of DNA (probes) that are complementary to known sequences in the genome (Fig. 3.14B). Each probe is fixed at a known position on the array (often printed on to a specially coated glass slide). The patient’s fluorescently labelled DNA sample is hybridised to the array, and results for each probe are read by a laser scanner. This allows a copy number map of the patient’s DNA to be constructed and abnormalities to be identified. Many clinically recognisable syndromes are the result of aneuploidy. The specific phenotype associated with individual deletion syndromes is the result of loss of one copy of several adjacent genes—a contiguous gene syndrome (see Box 3.5). Fluorescent in situ hybridisation (FISH, Fig. 3.14A) can be used to confirm specific deletions or duplications on metaphase chromosomes as a follow-up to microarray analysis.

Non-DNA-based methods of assessment

Although DNA-based diagnostic tools are used in the majority of patients with suspected genetic disease, direct analysis of protein function, such as measurement of specific enzyme activity, can also be used to diagnose single-gene disorders. An example of this is the investigation of myopathy thought to be due to defects in mitochondrial complex 1 proteins (Box 3.7). Complex 1 is made up of at least 36 nuclear-encoded and 7 mitochondrial DNA-encoded subunits, and mutations in any of these subunits can cause the disorder, which makes sequence analysis impractical as a first-line clinical test. Conversely, the biochemical measurement of respiratory chain complex 1 proteins can easily be analysed in muscle biopsies, and this can be diagnostic of a specific mitochondrial cytopathy (see Fig. 3.3, p. 45, and Box 3.1, p. 46).

Genetic testing in pregnancy and pre-implantation genetic testing

Genetic testing may be performed during pregnancy. Invasive tests, such as amniocentesis and chorionic...
villus sampling, are most often carried out to diagnose conditions that result in early infant death or severe disability. Such tests are only offered after careful explanation of the risks involved. Many couples will use the result of such tests to decide about termination of pregnancy. Some indications for testing are listed in Box 3.8; a previous child with a detectable chromosome abnormality or a parent with a chromosome abnormality such as a balanced translocation; a parent or child with a genetic disease for which testing is available; or advanced maternal age and a high-risk serum screening result.

### Ethical issues

Ethical issues often arise with regard to genetic testing of children. For conditions with onset during childhood and for which useful medical interventions are available, it is clearly important to test a child. An example of this is neonatal testing for cystic fibrosis, when early therapy reduces disease progression (p. 680), or in multiple endocrine neoplasia type 2B (MEN 2B), when early thyroidectomy prevents medullary thyroid carcinoma (p. 755). However, testing a healthy child for an adult-onset disorder where no benefit from early intervention exists should be avoided. Instead, the child should be left to make his or her own informed decision as an adult.

### Identifying a disease gene in families

In families with a genetic disease for which the causative gene is unknown, single nucleotide polymorphisms (SNPs) can be used to track or ‘map’ disease genes using a technique called genome-wide linkage analysis. Microarray-based techniques allow more than 500 000 SNPs to be typed in a single experiment, and by comparison of the segregation of patterns of contiguous SNPs (called haplotypes) in affected and unaffected individuals, the ‘locus’ of DNA where the responsible gene resides can be identified. The confidence of association (‘linkage’) with the disease in question is influenced by the number of subjects studied, the strength of the effect of the gene on the disease, and the closeness of the SNP to the disease gene in question. The confidence can be expressed as a LoD (logarithm of the odds) score, which is −log10 of the probability (p value) of linkage; by convention, a LoD score of more than 3 (p < 0.001) is taken to be statistically significant. Once a locus has been identified, more detailed mapping within the locus can be undertaken and the relevant mutation confirmed by sequencing the relevant gene. Over recent years, next-generation sequencing of every exon in the genome (exome sequencing) has been used as an alternative to linkage analysis in identifying disease-causing mutations in families. Typically affected individuals within the family are sequenced and the results compared with unaffected family members and controls from the general population. For a fully penetrant disorder, the disease-causing mutation will be present in affected individuals and not present in unaffected family members or unrelated controls.
Genetic investigation in populations

Genetic screening may be applied to whole populations. The criteria for the use of population screening are well established; they depend on the incidence of specific conditions in individual populations and on whether an intervention is available to ameliorate the effects of the disease. In the UK, examples include screening for phenylketonuria and cystic fibrosis in the newborn, and prenatal screening for neural tube defects and Down’s syndrome in pregnant women (see Box 3.10). Screening for carriers of haemoglobinopathies and Tay–Sachs disease is also carried out in some countries where the incidence of these conditions may be high enough to merit screening the entire population (p. 1031).

Predictive genetic testing

In the absence of symptoms or signs of disease in an individual at risk, a genetic test can be used to determine whether that individual carries the disease-causing mutation. This is known as pre-symptomatic or predictive genetic testing. Predictive tests are usually carried out for adult-onset disorders such as familial cancer syndromes and neurodegenerative disorders such as Huntington’s disease (Box 3.11), or when a positive result in children will affect screening and management, such as in familial polyposis coli (p. 911). However, many complicated ethical issues arise with testing of children and such tests should only be carried out by clinicians experienced in their use.

Whilst a negative predictive test is clearly a favourable outcome for the individual concerned, a positive test may have significant negative consequences. These should have been explained fully in the counselling process (see below), and include employment discrimination and psychological effects. Providing this is done, current evidence suggests that serious psychological sequelae are uncommon.

PRESENTING PROBLEMS IN GENETIC DISEASE

There are many thousands of known single-gene diseases. Individually these are rare, but collectively they are relatively common. This diversity makes clinical genetics a fascinating clinical specialty but it does mean that it is difficult, if not impossible, for any individual clinician to memorise the features associated with all these disorders. It is therefore important to have an awareness of the existence of genetic diseases and some general rules or ‘triggers’ in mind. Although single-gene disorders can present at any age (Box 3.12) and affect any tissue or organ system, they share some general characteristics:

- positive family history
- early age of onset
- multisystem involvement
- no obvious non-genetic explanation.

It is important to recognise any unusual clinical presentation and to consider genetic disease in the context of the clinical findings and the family history. Publicly accessible online catalogues of Mendelian diseases can be useful sources of potential diagnoses.

3.12 Genetic disease and counselling in old age

- Genetic disease: may present for the first time in elderly patients, e.g. Huntington’s disease.
- Family investigation: remains essential in the management of genetic disease presenting in old age and referral to clinical genetics services should be considered.

MAJOR CATEGORIES OF GENETIC DISEASE

It is clearly impossible to discuss all Mendelian disease in this chapter, as there are many thousands of single-gene disorders. However, the major categories of genetic disease that are commonly encountered by clinical geneticists in adult practice are discussed below.

Inborn errors of metabolism

Inborn errors of metabolism (IEM) are caused by mutations that disrupt the normal function of a biochemical pathway. A good example is the glycogen storage diseases (see Box 16.23, p. 450), which are caused by mutations in various genes involved in regulating glucose metabolism. Most IEM are due to autosomal or X-linked recessive loss-of-function mutations in genes encoding specific enzymes or enzymatic co-factors. Knowledge of the biochemical pathway involved means that specific blocks have predictable consequences, including deficiency of the end product and build-up of intermediary compounds. Many hundreds of different IEM have been identified and these disorders have contributed a great deal to our understanding of human biochemistry. Most IEM are very rare and some are restricted to paediatric practice; however, a growing number may now present during adult life and some of these are discussed below.

Intoxicating IEM

A subgroup of IEM, termed ‘intoxicating IEM’, can present as a sudden deterioration in a previously well individual. Such deteriorations are usually precipitated
by physiological stress, such as infection, pregnancy, exercise or changes in diet. The intoxication is due to the build-up of intermediary, water-soluble compounds, which will vary according to the pathway involved. For example, in urea cycle disorders ammonia is the toxic substance, whereas in maple syrup urine disease it is branched-chain amino acids. The intoxication often associated with derangement of the acid-base balance and, if not recognised and treated, will often proceed to multi-organ failure, coma and death. In the porphyrias (Box 16.32, p. 460), the intoxication is caused by a build-up in the metabolites involved in haem synthesis. The diagnosis of these disorders requires specialist biochemical analysis of blood and/or urine. In some disorders, treatment relies on removal of the toxic substance using haemodialysis or chemical conjugation, or prevention of further accumulation by restricting intake of the precursors, such as total protein restriction in urea cycle disorders and avoidance of branched-chain amino-acid intake in maple syrup urine disease. In other disorders, such as the porphyrias, treatment is based on avoiding precipitating factors and supportive care (p. 460).

**Mitochondrial disorders**

Disorders of energy production are the most common type of IEM presenting in adult life, and some of these disorders have been mentioned in the section on mitochondrial function (see Fig. 3.3, p. 45, and Box 3.1, p. 46). The tissues that are most commonly affected in this group of disorders are those with the highest metabolic energy requirements, such as muscle, heart, retina and brain. Therapy in this group of disorders is based on giving antioxidants and co-factors, such as vitamin C and ubiquinone, that can improve the function of the respiratory chain.

**Storage disorders**

Storage disorders involve enzyme deficiency in lysosomal degradation pathways. The clinical consequences depend on the specific enzyme involved. For example, Fabry disease, an X-linked recessive deficiency of α-galactosidase A, results in abdominal pain, episodic diarrhoea, renal failure and angiokeratoma. Niemann–Pick disease type C is caused by autosomal recessive loss-of-function mutations in either the NPC1 or NPC2 gene. These result in lysosomal cholesterol accumulation, causing hepatosplenomegaly, dysphagia, loss of speech, very early dementia, spasticity and dystonia. An increasing number of storage disorders are treatable with enzyme replacement or substrate depletion therapies, making awareness and diagnosis more important. More details of specific disorders are provided in Chapter 16 (Box 16.24, p. 451).

**Neurological disorders**

Progressive neurological deterioration is one of the most common presentations of adult genetic disease. These diseases are mostly autosomal dominant and can be grouped into specific neurological syndromes and early-onset forms of well-known, non-Mendelian clinical entities. In the latter group, the best examples would be early-onset familial forms of dementia, Parkinson’s disease and motor neuron disease. The triplet repeat disorders cause an interesting group of syndromes and have specific features that are dealt with below.

**Huntington’s disease**

Huntington’s disease (HD) is the paradigm of triplet repeat disorders. This condition can present with a movement disorder, weight loss or psychiatric symptoms (depression, addiction, psychosis, dementia), or with a combination of all three. The disease is the result of a [CAG]n triplet repeat expansion mutation in the HD gene on chromosome 4. Since CAG is a codon for glutamine and this mutation is positioned in the ORF, this results in an expansion of a polyglutamine tract in the protein. The mutation probably leads to gain of function, as deletions of the gene do not cause HD. The function of the protein encoded by the HD gene is not fully understood, but expansion of the repeat to above the normal range of 3–35 results in neurological disease. In general, the severity of disease and age at onset are related to the repeat length. In HD, atrophy of the caudate nuclei and the putamen is obvious on magnetic resonance imaging (MRI) of the brain, and in later stages cerebral atrophy is also apparent. There is currently no therapy that will alter the progression of the disease, which will often be the cause of the patient’s death. Within families there is a tendency for disease severity to increase and age at onset to fall due to further expansion of the repeat, a phenomenon known as anticipation. The mutation is more likely to expand through the male germ line than through female meiosis.

**Other triplet repeat disorders**

Other progressive neurological disorders caused by triplet repeat expansion mutations in different genes include several forms of autosomal dominant spinocerebellar ataxias, dentatorubral-pallidolysian atrophy (DRPLA), Machado–Joseph disease and Kennedy disease. These polyglutamine disorders all show intracellular inclusions in affected cells. It is thought that this accumulation may, in itself, be deleterious and is the result of defective protein degradation. Myotonic dystrophy and Friedreich’s ataxia are also triplet repeat expansion disorders but the pathogenetic mechanism associated with these diseases is different, as these repeats do not lie within the coding regions of the affected genes.

**Connective tissue disorders**

Mutations in different types of collagen, fibrillin and elastin make up the majority of connective tissue disorders. The clinical features of these disorders vary, depending on the structural function and tissue distribution of the protein that is mutated. For example, autosomal dominant loss-of-function mutations in the gene encoding elastin cause either supravalvular aortic stenosis, cutis laxa or a combination of both conditions. The most commonly involved systems are:

- **skin** (increased or decreased elasticity, poor wound healing)
- **eyes** (myopia, lens dislocation)
- **blood vessels** (vascular fragility)
- **bones** (osteoporosis, skeletal dysplasia)
- **joints** (hypermobility, dislocation, arthropathy).
Learning disability, dysmorphism and malformations

Congenital global cognitive impairment (also called mental handicap or learning disability) affects about 3% of the population. It is commonly divided into broad categories of mild to moderate (IQ 50–70), moderate to severe (IQ 20–50), and severe to profound (IQ < 20). There are important ‘environmental’ causes of global cognitive impairment, including:

- teratogen exposure during pregnancy (alcohol, anticonvulsants)
- congenital infections (cytomegalovirus, rubella, toxoplasmosis, syphilis)
- the sequelae of prematurity (intraventricular haemorrhage)
- birth injury (hypoxic ischaemic encephalopathy).

Genetic disorders contribute very significantly to the aetiology of global cognitive impairment. Given the complexity of brain development, it is not surprising that global cognitive impairment shows extreme locus heterogeneity. The three most important groups of disorder are reviewed below.

Chromosome disorders

Any significant gain or loss of autosomal chromosomal material (known as aneuploidy) usually results in learning disability and other phenotypic abnormalities (see Fig. 3.11, p. 57). Down’s syndrome is the most frequently found and best known of these disorders, and is caused by an increased dosage of genes on chromosome 21. Most cases of Down’s syndrome are due to a numerical chromosome abnormality with trisomy of chromosome 21, e.g. 47,XX,+21 or 47,XY,+21. The clinical features are:

- globally delayed development
- characteristic facial appearance
- a significant risk of specific malformations (atrioventricular septal defect, duodenal atresia)
- a predisposition to several late-onset disorders, including hypothyroidism, acute leukaealias and Alzheimer’s disease.

Recent surveys have shown that DNA microarray analysis can identify causative structural chromosome abnormalities in 10–25% of cases of significant learning disability. These deletions and duplications are mostly de novo and unique. An interest group of recurrent deletions and duplication caused by non-allelic homologous recombination events has been mentioned above. These result in specific microdeletion or microduplication syndromes, such as:

- velocardiofacial syndrome due to deletion of 22q11.2 (learning disability, malformations of the cardiac outflow tract, cleft palate, distinctive facial appearance and immune disorders)
- Williams’s syndrome due to deletion of 7q11.23 (learning disability, supravalvular aortic stenosis and mild cutis laxa as a result of deletion of the elastin gene, distinctive facial appearance and over-friendly, chatty personality).

Dysmorphic syndromes

There are several thousand different dysmorphic syndromes; all are rare but they are characterised by the occurrence of cognitive impairment, malformations and a distinctive facial appearance – or ‘gestalt’ – associated with various other clinical features. Making the correct diagnosis is important, as it has profound implications on immediate patient management, detection of future complications and assessment of recurrence risks in the family. Clinical examination remains the mainstay of diagnosis and the patient often needs to be evaluated by a clinician who specialises in the diagnosis of these syndromes. The differential diagnosis in dysmorphic syndromes is often very wide and this has resulted in computer-aided diagnosis becoming an established clinical tool. Dysmorphology databases such as POSSUM and LMD have been established that are curated catalogues of the many thousands of known syndrome entities; they can be searched to identify possible explanations of unusual combinations of clinical features. The clinical diagnosis may then be confirmed by specific genetic investigations, as the genetic basis of a wide range of dysmorphic syndromes has been identified.

X-linked mental handicap

X-linked mental handicap (XLMH) accounts for approximately 10% of cases of moderate to severe learning disability. There are over 100 genes on the X chromosome that can cause learning disability but the most common disorder is fragile X syndrome, characterised by a distinctive facial appearance, attention deficit, joint hypermobility, macro-orchism (increased testicular size) and a non-staining gap on the X chromosome on chromosome analysis. Fragile X is caused by a triplet repeat expansion mutation but of a different type from the polyglutamine repeat disorders mentioned above. The repeat in fragile X syndrome is not in the coding region and is a [CGG]n expansion (see Box 3.4, p. 56). Methylation of the expanded repeat results in silencing of a specific gene called FMR1, which encodes an RNA-binding protein.

De novo mutations

Next-generation sequencing technology has made possible trio-based, whole-exome sequencing, in which the affected individual and both of their parents are analysed. It has recently become clear that de novo mutations in the coding regions of one of the many genes that are involved in normal brain development are collectively responsible for severe intellectual disability in at least 25% of affected patients. Trio-exome sequencing is thus likely to become a first-line diagnostic test for such cases in the near future.

Familial cancer syndromes

Most cancers are not inherited but occur as the result of an accumulation of somatic mutations, as discussed previously in this chapter. However, it has been recognised for many decades that some families are prone to one or more specific types of cancer. Affected individuals tend to present with tumours at an early age and are more likely to have multiple primary foci of carcinogenesis.

Retinoblastoma

Patients with autosomal dominant familial retinoblastoma have an inherited mutation in one copy of the RB
gene, which is a tumour suppressor. This strongly predisposes individuals to the formation of retinoblastoma in one or both eyes. It is possible for more than one primary tumour to form in the same eye and for retinoblastoma to occur in the pineal gland. From a clinical perspective, it is important to screen the eyes and pineal gland of such individuals regularly so that tumours can be treated early and sight preserved. This gene is widely expressed and it is not clear why the retina is the main site of oncogenesis in this syndrome. An increased incidence of osteogenic sarcoma is also seen in affected individuals.

Familial adenomatous polyposis coli
Familial adenomatous polyposis coli (FAP) is an autosomal dominant condition due to inactivation mutations in the FAP tumour suppressor gene on 5q. The gene product is thought to modulate a specific signalling cascade (Wnt signalling) that regulates cell proliferation. Mutation carriers usually develop many thousands of intestinal polyps in their second and third decades and have a very high risk of malignant change in the colon. Prophylactic colectomy in the third decade is necessary in most cases. Regular screening for polyps in the upper gastrointestinal tract is also recommended.

Li–Fraumeni syndrome
Heterozygous loss-of-function mutations in the gene encoding p53 cause Li–Fraumeni syndrome. Families with this condition have a very significant increased predisposition to early-onset leukaemias, sarcomas, and breast and brain malignancies. Screening for pre-symptomatic tumours in this condition is very difficult and of unproven benefit, as almost any tissue can be affected.

Hereditary non-polyposis colorectal cancer
Hereditary non-polyposis colorectal cancer (HNPCC) is an autosomal dominant disorder that presents with early-onset familial colon cancer, particularly affecting the proximal colon. Other cancers, such as endometrial cancer, are often observed in affected families. This disorder shows marked locus heterogeneity, as mutations can occur in several different genes encoding proteins involved in DNA mismatch repair.

Familial breast cancer
Familial breast cancer is an autosomal dominant disorder that is most often due to mutations in genes encoding either BRCA1 or BRCA2. Both of these proteins are involved in DNA repair. Individuals who carry a BRCA1 or BRCA2 mutation are at high risk of early-onset breast and ovarian tumours, and require regular screening for both of these conditions. Because of the very high risk of cancer, many women who carry these mutations elect to have prophylactic bilateral mastectomy and oophorectomy in the absence of a detectable tumour.

Xeroderma pigmentosum
Xeroderma pigmentosum (XP) is the name given to a group of rare disorders in which there are autosomal recessive defects in DNA repair genes that deal primarily with the effects of non-ionising radiation. The skin is particularly involved, and affected patients develop skin cancers with increased frequency.

**GENETIC COUNSELLING**

Genetic counselling provides information about the medical and family implications of a specific disease in a clear and non-directive manner. Such counselling aims to help individuals make informed decisions about planning a family, taking part in screening programmes and accepting prophylactic therapies. Genetic counselling may be provided by a medical geneticist, a specialist nurse, or a clinician with particular skills in this area, such as an obstetrician or paediatrician (Box 3.13). Perception of genetic risks clearly depends on perceived hazard. For example, a 5% (or 1:20) risk of genetic disease may be perceived as low if the disease is treatable, but unacceptably high if not.

Specific problems encountered in genetic counselling include:
- accurate assessment of genetic risk
- identification of children at risk of genetic disorders
- the increase in genetic risks associated with consanguinity
- non-paternity as an incidental finding in DNA.

### 3.13 Clinical genetics services

<table>
<thead>
<tr>
<th>Component</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical geneticist</td>
<td>Diagnosis and management of genetic disease, assessment of genetic risk, managing screening programmes, interpretation of genetic test results. Subspecialties include prenatal genetics, dysmorphology (syndrome identification), cancer genetics</td>
</tr>
<tr>
<td>Genetic counsellor</td>
<td>Assessing genetic risk, provision of genetic counselling (providing accurate risk information in a comprehensible format), predictive testing for genetic disease and provision of information and support</td>
</tr>
<tr>
<td>DNA diagnostic laboratory</td>
<td>Identifying and reporting disease-causing mutations in validated disease genes. Some laboratories also provide linkage analysis to track diseases in families. Laboratories often work in a consortium, as so many different disease genes have now been identified</td>
</tr>
<tr>
<td>Cytogenetics laboratory</td>
<td>Identifying pathogenic numerical and structural chromosome anomalies in prenatal, postnatal and oncology samples</td>
</tr>
<tr>
<td>Biochemical genetics laboratory</td>
<td>Metabolite and enzymatic-based diagnosis of IEM. Metabolite-based monitoring of treatment of IEM</td>
</tr>
<tr>
<td>Newborn screening laboratory</td>
<td>Provision of population-based newborn screening, e.g. PKU, cystic fibrosis, etc.</td>
</tr>
</tbody>
</table>

(IEM = inborn errors of metabolism; PKU = phenylketonuria)
Genetic tests are increasingly used for the diagnosis and prediction of Mendelian disease in a medical context, and such skills will become increasingly important for many clinicians.

Genetic risk is often calculated using Bayes’ theorem (Box 1.6, p. 6), which takes prior risk into account to calculate future risk. A simple Bayesian calculation is illustrated here. Consider a woman who is at risk of being a carrier of an X-linked recessive disease. Her grandfather and brother are affected, which makes her mother an obligate gene carrier. Her risk of being a carrier is therefore 50%. However, she has two unaffected sons. This information can be used to modify her risk. The prior probability that she is a carrier is 1:2 and that she is not a carrier also 1:2. The conditional probability that she would have two normal sons if she were a carrier is $1/2 \times 1/2$, i.e. 1/4. If she were not a carrier, the probability of having normal sons is 1. From this, the joint probability for each outcome can be calculated (the prior risk x the conditional risk): $1/2 \times 1/4$ (1/8) for being a carrier and $1/2 \times 1$ (1/2) for not being a carrier. The final risk, or relative probability, for each outcome can then be obtained by dividing the joint probability for that outcome by the sum of the joint probabilities. The probability that she is a carrier is therefore $1/8(1/8 + 1/2) = 1/5$ (20%).

### GENETICS OF COMMON DISEASES

Many common disorders, such as diabetes, atherosclerosis, hypertension, cancer, osteoarthritis, inflammatory bowel disease and osteoporosis, have an important genetic component but are not caused by a single mutation. Techniques are now available both to measure the contribution and to identify genes with significant effects. This means that the result of genetic testing is beginning to have an impact on diagnosis, prognosis and therapy for common diseases, and this trend is likely to expand significantly in the years to come. Some of the most useful approaches to clinical interpretation of the genetic aspects of common disorders are outlined below.

### Measuring the genetic contribution to complex disease

Genetic contributions to complex disease can be detected and quantified by twin studies and/or by analysing familial clustering. Twin studies use the difference in disease concordance between monozygotic (MZ) and dizygotic (DZ) twins to calculate genetic contribution. MZ twins are genetically identical, whereas DZ twins, like all siblings, are identical for only about 50% of their genetic variation. However, both MZ and DZ twins share an almost identical intrauterine environment and similar postnatal environment. Thus, any evidence of a higher concordance of the disease in MZ compared to DZ twins is assumed to be evidence of genetic contribution. Many common diseases and quantitative traits, such as height, weight, blood pressure and bone mineral density, show higher concordance rates in MZ twins compared to DZ twins. Genetic contributions to common diseases can also be assessed by studying the incidence of the disease in first-degree relatives of affected individuals, as compared with the general population (Fig. 3.15). The difference in incidence is used to calculate a disease risk, which is measured by the $\lambda_s$ value (Box 3.14).

### 3.14 Risk to siblings of affected patients for common polygenic diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>$\lambda_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1 diabetes mellitus</td>
<td>15</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>10–20</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>20–40</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>10</td>
</tr>
<tr>
<td>Ischaemic heart disease</td>
<td>4–12</td>
</tr>
</tbody>
</table>

Fig. 3.15 The spectrum of genetic disease: how the genotype influences the phenotype. A particular characteristic or disease in an individual may be due to a specific genetic abnormality (monogenic disease) or may reflect several predisposing genes (polygenic disease). In each case, environmental factors may further influence the phenotype; in their absence, genetic factors alone may be insufficient to allow the disease to develop, resulting in non-penetrance or reduced penetrance (see text).
Genetic testing in complex disease

Most common diseases are determined by interactions between a number of genes and the environment. In this situation, the genetic contribution to disease is termed polygenic. Until recently, very little progress had been made in identifying the genetic variants that predispose to common diseases, but this has been changed by the advent of genome-wide association studies (GWAS). A GWAS typically involves genotyping many (> 500,000) genetic markers spread across the genome in a large group of individuals with the disease and controls. By comparing the genotypes in cases and controls, it is possible to identify regions of the genome and candidate genes that contribute to the disease under study. Some of the candidate genes for common diseases identified by this approach are listed in Box 3.15.

Pharmacogenomics

Pharmacogenomics is the science of dissecting the genetic determinants of drug kinetics and effects using information from the human genome. For more than 50 years, it has been appreciated that polymorphic mutations within genes can affect individual responses to some drugs, such as loss-of-function mutations in CYP2D6 causing hypersensitivity to debrisoquine, an adrenergic-blocking medication formerly used for the treatment of hypertension, in 3% of the population. This gene is part of a large family of highly polymorphic genes encoding cytochrome P450 proteins, mostly expressed in the liver, which determine the metabolism of a host of specific drugs. Polymorphisms in the CYP2D6 gene also determine codeine activation, while those in the CYP2C9 gene affect warfarin inactivation. Polymorphisms in these and other drug metabolic genes determine the persistence of drugs and, therefore, should provide information about dosages and toxicity. At the present time, genetic testing for assessment of drug response is seldom used routinely, but in the future it may be possible to predict the best specific drugs and dosages for individual patients based on genetic profiling: so-called ‘personalised medicine’. An example is the enzyme thiopurine methyltransferase (TPMT), which catabolises azathioprine, a drug that is used in the treatment of autoimmune diseases and in cancer chemotherapy. Genetic screening for polymorphic variants of TPMT can be useful in identifying patients who have increased sensitivity to the effects of azathioprine and who can be treated with lower doses than normal.

RESEARCH FRONTIERS IN MOLECULAR MEDICINE

Gene therapy

Replacing or repairing mutated genes (gene therapy) is very difficult in humans. Retroviral-mediated ex vivo replacement of the defective gene in bone marrow cells for the treatment of severe combined immune deficiency syndrome (p. 80) has been partially successful. There have been two major problems with the clinical trials of virally delivered gene therapy conducted to date:

- The random integration of the retroviral DNA (which contains the replacement gene) into the genome has caused leukaemia in some treated children via activation of proto-oncogenes.
- A severe immune response to the viral vector may be induced. It has not yet been possible to use non-viral means to introduce sufficient numbers of copies of replacement genes to produce significant biological effects.

Other therapies for genetic disease include PTC124, a compound that can ‘force’ cells to read through a mutation that results in a premature termination codon in an ORF with the aim of producing a near-normal protein product. This therapeutic approach could be applied to any genetic disease caused by nonsense mutations.

Induced pluripotent stem cells and regenerative medicine

Adult stem cell therapy has been in wide use for decades in the form of haematopoietic stem cell transplantation. The identification of adult stem cells for other tissues, coupled with the ability to purify and maintain such cells in vitro, now offers exciting therapeutic potential for other diseases. Many different adult cell types can be transdifferentiated to form cells termed induced pluripotent stem cells (iPScells) with almost all the characteristics of embryonic stem cells. In mammals, iPScells can be used to regenerate various tissues such as the
heart and brain. They have great potential both to
develop tissue models of human disease and for regen-
erative medicine. In mammalian model species, such
cells can be taken and used to regenerate differentiated
tissue cells, such as in heart and brain.

Pathway medicine

The ability to manipulate pathways that have been altered in genetic disease has tremendous therapeutic
potential for Mendelian disease, but a firm understand-
ing of both disease pathogenesis and drug action at a
biochemical level is required. An exciting example of
this has been the discovery that the vascular pathology
associated with Marfan’s syndrome is due to the defect-
tive fibrillin molecules causing up-regulation of trans-
forming growth factor (TGF)-β signalling in the vessel
wall. Losartan is an antihypertensive drug that is mar-
keted as an angiotensin II receptor antagonist. However,
it also acts as a partial antagonist of TGF-β signalling
and is effective in preventing aortic dilatation in a mouse
model of Marfan’s syndrome, showing promising effects
in early human clinical trials.

Further information

Books and journal articles

Read A, Donnai D. New clinical genetics. 2nd edn. Banbury:
Scion; 2010.
Strachan T, Read A. Human molecular genetics. 3rd edn.

Websites

www.bshg.org.uk British Society for Human Genetics; has
report on genetic testing of children.
www.ensembl.org Annotated genome databases from multiple
organisms.
www.genome.ucsc.edu Excellent source of genomic information.
(OMIM).
Immunological factors in disease

**Functional anatomy and physiology of the immune system** 72
- The innate immune system 72
- The adaptive immune system 76

**Immune deficiency** 78
- Presenting problems in immune deficiency 79
- Recurrent infections 79
- Primary phagocyte deficiencies 79
- Complement pathway deficiencies 79
- Primary deficiencies of the adaptive immune system 80
- Secondary immune deficiencies 82

**The inflammatory response** 82
- Physiology and pathology of inflammation 82
- Investigations in inflammation 83
- Presenting problems in inflammation 85
- Unexplained raised ESR 85
- Periodic fever syndromes 85
- Amyloidosis 86

**Autoimmune disease** 86
- Pathophysiology of autoimmunity 86
- Investigations in autoimmunity 88

**Allergy** 89
- Pathology of allergy 89
- Presenting problems in allergy 90
  - A general approach to the allergic patient 90
  - Anaphylaxis 91
  - Angioedema 93
- Specific allergies 94
- C1 inhibitor deficiency 94

**Transplantation immunology** 94
- Transplant rejection 94
- Complications of transplant immunosuppression 95
- Organ donation 96
The immune system has evolved to protect the host from pathogens while minimising damage to self tissue. Despite the ancient observation that recovery from some diseases results in protection against that condition, the existence of the immune system as a functional entity was not recognised until the end of the 19th century. More recently, it has become clear that the immune system not only protects against infection, but also influences healing and governs the responses that can lead to autoimmune diseases. Dysfunction or deficiency of the immune response leads to a wide variety of diseases, involving every organ system in the body.

The aim of this chapter is to provide a general understanding of immunology and how it contributes to human disease. A review of the key components of the immune response is followed by five sections that illustrate the clinical presentation of the most common forms of immune dysfunction. Clinical immunologists are usually involved in managing patients with allergy and immune deficiency. More detailed discussion of individual conditions can be found in the relevant organ-specific chapters of this book.

FUNCTIONAL ANATOMY AND PHYSIOLOGY OF THE IMMUNE SYSTEM

The immune system consists of an intricately linked network of cells, proteins and lymphoid organs that are strategically placed to ensure maximal protection against infection. Immune defences are normally categorised into the innate immune response, which provides immediate protection against an invading pathogen, and the adaptive or acquired immune response, which takes more time to develop but confers exquisite specificity and long-lasting protection.

The innate immune system

Innate defences against infection include anatomical barriers, phagocytic cells, soluble molecules, such as complement and acute phase proteins, and natural killer cells. The innate immune system recognises generic microbial structures present on non-mammalian tissue and can be mobilised within minutes. A specific stimulus will elicit essentially identical responses in different individuals (in contrast with antibody and T-cell responses, which vary greatly between individuals).

Constitutive barriers to infection

The tightly packed, highly keratinised cells of the skin constantly undergo renewal and replacement, which physically limits colonisation by microorganisms. Microbial growth is inhibited by physiological factors, such as low pH and low oxygen tension, and sebaceous glands secrete hydrophobic oils that further repel water and microorganisms. Sweat also contains lysozyme, an enzyme that destroys the structural integrity of bacterial cell walls; ammonia, which has antibacterial properties; and several antimicrobial peptides such as defensins. Similarly, the mucous membranes of the respiratory, gastrointestinal and genitourinary tract provide a constitutive barrier to infection. Secreted mucus acts as a physical barrier to trap invading pathogens, and immunoglobulin A (IgA) prevents bacteria and viruses attaching to and penetrating epithelial cells. As in the skin, lysozyme and antimicrobial peptides within mucosal membranes can directly kill invading pathogens, and additionally lactoferrin acts to starve invading bacteria of iron. Within the respiratory tract, cilia directly trap pathogens and contribute to removal of mucus, assisted by physical manoeuvres, such as sneezing and coughing. In the gastrointestinal tract, hydrochloric acid and salivary amylase chemically destroy bacteria, while normal peristalsis and induced vomiting or diarrhoea assist clearance of invading organisms.

Endogenous commensal bacteria provide an additional constitutive defence against infection (p. 136). They compete with pathogenic microorganisms for space and nutrients, and produce fatty acids and bacteriocidins that inhibit the growth of many pathogens. In addition, commensal bacteria help to shape the immune response by inducing specific regulatory T cells within the intestine (p. 78).

These constitutive barriers are highly effective, but if external defences are breached by a wound or pathogenic organism, the specific soluble proteins and cells of the innate immune system are activated.

Phagocytes

Phagocytes (‘eating cells’) are specialised cells which ingest and kill microorganisms, scavenge cellular and infectious debris, and produce inflammatory molecules which regulate other components of the immune system. They include neutrophils, monocytes and macrophages, and are particularly important for defence against bacterial and fungal infections.

Phagocytes express a wide range of surface receptors that allow them to identify microorganisms. These pattern recognition receptors include the Toll-like receptors, NOD (nucleotide-oligomerisation domain protein)-like receptors and mannose receptors. They recognise generic molecular motifs not present on mammalian cells, including bacterial cell wall components, bacterial DNA and viral double-stranded RNA. While phagocytes can recognise microorganisms through pattern recognition receptors alone, engulfment of microorganisms is greatly enhanced by opsonisation. Opsonins include acute phase proteins such as C-reactive protein (CRP), antibodies and complement. They bind both to the pathogen and to phagocyte receptors, acting as a bridge between the two to facilitate phagocytosis (Fig. 4.1).

Neutrophils

Neutrophils, also known as polymorphonuclear leucocytes, are derived from the bone marrow (Fig. 4.2). They are short-lived cells with a half-life of 6 hours in the blood stream, and are produced at the rate of approximately 10¹³ cells daily. Their functions are to kill microorganisms directly, facilitate the rapid transit of cells through tissues, and non-specifically amplify the immune response. This is mediated by enzymes contained in granules which also provide an intracellular milieu for the killing and degradation of microorganisms.

The two main types of granule are primary or azurophil granules, and the more numerous secondary or specific granules. Primary granules contain myeloperoxidase and other enzymes important for
protected compartment, killing of the organism occurs through a combination of oxidative and non-oxidative killing. Oxidative killing, also known as the respiratory burst, is mediated by the NADPH (nicotinamide adenine dinucleotide phosphate) oxidase enzyme complex. This converts oxygen into reactive oxygen species such as hydrogen peroxide and superoxide that are lethal to microorganisms. When combined with myeloperoxidase, hypochlorous ions (HOCl$, \text{analogous to bleach}$) are produced, which are highly effective oxidants and antimicrobial agents. Non-oxidative (oxygen-independent) killing occurs through the release of bactericidal enzymes into the phagolysosome. Each enzyme has a distinct antimicrobial spectrum, providing broad coverage against bacteria and fungi.

The process of phagocytosis depletes neutrophil glycogen reserves and is followed by neutrophil cell death. As the cells die, their contents are released and lysosomal enzymes degrade collagen and other components.
of the interstitium, causing liquefaction of closely adjacent tissue. The accumulation of dead and dying neutrophils results in the formation of pus, which, if extensive, may result in abscess formation.

**Monocytes and macrophages**

Monocytes are the precursors of tissue macrophages. They are produced in the bone marrow and constitute about 5% of leucocytes in the circulation. From the blood stream, they migrate to peripheral tissues, where they differentiate into tissue macrophages and reside for long periods. Specialised populations of tissue macrophages include Kupffer cells in the liver, alveolar macrophages in the lung, mesangial cells in the kidney, and microglial cells in the brain. Macrophages, like neutrophils, are capable of phagocytosis and killing of microorganisms but also play an important role in the amplification and regulation of the inflammatory response (Box 4.1). They are particularly important in tissue surveillance, monitoring their immediate surroundings for signs of tissue damage or invading organisms.

**Dendritic cells**

Dendritic cells are specialised antigen-presenting cells which are prevalent in tissues in contact with the external environment, such as the skin and mucosa. They can also be found in an immature state in the blood. They sample the environment for foreign particles, and once activated, carry microbial antigens to regional lymph nodes, where they interact with T cells and B cells to initiate and shape the adaptive immune response.

**Cytokines**

Cytokines are small soluble proteins that act as multi-purpose chemical messengers. Examples are listed in Box 4.2. They are produced by cells involved in immune responses and by stromal tissue. More than 100 cytokines have been described, with overlapping, complex roles in intercellular communication. Their clinical importance is demonstrated by the efficacy of ‘biological’ therapies (often abbreviated to ‘biologics’) that target specific cytokines (pp. 1102 and 18).

**Complement**

The complement system is a group of more than 20 tightly regulated, functionally linked proteins that act to promote inflammation and eliminate invading pathogens. Complement proteins are produced in the liver and are present in the circulation as inactive molecules. When triggered, they enzymatically activate other proteins in a rapidly amplified biological cascade analogous to the coagulation cascade (p. 995). There are three mechanisms by which the complement cascade may be triggered (Fig. 4.3):
Mast cells and basophils

Mast cells and basophils are bone marrow-derived cells which play a central role in allergic disorders. Mast cells reside predominantly in tissues exposed to the external environment, such as the skin and gut, while basophils are located in the circulation and are recruited into tissues in response to inflammation. Both contain large cytoplasmic granules which contain preformed vasoactive substances such as histamine (see Fig. 4.9, p. 89).

Mast cells and basophils express IgE receptors on their cell surface (see Fig. 4.5). On encounter with specific antigen, the cell is triggered to release preformed mediators and synthesise additional mediators, including leukotrienes, prostaglandins and cytokines. These trigger an inflammatory cascade which increases local blood flow and vascular permeability, stimulates smooth muscle contraction, and increases secretion at mucosal surfaces.

Natural killer cells

Natural killer (NK) cells are large granular lymphocytes which play a major role in defence against tumours and viruses. They exhibit features of both the adaptive and innate immune systems: they are morphologically similar to lymphocytes and recognise similar ligands, but they are not antigen-specific and cannot generate immunological memory.

NK cells express a variety of cell surface receptors. Some recognise stress signals, while others recognise the absence of human leucocyte antigen (HLA) molecules on cell surfaces (down-regulation of HLA molecules by viruses and tumour cells is an important mechanism by which they evade T lymphocytes). NK cells can also be activated by binding of antigen-antibody complexes to surface receptors. This physically links the NK cell to its target in a manner analogous to opsonisation, and is known as antibody-dependent cellular cytotoxicity (ADCC).

Activated NK cells can kill their targets in various ways. Pore-forming proteins, such as perforin, induce
direct cell lysis, while granzymes are proteolytic enzymes which stimulate apoptosis. In addition, NK cells produce a variety of cytokines, such as tumour necrosis factor (TNF)-α and interferon-γ (IFN-γ), which have direct antiviral and antitumour effects.

**The adaptive immune system**

If the innate immune system fails to provide effective protection against an invading pathogen, the adaptive immune system (Fig. 4.4) is mobilised. This has three key characteristics:

- It has exquisite specificity and is able to discriminate between very small differences in molecular structure.
- It is highly adaptive and can respond to an unlimited number of molecules.
- It possesses immunological memory, such that subsequent encounters with a particular antigen produce a more effective immune response than the first encounter.

There are two major arms of the adaptive immune response: humoral immunity involves antibodies produced by B lymphocytes; cellular immunity is mediated by T lymphocytes, which release cytokines and kill immune targets. These interact closely with each other and with the innate immune system, to maximise the effectiveness of the response.

**Lymphoid organs**

- **Primary lymphoid organs.** The primary lymphoid organs are involved in lymphocyte development. They include the bone marrow, where both T and B lymphocytes are derived from haematopoietic stem cells (p. 993) and where B lymphocytes also mature, and the thymus, where T lymphocytes mature.
- **Secondary lymphoid organs.** After maturation, lymphocytes migrate to the secondary lymphoid organs. These include the spleen, lymph nodes and mucosa-associated lymphoid tissue. These organs trap and concentrate foreign substances, and are the major sites of interaction between naïve lymphocytes and microorganisms.

**The thymus**

The thymus is a bilobed structure organised into cortical and medullary areas. The cortex is densely populated with immature T cells, which migrate to the medulla to undergo selection and maturation. The thymus is most active in the fetal and neonatal period, and involutes after puberty. Failure of thymic development is associated with profound T-cell immune deficiency (p. 80), but surgical removal of the thymus in childhood (usually in the context of major cardiac surgery) is not associated with significant immune dysfunction.

**The spleen**

The spleen is the largest of the secondary lymphoid organs. It is highly effective at filtering blood and is an important site of phagocytosis of senescent erythrocytes, bacteria, immune complexes and other debris. It is also a major site of antibody synthesis. It is particularly important for defence against encapsulated bacteria, and asplenic individuals are at risk of overwhelming *Streptococcus pneumoniae* and *H. influenzae* infection (see Box 24.40, p. 1028).

**Lymph nodes and mucosa-associated lymphoid tissue**

Lymph nodes are positioned to maximise exposure to lymph draining from sites of external contact. Their structure is highly organised, as shown in Figure 4.4B.

More diffuse unencapsulated lymphoid cells and follicles are also present on mucosal surfaces: for example, in Peyer’s patches in the small intestine.

---

**Fig. 4.4 Anatomy of the adaptive immune system.**

A** Macroanatomy

B Anatomy of a lymph node.

Streptococcus pneumoniae and H. influenzae infection (see Box 24.40, p. 1028).
**Lymphatics**

Lymphoid tissues are physically connected by a network of lymphatics, which has three major functions: it provides access to lymph nodes, returns interstitial fluid to the venous system, and transports fat from the small intestine to the blood stream (see Fig. 16.14, p. 452). The lymphatics begin as blind-ending capillaries, which come together to form lymphatic ducts. These enter and then leave regional lymph nodes as afferent and efferent ducts respectively. They eventually coalesce and drain into the thoracic duct and thence into the left subclavian vein. Lymphatics may be either deep or superficial, and, in general, follow the distribution of major blood vessels.

**Humoral immunity**

*B lymphocytes*

These specialised cells arise in the bone marrow. Mature B lymphocytes (also known as B cells) are found in bone marrow, lymphoid tissue, spleen and, to a lesser extent, the blood stream. They express a unique immunoglobulin receptor on their cell surface (the B-cell receptor), which binds to soluble antigen. Encounters with antigen usually occur within lymph nodes, where, if provided with appropriate signals from nearby T lymphocytes, stimulated antigen-specific B cells respond by proliferating rapidly in a process known as clonal expansion. This is accompanied by a highly complex series of genetic rearrangements, which generates B-cell populations that express receptors with greater affinity for antigen than the original. These cells differentiate into either long-lived memory cells, which reside in the lymph nodes, or plasma cells, which produce antibody.

**Immunoglobulins**

Immunoglobulins (Ig) are soluble proteins made up of two heavy and two light chains (Fig. 4.5). The heavy chain determines the antibody class or isotype, i.e. IgG, IgA, IgM, IgE or IgD. Subclasses of IgG and IgA also occur. The antigen is recognised by the antigen-binding regions (F_{ab}) of both heavy and light chains, while the consequences of antibody-binding are determined by the constant region of the heavy chain (F_{c}) (Box 4.3).

Antibodies can initiate a number of different actions. They facilitate phagocytosis by acting as opsonins (see Fig. 4.1), and can also facilitate cell killing by cytotoxic cells (ADCC, p. 75). Binding of antibodies to antigen can trigger activation of the classical complement pathway (see Fig. 4.3). In addition, antibodies may act directly to neutralise the biological activity of toxins. This is a particularly important feature of IgA antibodies, which act predominantly at mucosal surfaces.

The humoral immune response is characterised by immunological memory: that is, the antibody response to successive exposures to antigen is qualitatively and quantitatively different from that on first exposure. When a previously unstimulated (naive) B lymphocyte is activated by antigen, the first antibody to be produced is IgM, which appears in the serum after 5–10 days. Depending on additional stimuli provided by T lymphocytes, other antibody classes (IgG, IgA and IgE) are produced 1–2 weeks later. If, some time later, a memory B cell is re-exposed to antigen, the lag time between antigen exposure and the production of antibody is decreased (to 2–3 days), the amount of antibody produced is increased, and the response is dominated by IgG antibodies of high affinity. Furthermore, in contrast

<table>
<thead>
<tr>
<th>Antibody/proPERTIES</th>
<th>Concentration in adult serum</th>
<th>Complement activation*</th>
<th>Opsonisation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IgG</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 subclasses: IgG1, IgG2, IgG3, IgG4</td>
<td>8.0–16.0 g/L</td>
<td>IgG1 +++</td>
<td>IgG1 ++</td>
</tr>
<tr>
<td>Distributed equally between blood and extracellular fluid, and transported across placenta</td>
<td>IgG2 +</td>
<td>IgG3 +++</td>
<td></td>
</tr>
<tr>
<td>IgG2 is the predominant antibody produced against polysaccharides</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IgA</strong></td>
<td>1.5–4.0 g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 subclasses: IgA1, IgA2</td>
<td>Highly effective at neutralising toxins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Particularly important at mucosal surfaces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IgM</strong></td>
<td>0.5–2.0 g/L</td>
<td>++++</td>
<td></td>
</tr>
<tr>
<td>Highly effective at agglutinating pathogens</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IgE</strong></td>
<td>0.003–0.04 g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mostly bound to mast cells, basophils and eosinophils</td>
<td>Important in allergic disease and defence against parasite infection</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IgD</strong></td>
<td>Function unknown</td>
<td>Not detected</td>
<td></td>
</tr>
</tbody>
</table>

*Refers to activation of the classical complement pathway, also called ‘complement fixation’.
IMMUNOLOGICAL FACTORS IN DISEASE

1. Cellular immunity

T lymphocytes (also known as T cells) mediate cellular immunity and are important for defence against viruses, fungi and intracellular bacteria. They also play an important immunoregulatory role, orchestrating and regulating the responses of other components of the immune system. T-lymphocyte precursors arise in bone marrow and are exported to the thymus while still immature (see Fig. 4.6 below). Within the thymus, each cell expresses a T-cell receptor with a unique specificity. These cells undergo a process of stringent selection to ensure that autoreactive T cells are deleted. Mature T lymphocytes leave the thymus and expand to populate secondary lymphoid tissues. HLA molecules exhibit extreme polymorphism; as each HLA molecule has the capacity to process and present antigens, specialised antigen-presenting cells include dendritic cells, macrophages and B lymphocytes. HLA molecules exhibit extreme polymorphism; as each HLA molecule has the capacity to present a subtly different peptide repertoire to T lymphocytes, this ensures enormous diversity in recognition of antigens within the population.

T lymphocytes can be segregated into two subgroups on the basis of function and recognition of HLA molecules. These are designated CD4+ and CD8+ T cells, according to the ‘cluster of differentiation’ (CD) antigen expressed on their cell surface. CD8+ T cells recognise antigens in association with HLA class I molecules (HLA-A, HLA-B, HLA-C). They kill infected cells directly through the production of pore-forming molecules such as perforin, or by triggering apoptosis of the target cell, and are particularly important in defence against viral infection. CD4+ T cells recognise peptides presented on HLA class II molecules (HLA-DR, HLA-DP and HLA-DQ) and have mainly immunoregulatory functions. They produce cytokines and provide co-stimulatory signals that support the activation of CD8+ T lymphocytes and assist the production of mature antibody by B cells. In addition, their close interaction with phagocytes determines cytokine production by both cell types.

CD4+ lymphocytes can be further subdivided into subsets on the basis of the cytokines they produce:

- Typically, Th1 cells produce IL-2, IFN-γ and TNF-α, and support the development of delayed type hypersensitivity responses (p. 87).
- Th2 cells typically secrete IL-4, IL-5 and IL-10, and promote allergic responses (p. 89).
- A further subset of specialised CD4+ lymphocytes known as regulatory cells are important in immune regulation of other cells and the prevention of autoimmune disease.

4. Immune deficiency

The consequences of deficiencies of the immune system include recurrent infections, autoimmunity and susceptibility to malignancy. Immune deficiency may arise through intrinsic defects in immune function, but is much more commonly due to secondary causes, including infection, drug therapy, malignancy and ageing. This chapter gives an overview of the rare primary immune deficiencies. More than a hundred genetically determined deficiencies have been described, most of which present in childhood or adolescence. The clinical manifestations are dictated by the component of the immune system involved (Box 4.4), but there is considerable overlap and redundancy in the immune

<table>
<thead>
<tr>
<th>4.4 Immune deficiencies and common patterns of infection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phagocyte deficiency</strong></td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Fungi</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Viruses</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Protozoa</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

IMMUNE DEFICIENCY
4.5 Warning signs of primary immune deficiency*

The presence of ≥ 2 warning signs may indicate an underlying primary immunodeficiency:

- ≥ 4 new ear infections within 1 yr
- ≥ 2 serious sinus infections within 1 yr
- ≥ 2 mths on antibiotics with little effect
- ≥ 2 pneumonias within 1 yr
- Failure of an infant to gain weight or grow normally
- Recurrent, deep skin or organ abscesses
- Persistent thrush in mouth or fungal infection on skin
- Need for intravenous antibiotics to clear infections
- ≥ 2 deep-seated infections, including septicaemia
- A family history of primary immune deficiency

*Developed by the Jeffrey Modell Foundation (www.info4pi.org).

network so some diseases do not fall easily into this classification.

Presenting problems in immune deficiency

Recurrent infections

Most patients with an immune deficiency present with recurrent infections. While there is no accepted definition of ‘too many’ infections, features that may indicate immune deficiency are shown in Box 4.5. Frequent, severe infections or infections caused by unusual organisms or at unusual sites are the most useful indicator.

Baseline investigations include full blood count with white cell differential, acute phase reactants (CRP, see below), renal and liver function tests, urine dipstick, serum immunoglobulins with protein electrophoresis, and total IgE level. Additional microbiological, virological and radiological tests may be appropriate. At this stage, it may be clear which category of immune deficiency should be considered, and specific investigation can be undertaken, as described below.

If an immune deficiency is suspected but has not yet been formally characterised, patients should not receive live vaccines because of the risk of vaccine-induced disease. Discussion with specialists will help determine whether additional preventative measures, such as prophylactic antibiotics, are indicated.

Primary phagocyte deficiencies

Primary phagocyte deficiencies (see Fig. 4.2, p. 73) usually present with recurrent bacterial and fungal infections which may affect unusual sites. Aggressive management of existing infections, including intravenous antibiotics and surgical drainage of abscesses, and long-term prophylaxis with antibacterial and antifungal agents, is required. Specific treatment depends upon the nature of the defect; haematopoietic stem cell transplantation may be considered (p. 1017).

Leucocyte adhesion deficiencies

These are rare disorders of phagocyte migration, when failure to express adhesion molecules on the surface of leucocytes results in their inability to exit the blood stream. They are characterised by recurrent bacterial infections with high blood neutrophil counts but sites of infection lack pus or other evidence of neutrophil infiltration.

Chronic granulomatus disease

This is caused by mutations in the genes encoding the NADPH oxidase enzymes, which results in failure of oxidative killing. The defect leads to susceptibility to catalase-positive organisms, such as Staphylococcus aureus, Burkholderia cenocepacia and Aspergillus. Intracellular killing of mycobacteria is also impaired. Infections most commonly involve the lungs, lymph nodes, soft tissues, bone, skin and urinary tract, and are characterised histologically by granuloma formation.

Defects in cytokines and cytokine receptors

Defects of cytokines such as IFN-γ, IL-12 or their receptors also result in failure of intracellular killing, with particular susceptibility to mycobacterial infections.

Complement pathway deficiencies

Genetic deficiencies of almost all the complement pathway proteins (see Fig. 4.3, p. 75) have been described. Many present with recurrent infection with encapsulated bacteria, particularly Neisseria species, reflecting the importance of the membrane attack complex in defence against these bacteria. In addition, genetic deficiencies of the classical complement pathway (C1, C2 and C4) are associated with a high prevalence of autoimmune disease, particularly systemic lupus erythematosus (SLE, p. 1109).

In contrast to other complement deficiencies, mannose-binding lectin deficiency is very common (5% of the northern European population). Complete deficiency may predispose to bacterial infections in the presence of an additional cause of immune compromise, such as premature birth or chemotherapy, but is otherwise well tolerated. Deficiency of the complement regulatory protein Cl inhibitor is not associated with recurrent infections but causes recurrent angioedema (p. 93).

Investigations and management

Complement C3 and C4 are the only complement components that are routinely measured. Screening for complement deficiencies is performed using more specialised functional tests of complement-mediated haemolysis, known as CH50 and AP50 (classical haemolytic pathway 50 and alternative pathway 50). If abnormal, these haemolytic tests should be followed by measurement of individual complement components.

There is no definitive treatment for complement deficiencies. Patients should be vaccinated with meningococcal, pneumococcal and H. influenzae B vaccines in order to boost their adaptive immune responses. Lifelong prophylactic penicillin to prevent meningococcal infection is recommended. At-risk family members should also be screened.
**Primary deficiencies of the adaptive immune system**

**Primary T-lymphocyte deficiencies**

These are characterised by recurrent viral, protozoal and fungal infections (see Box 4.4). In addition, many T-cell deficiencies are associated with defective antibody production because of the importance of T cells in regulating B cells. These disorders generally present in childhood and are illustrated in Figure 4.6.

**DiGeorge syndrome**

This results from failure of development of the 3rd/4th pharyngeal pouch, usually caused by a deletion of 22q11. It is associated with multiple abnormalities, including congenital heart disease, hypocalcaemia, tracheo-oesophageal fistulae, cleft lip and palate, and absent thymic development. The immune deficiency is characterised by very low numbers of circulating T cells, despite normal development in the bone marrow.

**Bare lymphocyte syndromes**

These are caused by absent expression of HLA molecules within the thymus. If HLA class I molecules are affected, CD8+ lymphocytes fail to develop, while absent expression of HLA class II molecules affects CD4+ lymphocyte maturation. In addition to recurrent infections, failure to express HLA class I is associated with systemic vasculitis caused by uncontrolled activation of NK cells.

**Autoimmune lymphoproliferative syndrome**

This is caused by failure of normal lymphocyte apoptosis (p. 50), leading to non-malignant accumulation of autoreactive cells. This results in lymphadenopathy, splenomegaly and a variety of autoimmune diseases.

**Investigations and management**

The principal tests for T-lymphocyte deficiencies are a total blood lymphocyte count and quantitation of lymphocyte subpopulations by flow cytometry. Serum immunoglobulins should also be measured. Line, functional tests of T-cell activation and proliferation may be indicated. Patients in whom T-lymphocyte deficiencies are suspected should be tested for human immunodeficiency (HIV) infection (p. 392).

Anti-\textit{Pneumocystis} and antifungal prophylaxis, and aggressive management of infections, are required. Immunoglobulin replacement may be indicated if antibody production is impaired. Haematopoietic stem cell transplantation (HSCT, p. 1017) may be appropriate.

**Combined B- and T-lymphocyte immune deficiencies**

Severe combined immune deficiency (SCID) is caused by defects in lymphoid precursors and results in combined failure of B- and T-cell maturation. The absence of an effective adaptive immune response causes recurrent bacterial, fungal and viral infections soon after birth. HSCT (p. 1017) is the only current treatment, although gene therapy is under investigation.

**Primary antibody deficiencies**

Primary antibody deficiencies (Fig. 4.7) are characterised by recurrent bacterial infections, particularly of the respiratory and gastrointestinal tract. The most common causative organisms are encapsulated bacteria, such as \textit{Strep. pneumoniae} and \textit{H. influenzae}. These disorders may present in infancy, when the protective benefit of transferred maternal immunoglobulin has waned. However, three forms of primary antibody deficiency can also present in adulthood:
There is overlap between specific antibody deficiency, IgA deficiency and CVID, and some patients may progress to a more global antibody deficiency over time.

### Investigations

Investigations include serum immunoglobulins (Box 4.6), with protein and urine electrophoresis to exclude secondary causes of hypogammaglobulinaemia, and B and T lymphocyte counts in blood by flow cytometry. Specific antibody responses to known pathogens can be assessed by measuring IgG antibodies against tetanus, H. influenzae and Strep. pneumoniae (most patients will have been exposed to these antigens through infection or immunisation). If specific antibody levels are low, immunisation with the appropriate killed vaccine should be followed by repeat antibody measurement 6–8 weeks later; failure to mount a response indicates a defect in antibody production. These functional tests have superseded IgG subclass quantitation.

### Management

With the exception of individuals with selective IgA deficiency, patients with antibody deficiencies require aggressive treatment of infections, and prophylactic antibiotics may be indicated. The mainstay of treatment is life-long immunoglobulin replacement therapy. This

---

**Selective IgA deficiency** is the most common primary immune deficiency, affecting 1:600 northern Europeans. In most patients, low (< 0.05 g/L) or undetectable IgA is an incidental finding with no clinical sequelae. However, 30% of individuals experience recurrent mild respiratory and gastrointestinal infections. In some patients, there is a compensatory increase in serum IgG levels. Specific treatment is generally not required.

**Common variable immune deficiency (CVID)** is a heterogeneous primary immune deficiency of unknown cause. It is characterised by low serum IgG levels and failure to make antibody responses to exogenous pathogens. Paradoxically, antibody-mediated autoimmune diseases, such as autoimmune haemolytic anaemia, are common. CVID is also associated with an increased risk of malignancy, particularly lymphoproliferative disease.

**Specific antibody deficiency or functional IgG antibody deficiency** is a poorly characterised condition which causes defective antibody responses to polysaccharide antigens. Some patients are deficient in antibody subclasses IgG2 and IgG4, and this condition was previously called IgG subclass deficiency.

---

**Fig. 4.7 B lymphocytes and primary antibody deficiencies (green boxes).**

### 4.6 Investigation of primary antibody deficiencies

<table>
<thead>
<tr>
<th>Serum immunoglobulin concentrations</th>
<th>Blood cell count</th>
<th>Test immunisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgM</td>
<td>IgG</td>
<td>IgA</td>
</tr>
<tr>
<td>Selective IgA deficiency</td>
<td>Normal</td>
<td>Often elevated</td>
</tr>
<tr>
<td>Common variable immune deficiency</td>
<td>Normal or low</td>
<td>Low</td>
</tr>
<tr>
<td>Specific antibody deficiency</td>
<td>Normal</td>
<td>Normal</td>
</tr>
</tbody>
</table>
IMMUNOLOGICAL FACTORS IN DISEASE

is derived from pooled plasma (p. 1011) and contains IgG antibodies to a wide variety of common organisms. Immunoglobulin replacement may be administered either intravenously or subcutaneously, often by the patient, with the aim of maintaining trough IgG levels within the normal range. Immunisation is generally not effective because of the defect in IgG antibody production. As with all primary immune deficiencies, live vaccines should be avoided (p. 148).

Secondary immune deficiencies

Secondary immune deficiencies are much more common than primary immune deficiencies (Box 4.7). Common causes include infections, such as HIV and measles, and cytotoxic and immunosuppressive drugs, particularly those used in the management of transplantation, autoimmunity and cancer. Physiological immune deficiency occurs at the extremes of life; the decline of the immune response in the elderly is known as immune senescence (Box 4.8). Management of secondary immune deficiency is described in the relevant chapters on infectious diseases (Ch. 13), HIV (Ch. 14), oncology (Ch. 11) and haematological disorders (Ch. 24).

<table>
<thead>
<tr>
<th>4.7 Causes of secondary immune deficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physiological</strong></td>
</tr>
<tr>
<td>• Ageing</td>
</tr>
<tr>
<td>• Prematurity</td>
</tr>
<tr>
<td><strong>Infection</strong></td>
</tr>
<tr>
<td>• HIV</td>
</tr>
<tr>
<td>• Measles</td>
</tr>
<tr>
<td><strong>Iatrogenic</strong></td>
</tr>
<tr>
<td>• Immunosuppressive therapy</td>
</tr>
<tr>
<td>• Antineoplastic agents</td>
</tr>
<tr>
<td>• Corticosteroids</td>
</tr>
<tr>
<td><strong>Malignancy</strong></td>
</tr>
<tr>
<td>• B-cell malignancies</td>
</tr>
<tr>
<td>• Including leukaemia, lymphoma and myeloma</td>
</tr>
<tr>
<td><strong>Biochemical and nutritional disorders</strong></td>
</tr>
<tr>
<td>• Malnutrition</td>
</tr>
<tr>
<td>• Renal insufficiency/dialysis</td>
</tr>
<tr>
<td>• Diabetes mellitus</td>
</tr>
<tr>
<td><strong>Other conditions</strong></td>
</tr>
<tr>
<td>• Burns</td>
</tr>
<tr>
<td>• Asplenia/hyposplenism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.8 Ageing and immune senescence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T-cell responses</strong>: decline, with reduced delayed type hypersensitivity responses.</td>
</tr>
<tr>
<td><strong>Antibody production</strong>: decreased for many exogenous antigens. Although autoantibodies are frequently detected, autoimmune disease is less common.</td>
</tr>
<tr>
<td><strong>Response to vaccination</strong>: reduced, e.g. 30% of healthy older people may not develop protective immunity after influenza vaccination.</td>
</tr>
<tr>
<td><strong>Allergic disorders and transplant rejection</strong>: less common.</td>
</tr>
<tr>
<td><strong>Susceptibility to infection</strong>: increased, e.g. community-acquired pneumonia by threefold and urinary tract infection by 20-fold. Latent infections, including tuberculosis and herpes zoster, may be reactivated.</td>
</tr>
<tr>
<td><strong>Manifestations of inflammation</strong>: may be absent, e.g. lack of pyrexia or leucocytosis.</td>
</tr>
</tbody>
</table>

THE INFLAMMATORY RESPONSE

Inflammation is the response of tissues to injury or infection, and is necessary for normal repair and healing. This section focuses on the generic inflammatory response and its multisystem manifestations. The role of inflammation in specific diseases is illustrated in many other chapters of this book.

Physiology and pathology of inflammation

Acute inflammation

Acute inflammation is the result of rapid and complex interplay between the cells and soluble molecules of the innate immune system. The classical external signs include heat, redness, pain and swelling (calor, rubor, dolor and oedema, Fig. 4.8).

The inflammatory process is initiated by local tissue injury or infection. Damaged epithelial cells produce cytokines and antimicrobial peptides, causing early infiltration of phagocytic cells. As a result, there is production of leukotrienes, prostaglandins, histamine, kinins, anaphylotoxins and inducible nitric oxide synthase within inflamed tissue. The effect is vasodilatation and increased local vascular permeability, which increases trafficking of fluid and cells to the affected tissue. In addition, pro-inflammatory cytokines produced at the site of injury have profound systemic effects. IL-1, TNF-α and IL-6 act on the hypothalamus to raise the temperature set-point, causing fever, and also stimulate the production of acute phase proteins.

Acute phase proteins

Acute phase proteins are produced by the liver in response to inflammatory stimuli and have a wide range of activities. CRP and serum amyloid A may be increased 1000-fold, contributing to host defence and stimulating repair and regeneration. Fibrinogen plays an essential role in wound healing, and α1-antitrypsin and α1-antichymotrypsin control the pro-inflammatory cascade by neutralising the enzymes produced by activated neutrophils, preventing widespread tissue destruction. In addition, antioxidants, such as haptoglobin and manganese superoxide dismutase, scavenge for oxygen free radicals, while increased levels of iron-binding proteins, such as ferritin and lactoferrin, decrease the iron available for uptake by bacteria (p. 1023). Immunoglobulins are not acute phase proteins but are often increased in chronic inflammation.

Resolution of inflammation

Resolution of an inflammatory response is crucial for normal healing. This involves active down-modulation
of inflammatory stimuli and repair of bystander damage to local tissues. Extravasated neutrophils undergo apoptosis and are phagocytosed by macrophages, along with the remains of microorganisms. Macrophages also synthesise collagenase and elastase, which break down local connective tissue and remove debris. Macrophage-derived cytokines, including transforming growth factor (TGF)-β and platelet-derived growth factor, attract fibroblasts and promote synthesis of new collagen, while angiogenic factors stimulate new vessel formation.

**Sepsis and septic shock**

Septic shock is the clinical manifestation of overwhelming inflammation (p. 190). Failure of normal inhibitory mechanisms results in excessive production of pro-inflammatory cytokines by macrophages, causing hypertension, hypovolaemia, decreased perfusion and tissue oedema. In addition, uncontrolled neutrophil activation causes release of proteases and oxygen free radicals within blood vessels, damaging the vascular endothelium and further increasing capillary permeability. Direct activation of the coagulation pathway combines with endothelial cell disruption to form clots within the damaged vessels. The clinical consequences include cardiovascular collapse, acute respiratory distress syndrome, disseminated intravascular coagulation, multiorgan failure and often death. Septic shock most frequently results from infection with Gram-negative bacteria, because lipopolysaccharide is particularly effective at activating the inflammatory cascade.

**Chronic inflammation**

In most instances, the development of an active immune response results in either clearance or control of the inflammatory stimulus with minimal local damage. Failure of elimination may result in chronic inflammation. Persisting microorganisms stimulate the ongoing accumulation of neutrophils, macrophages and activated T lymphocytes. If this is associated with local deposition of fibrous connective tissue, a granuloma may form. Granulomas are characteristic of infections such as tuberculosis and leprosy, in which the microorganism is protected by a robust cell wall which shields it from killing, despite phagocytosis.

Too vigorous or prolonged immune responses may cause bystander tissue damage, known as hypersensitivity responses. The Gell and Coombs classification of hypersensitivity disorders is discussed on page 87.

**Investigations in inflammation**

Changes associated with inflammation are reflected in many laboratory investigations. Leucocytosis is common,
and reflects the transit of activated neutrophils and monocytes to the site of infection. The platelet count may also be increased. The most widely used laboratory measure of acute inflammation is the C-reactive protein (see below). Plasma levels of many other acute phase reactants, including fibrinogen, ferritin and complement components, are increased in response to acute inflammation, while albumin levels are reduced. Chronic inflammation is frequently associated with a normocytic normochromic anaemia of chronic disease (p. 1023).

**C-reactive protein**

C-reactive protein (CRP) is an acute phase reactant synthesised by the liver, which opsonises invading pathogens. Levels of CRP increase within 6 hours of an inflammatory stimulus and may rise up to 1000-fold. Measurement of CRP provides a direct index of acute inflammation and reflects the transit of activated neutrophils and monocytes to the site of infection. The platelet count may also be increased. The most widely used laboratory measure of acute inflammation is the C-reactive protein (see below). Plasma levels of many other acute phase reactants, including fibrinogen, ferritin and complement components, are increased in response to acute inflammation, while albumin levels are reduced. Chronic inflammation is frequently associated with a normocytic normochromic anaemia of chronic disease (p. 1023).

**Erythrocyte sedimentation rate**

In contrast to the CRP, the erythrocyte sedimentation rate (ESR) is an indirect measure of inflammation. It measures how fast erythrocytes fall through anticoagulated blood, and is determined by a combination of the composition of plasma proteins and the morphology of circulating erythrocytes. These factors govern the propensity of red cells to aggregate, which is the major determinant of the ESR. Erythrocytes are inherently negatively charged, and this prevents them clumping together in the blood stream. Plasma proteins are positively charged and an increase in plasma proteins neutralises the surface charge of erythrocytes, overcoming their inherent repulsive forces and causing them to aggregate, or stack like tyres, forming rouleaux. Rouleaux have a higher mass/surface area ratio than single red cells, and therefore sediment faster.

The most common cause of an increased ESR is an acute phase response, which leads to an increase in the concentration of acute phase reactants, including CRP. However, other conditions that do not affect acute phase proteins may alter the composition and concentration of other plasma proteins (see Box 4.9). For example, immunoglobulins comprise a significant proportion of plasma proteins, but do not participate in the acute phase response. Thus, any condition that causes a monoclonal or polyclonal increase in serum immunoglobulins will increase the ESR without a corresponding rise in CRP. In addition, changes in erythrocyte surface area and density influence sedimentation, and abnormal red cell morphology can make rouleaux formation impossible. For these reasons, an inappropriately low ESR occurs in spherocytosis and sickle cell anaemia.

As CRP is a simple and sensitive early indicator of the acute phase response, it is increasingly used in preference to the ESR. If both ESR and CRP are used, any discrepancy should be resolved by assessing the individual determinants of the ESR, i.e. full blood count and blood film, serum immunoglobulins (IgG, IgA and IgM) and protein electrophoresis. The IgE concentration in

---

**4.9 Conditions commonly associated with abnormal CRP and/or ESR**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Consequence</th>
<th>Effect on CRP</th>
<th>Effect on ESR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute bacterial, fungal or viral infection</td>
<td>Acute phase response</td>
<td>Increased (range 50–150 mg/L; in severe infections may be &gt; 300 mg/L)</td>
<td>Increased</td>
</tr>
<tr>
<td>Necrotising bacterial infection</td>
<td>Profound acute inflammatory response</td>
<td>Increased +++ (may be &gt; 300 mg/L)</td>
<td>Increased</td>
</tr>
<tr>
<td>Acute inflammatory diseases, e.g. Crohn’s disease, polymyalgia rheumatica</td>
<td>Acute phase response</td>
<td>Increased (range 50–150 mg/L)</td>
<td>Increased</td>
</tr>
<tr>
<td>Chronic bacterial or fungal infection, e.g. localised abscess, bacterial endocarditis or tuberculosis</td>
<td>Acute and chronic inflammatory response: increased acute phase proteins with polyclonal increase in immunoglobulins</td>
<td>Increased (range 50–150 mg/L)</td>
<td>Increase disproportionate to CRP</td>
</tr>
<tr>
<td>SLE, Sjögren’s syndrome</td>
<td>Chronic inflammatory response with polyclonal increase in immunoglobulins</td>
<td>Normal (paradoxically)</td>
<td>Increased</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>Monoclonal increase in immunoglobulin without acute inflammation</td>
<td>Normal</td>
<td>Increased</td>
</tr>
<tr>
<td>Pregnancy, old age, end-stage renal disease</td>
<td>Normal immunoglobulins but increased fibrinogen</td>
<td>Normal</td>
<td>Moderately increased</td>
</tr>
</tbody>
</table>

| Reference range: adult males < 10 mm/hr, adult females < 20 mm/hr. |
plasma is very low and does not contribute significantly to the ESR.

**Plasma viscosity**
Plasma viscosity is another surrogate measure of plasma protein concentration. Like the ESR, it is affected by the concentration of large plasma proteins, including fibrinogen and immunoglobulins. However, it is not affected by properties of erythrocytes and is generally considered to be more reliable than the ESR.

**Presenting problems in inflammation**
In most patients presenting with the manifestations of acute inflammation shown in Figure 4.8, it is possible to identify the source of the problem quickly and to assess the consequences, as discussed in other chapters. Systemic manifestations of inflammation include fever (p. 296), leucocytosis (p. 1005) and shock (p. 190).

**Unexplained raised ESR**
The ESR should not be used to screen asymptomatic patients for the presence of disease. However, in the era of frequent routine laboratory testing, an unexplained raised ESR is a common problem.

**Clinical assessment**
A comprehensive history and examination are crucial. Extreme elevations in the ESR (> 100 mm/hr) rarely occur in the absence of significant disease (see Box 4.9).

**Investigations**
Assessing the CRP, serum immunoglobulins and electrophoresis, and urine electrophoresis will help determine whether the elevation in ESR is due to an inflammatory process (see Box 4.9).

A full blood count and film may show a normocytic, normochromic anaemia, which occurs in many chronic diseases. Leucocytosis may reflect infection, inflammatory disease or tissue necrosis. Neutrophilia suggests infection or acute inflammation. Atypical lymphocytes may occur in some chronic infections, such as cytomegalovirus (CMV) and Epstein-Barr virus (EBV).

Abnormalities in liver function suggest either a local infective process (hepatitis, hepatic abscess or biliary sepsis) or systemic disease, including malignancy.

Blood and urine cultures should be performed. It may be relevant to measure antinuclear and antineutrophil cytoplasmic antibodies, and to exclude chronic infections, including HIV and syphilis.

In the unusual circumstances when ESR is elevated but both CRP and immunoglobulins are normal, fibrinogen should be measured. Elevated fibrinogen causes a higher ESR in older people, women, and patients with renal or heart failure, obesity and diabetes mellitus.

**Imaging**
If indicated by the clinical and laboratory features, a chest X-ray and abdominal computed tomography (CT) scan may identify a source of unknown infection or malignancy. An abdominal and pelvic ultrasound may identify hepatic lesions, abdominal nodes and local intra-abdominal or pelvic abscesses. Magnetic resonance imaging (MRI) is more appropriate for the diagnosis of soft tissue or bone/joint infections. Echocardiography is used to look for vegetations and assess valve function in suspected bacterial endocarditis. White cell scans are rarely indicated but may be useful in identifying the site of pyogenic infection. An isotope bone scan may provide evidence of malignancy or focal bone infection.

**Periodic fever syndromes**
These rare disorders are characterised by recurrent episodes of fever and systemic inflammation, associated with an elevated acute phase response.

**Familial Mediterranean fever**
Familial Mediterranean fever (FMF) is the most common of the familial periodic fevers, predominantly affecting Mediterranean people, including Arabs, Turks, Sephardic Jews and Armenians. It results from mutations of the *MEFV* gene, which encodes a protein called pyrin. Pyrin regulates neutrophil-mediated inflammation by indirectly suppressing the production of IL-1.

FMF is characterised by recurrent painful attacks of fever associated with peritonitis, pleuritis and arthritis, which last for a few hours to 4 days and which are associated with markedly increased CRP levels. Symptoms resolve completely between episodes. The majority of individuals have their first attack before the age of 20 years. The major complication of FMF is AA amyloidosis (see below). Colchicine significantly reduces the number of febrile episodes in 90% of patients but is ineffective during acute attacks.

**Mevalonate kinase deficiency**
Mevalonate kinase deficiency (previously known as hyper-IgD syndrome, or HIDS) is an autosomal recessive disorder that causes recurrent attacks of fever, abdominal pain, diarrhoea, lymphadenopathy, arthralgia, skin lesions and aphthous ulceration. Most patients are from Western Europe, particularly the Netherlands and northern France. Mevalonate kinase is involved in the metabolism of cholesterol, but why mutations in its gene cause an inflammatory periodic fever remains unknown. Serum IgD and IgA levels are persistently elevated, and CRP levels are increased during acute attacks. Standard anti-inflammatory drugs (including colchicine and steroids) are ineffective.

**TNF receptor-associated periodic syndrome**
TNF receptor-associated periodic syndrome (TRAPS), also known as Hibernian fever, is an autosomal dominant syndrome, causing recurrent periodic fever, arthralgia, myalgia, serositis and rashes, which has been reported in many ethnic groups. Attacks may be prolonged (lasting over 1 week). During a typical attack, there is neutrophilia, increased CRP and elevated IgA levels. The diagnosis can be confirmed by low serum levels of the soluble type 1 TNF receptor and by analysis of the *TNFRSF1A* gene. As in FMF, the major complication is amyloidosis, and regular screening for proteinuria is advised. TRAPS responds to systemic corticosteroids and to biological therapies, including soluble TNF receptor therapy and IL-1 receptor antagonists (p. 1102).
Amyloidosis

The amyloidoses are a group of acquired and hereditary disorders characterised by extracellular deposition of insoluble proteins (Box 4.10). These complex deposits consist of fibrils of the specific protein involved, linked to glycosaminoglycans, proteoglycans and serum amyloid P (SAP). Protein accumulation may be localised or systemic, and the clinical manifestations depend upon the organ(s) affected. The diagnosis of amyloidosis should be considered in all cases of unexplained nephrotic syndrome (p. 476), cardiomyopathy (p. 636) and peripheral neuropathy (p. 1223).

Diagnosis

The diagnosis is established by biopsy, which may be of an affected organ, rectum or subcutaneous fat. The pathognomonic histological feature is apple-green birefringence of amyloid deposits when stained with Congo red dye and viewed under polarised light. Immunohistochemical staining can identify the type of Congo red dye and viewed under polarised light. Quantitative scintigraphy with radio-labelled SAP is a valuable tool in determining the overall load and distribution of amyloid deposits.

Management

The aims of treatment are to support the function of affected organs and, in acquired amyloidosis, to prevent further amyloid deposition through treatment of the primary cause. When the latter is possible, regression of existing amyloid deposits may occur. Liver transplantation may provide definitive treatment in selected patients with hereditary transthyretin amyloidosis.

### AUTOIMMUNE DISEASE

Autoimmunity can be defined as the presence of immune responses against self tissue. This may be a harmless phenomenon, identified only by the presence of low titre autoantibodies or autoreactive T cells. However, if these responses cause significant organ damage, this results in autoimmune diseases, which are a major cause of chronic morbidity and disability, affecting up to 1 in 30 adults at some time (Box 4.11).

### Pathophysiology of autoimmunity

**Immunological tolerance**

Autoimmunity results from the failure of immunological tolerance, the process by which the immune system recognises and accepts self tissue. There are a number of mechanisms of immune tolerance. Central tolerance occurs during lymphocyte development, when T and B lymphocytes that recognise self antigens are eliminated. This is a crucial mechanism of immune tolerance. Peripheral tolerance occurs during lymphocyte development, when T and B lymphocytes that recognise self antigens are eliminated. This is a crucial mechanism of immune tolerance.

### 4.10 Amyloid disorders

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Pathological basis</th>
<th>Predisposing conditions</th>
<th>Other features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired systemic amyloidosis</td>
<td>Increased production of serum amyloid A as part of prolonged or recurrent acute inflammatory response</td>
<td>Chronic infection (TB, bronchiectasis, chronic abscess, osteomyelitis) Chronic inflammatory diseases (untreated rheumatoid arthritis, FMF)</td>
<td>90% of patients present with non-selective proteinuria or nephrotic syndrome</td>
</tr>
<tr>
<td>Reactive (AA) amyloidosis</td>
<td>Increased production of monoclonal light chain</td>
<td>Monoclonal gammapathies, including myeloma, benign gammapathies and plasmacytoma</td>
<td>Restrictive cardiomyopathy, peripheral and autonomic neuropathy, carpal tunnel syndrome, proteinuria, spontaneous purpura, amyloid nodules and plaques. MacroGLOSSIA occurs rarely but is pathognomonic. Prognosis is poor</td>
</tr>
<tr>
<td>Light chain amyloidosis (AL)</td>
<td>Increased production of monoclonal light chain</td>
<td>Monoclonal gammapathies, including myeloma, benign gammapathies and plasmacytoma</td>
<td>90% of patients present with non-selective proteinuria or nephrotic syndrome</td>
</tr>
<tr>
<td>Dialysis-associated (Aβ2M) amyloidosis</td>
<td>Accumulation of circulating α2-microglobulin due to failure of renal catabolism in kidney failure</td>
<td>Renal dialysis</td>
<td>90% of patients present with non-selective proteinuria or nephrotic syndrome</td>
</tr>
<tr>
<td>Senile systemic amyloidosis</td>
<td>Normal transthyretin protein deposited in tissues</td>
<td>Age &gt; 70 yrs</td>
<td>90% of patients present with non-selective proteinuria or nephrotic syndrome</td>
</tr>
<tr>
<td>Hereditary systemic amyloidosis</td>
<td>Production of protein with an abnormal structure that predisposes to amyloid fibril formation. Most commonly due to mutations in transthyretin gene</td>
<td>Autosomal dominant inheritance</td>
<td>Peripheral and autonomic neuropathy, cardiomyopathy, Renal involvement unusual 10% of gene carriers are asymptomatic throughout life</td>
</tr>
</tbody>
</table>

(FMF = familial Mediterranean fever)

**Table 4.10 Amyloid disorders**

*Note: Table 4.10 Amyloid disorders is a concise summary of the various forms of amyloidosis, including their pathological basis, predisposing conditions, and other features.*
continues throughout life as immature lymphocytes are generated. Inevitably some autoreactive cells evade deletion and reach the peripheral tissues, where they are controlled by peripheral tolerance mechanisms. These include suppression of autoreactive cells by regulatory T cells, generation of functional hyporesponsiveness (‘anergy’) in lymphocytes which encounter antigen in the absence of the co-stimulatory signals that accompany inflammation, and T cell death by apoptosis.

Autoimmune diseases develop when self-reactive lymphocytes escape from these tolerance mechanisms and become activated.

Factors predisposing to autoimmune disease
Autoimmune diseases are much more common in women than in men, for reasons which remain unclear. Most autoimmune diseases have multiple genetic determinants (Box 4.12). Many are associated with variation at specific HLA loci, reflecting the importance of HLA genes in shaping lymphocyte responses. Other important susceptibility genes include those determining cytokine activity, co-stimulation and cell death. Even though some of these associations are the strongest that have been identified in polygenic diseases (p. 68), they have limited predictive value, and are not useful in determining disease risk for individual patients. Several acquired factors can trigger autoimmunity in genetically predisposed individuals, including infection, cigarette smoking and hormone levels. The most widely studied of these is infection, as occurs in acute rheumatic fever following bacterial infection. A number of mechanisms have been postulated, such as cross-reactivity between the infectious pathogen and self antigens (molecular mimicry), and release of sequestered antigens, which are not usually visible to the immune system, from damaged tissue. Alternatively, infection may result in the production of inflammatory cytokines, which overwhelm the normal control mechanisms that prevent bystander damage. Occasionally, the development of autoimmune disease is a side-effect of drug treatment. For example, the metabolic products of the anaesthetic agent halothane bind to liver enzymes, resulting in a structurally novel protein. This is recognised as a new (foreign) antigen by the immune system, and the autoantibodies and activated T cells directed against it may cause hepatic necrosis.

Classification of autoimmune diseases
The spectrum of autoimmune diseases is broad. They can be classified by organ involvement (see Box 4.11) or by the predominant mechanism responsible for tissue damage. The Gell and Coombs classification of hypersensitivity is the most widely used, and distinguishes four types of immune response which result in bystander tissue damage (Box 4.13).

- **Type I hypersensitivity** is relevant in allergy but is not associated with autoimmune disease.
- **In type II hypersensitivity**, injury is localised to a single tissue or organ and is mediated by specific autoantibodies.
- **Type III hypersensitivity** is a generalised reaction resulting from immune complex deposition which initiates activation of the classical complement cascade, as well as recruitment and activation of phagocytes and CD4+ lymphocytes. The site of immune complex deposition is determined by the relative amount of antibody, size of the immune complexes, nature of the antigen and local haemodynamics. Generalised deposition of immune complexes gives rise to systemic diseases such as SLE.
- **In type IV hypersensitivity**, activated T cells and macrophages mediate phagocytosis and tissue damage.

### 4.11 The spectrum of autoimmune disease

<table>
<thead>
<tr>
<th>Type</th>
<th>Disease</th>
<th>Page no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organ-specific</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immune response</td>
<td>Graves’ disease</td>
<td>747</td>
</tr>
<tr>
<td>directed against</td>
<td>Hashimoto’s thyroiditis</td>
<td>751</td>
</tr>
<tr>
<td>localised antigens</td>
<td>Addison’s disease</td>
<td>777</td>
</tr>
<tr>
<td></td>
<td>Pernicious anaemia</td>
<td>1025</td>
</tr>
<tr>
<td></td>
<td>Type 1 diabetes</td>
<td>803</td>
</tr>
<tr>
<td></td>
<td>Sympathetic ophthalmoplegia</td>
<td>1169</td>
</tr>
<tr>
<td></td>
<td>Multiple sclerosis</td>
<td>1188</td>
</tr>
<tr>
<td></td>
<td>Goodpasture’s syndrome</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>Pemphigus vulgaris</td>
<td>1294</td>
</tr>
<tr>
<td></td>
<td>Bullous pemphigoid</td>
<td>1292</td>
</tr>
<tr>
<td></td>
<td>Idiopathic thrombocytopenic purpura</td>
<td>1050</td>
</tr>
<tr>
<td></td>
<td>Autoimmune haemolytic anaemia</td>
<td>1029</td>
</tr>
<tr>
<td></td>
<td>Myasthenia gravis</td>
<td>1226</td>
</tr>
<tr>
<td></td>
<td>Primary antiphospholipid syndrome</td>
<td>1055</td>
</tr>
<tr>
<td></td>
<td>Rheumatoid arthritis</td>
<td>1096</td>
</tr>
<tr>
<td></td>
<td>Dermatomyositis</td>
<td>1114</td>
</tr>
<tr>
<td></td>
<td>Primary biliary cirrhosis</td>
<td>963</td>
</tr>
<tr>
<td></td>
<td>Autoimmune hepatitis</td>
<td>962</td>
</tr>
<tr>
<td></td>
<td>Sjögren’s syndrome</td>
<td>1114</td>
</tr>
<tr>
<td>Multisystem</td>
<td>Systemic sclerosis</td>
<td>1112</td>
</tr>
<tr>
<td>Immune response</td>
<td>Mixed connective tissue disease</td>
<td>1113</td>
</tr>
<tr>
<td>directed to</td>
<td></td>
<td></td>
</tr>
<tr>
<td>widespread target</td>
<td></td>
<td></td>
</tr>
<tr>
<td>antigens</td>
<td>SLE</td>
<td>1109</td>
</tr>
</tbody>
</table>

### 4.12 Some genetic variations predisposing to autoimmune diseases

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA complex</td>
<td>Key determinants of antigen presentation to T cells</td>
<td>Most autoimmune diseases</td>
</tr>
<tr>
<td>PTPN22</td>
<td>Regulates T- and B-cell receptor signalling</td>
<td>Rheumatoid arthritis, type 1 diabetes, SLE</td>
</tr>
<tr>
<td>CTLA4</td>
<td>Important co-stimulatory molecule which transmits inhibitory signals to T cells</td>
<td>Rheumatoid arthritis, type 1 diabetes</td>
</tr>
<tr>
<td>TNFRSF1A</td>
<td>Control of TNF network</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>ATG5</td>
<td>Autophagy</td>
<td>SLE</td>
</tr>
</tbody>
</table>
### 4.13 Gell and Coombs classification of hypersensitivity diseases

<table>
<thead>
<tr>
<th>Type</th>
<th>Mechanism</th>
<th>Example of disease in response to exogenous agent</th>
<th>Example of autoimmune disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>Immediate hypersensitivity</td>
<td>IgE-mediated mast cell degranulation</td>
<td>Allergic disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>None described</td>
</tr>
<tr>
<td>Type II</td>
<td>Antibody-mediated</td>
<td>Binding of cytotoxic IgG or IgM antibodies to antigens on cell surface causes cell killing</td>
<td>ABO blood transfusion reaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hyperacute transplant rejection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Autoimmune haemolytic anaemia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Idiopathic thrombocytopenic purpura</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Goodpasture’s disease</td>
</tr>
<tr>
<td>Type III</td>
<td>Immune complex-mediated</td>
<td>IgG or IgM antibodies bind soluble antigen to form immune complexes which trigger classical complement pathway activation</td>
<td>Serum sickness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Farmer’s lung</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SLE</td>
</tr>
<tr>
<td>Type IV</td>
<td>Delayed type</td>
<td>Activation of T cells and phagocytes</td>
<td>Acute cellular transplant rejection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nickel hypersensitivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Type 1 diabetes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hashimoto’s thyroiditis</td>
</tr>
</tbody>
</table>

### Investigations in autoimmunity

**Autoantibodies**

A number of autoantibodies can be identified in the laboratory and are used in disease diagnosis and monitoring, as discussed elsewhere in this book (e.g. p. 1067). Antibodies are quantified either by titre (the minimal dilution at which the antibody can be detected) or by concentration (in standardised units).

**Measures of complement activation**

Quantitation of complement components may be useful in the evaluation of immune complex-mediated diseases. Classical complement pathway activation leads to a decrease in circulating (unactivated) C4, and is often also associated with decreased C3 levels. Serial measurement of C3 and C4 is a useful surrogate measure of immune complex formation.

**Cryoglobulins**

Cryoglobulins are antibodies directed against other immunoglobulins, and which form immune complexes that precipitate in the cold. They are classified into three types on the basis of the properties of the immunoglobulin involved (Box 4.14). Testing for cryoglobulins requires the transport of a serum specimen to the laboratory at 37°C. Cryoglobulins should not be confused with cold agglutinins; the latter are autoantibodies specifically directed against the I/i antigen on the surface of red cells, which can cause intravascular haemolysis in the cold (p. 1030).
Susceptibility to allergic diseases

The incidence of allergic diseases is increasing. This trend is largely unexplained but one widely held theory is the ‘hygiene hypothesis’. This proposes that infections in early life are critically important in maturation of the immune response and bias the immune system against the development of allergies. It is suggested that the high prevalence of allergic disease is the penalty for the decreased exposure to infection that has resulted from improvements in sanitation and health care.

A number of factors predispose to allergic diseases, the strongest of which is a family history. A wide array of genetic determinants of disease susceptibility have been identified, including genes controlling innate immune responses, cytokine production, IgE levels and the ability of the epithelial barrier to protect against environmental agents. Contributory environmental factors include bacterial and viral infection, pollutants and cigarette smoke.

Pathology of allergy

Normally, the immune system does not make detectable responses to the many environmental substances to which it is exposed daily. However, in an allergic reaction, initial exposure to an otherwise harmless exogenous substance (known as an allergen) triggers the production of specific IgE antibodies by activated B cells (Fig. 4.9). These IgE antibodies bind to the surface of mast cells via high-affinity IgE receptors, a step that is not itself associated with clinical sequelae. However, upon re-exposure, the allergen binds to membrane-bound IgE which activates the mast cells, releasing a variety of mediators (the early phase response, Box 4.15). This type I hypersensitivity reaction is the basis of the symptoms of allergic reactions, which range from sneezing and rhinorrhoea to anaphylaxis (Box 4.16).

In some patients, the early phase response is followed by persistent activation of mast cells, manifest by ongoing swelling and local inflammation. This is known as the late phase reaction and is mediated by basophils, eosinophils and macrophages. Long-standing or recurrent allergic inflammation may give rise to a chronic inflammatory response characterised by a complex infiltrate of macrophages, eosinophils and T lymphocytes, in addition to mast cells and basophils. Once this has been established, inhibition of mast cell mediators with antihistamines is clinically ineffective.

Occasionally, mast cell activation may be non-specifically triggered through other signals, such as neuropeptides, anaphylotoxins and bacterial peptides.
sensitivity and specificity of specific IgE tests (previously known as radioallergosorbent tests, RAST) are lower than skin prick tests. However, IgE tests may be very useful if skin testing is inappropriate: for example, in patients taking antihistamines or those who have severe skin disease or dermatographism. They can also be used to test for cross-reactivity between insect venoms, and post mortem to identify allergens responsible for lethal anaphylaxis.

There is no indication for testing of specific IgG antibodies to allergens in the investigation of allergic diseases.

Supervised exposure to allergen (challenge test)
Allergen challenges are usually performed in specialist centres, and include bronchial provocation testing, nasal challenge and food challenge. These may be particularly useful in the investigation of occupational asthma or food allergy.

** Mast cell tryptase**
After a systemic allergic reaction, the circulating level of mast cell mediators increases dramatically. Tryptase is the most stable of these and serum levels peak at 1–2 hours. Measurement of serum mast cell tryptase is extremely useful in investigating a possible anaphylactic event. Ideally, measurements should be made at the time of the reaction, and 3 hours and 24 hours later.

**Non-specific markers of atopic disease: total serum IgE and eosinophilia**
Peripheral blood eosinophilia is common in atopic individuals. However, eosinophilia of more than 20% or an absolute eosinophil count over $1.5 \times 10^7/L$ should initiate a search for a non-atopic cause (p. 311).

Atopy is the most common cause of elevated total IgE in developed countries. However, there are many other causes, including parasite and helminth infections (pp. 369 and 381), lymphoma (p. 1041), drug reactions and Churg–Strauss vasculitis (p. 1118). Moreover, significant allergic disease can occur despite a normal total IgE level. Thus total IgE quantitation is not indicated in the routine investigation of allergic disease.

**Management**
- Avoidance of the allergen should be rigorously attempted, and the advice of specialist dietitians and occupational physicians may be required.
- Antihistamines block histamine $H_1$ receptors, thereby inhibiting the effects of histamine release. Long-acting, non-sedating preparations are particularly useful for prophylaxis against frequent attacks.
- Corticosteroids down-regulate pro-inflammatory cytokine production. They are highly effective in allergic disease and, if used topicaly, their adverse effects may be minimised.
- Sodium cromoglicate stabilises the mast cell membrane, inhibiting release of vasoactive mediators. It is effective as a prophylactic agent in asthma and allergic rhinitis, but has no role in acute attacks. It is poorly absorbed and therefore ineffective in the management of food allergies.
- Antigen-specific immunotherapy involves the sequential administration of escalating amounts of dilute allergen over a prolonged period of time. Its mechanism of action is unknown, but it is highly
effective in the prevention of insect venom anaphylaxis, and allergic rhinitis secondary to grass pollen (Box 4.17). The traditional route of administration is via subcutaneous injections, which carry a risk of anaphylaxis and should only be performed in specialised centres. More recently, sublingual immunotherapy has been shown to be effective in the management of moderate grass pollen allergy, and clinical trials of immunotherapy for food allergy are ongoing.

- Omalizumab, a monoclonal antibody against IgE, inhibits the binding of IgE to mast cells and basophils. It is effective in moderate and severe allergic asthma and rhinitis.

### EBM 4.17 Immunotherapy for allergy

‘Immunotherapy is effective for treatment of allergic rhinitis, allergic asthma and stinging insect hypersensitivity. Clinical studies to date do not support the use of allergen immunotherapy for food hypersensitivity, chronic urticaria and/or angioedema.’


For further information: www.cochrane.org/cochrane-reviews

---

### Differential diagnosis (examples)

- **Causes of loss of consciousness**
  - Vasovagal syncope
  - Cardiac arrhythmias
  - Myocardial infarction

- **Causes of respiratory distress**
  - Status asthmaticus

- **Causes of laryngeal obstruction**
  - C1 inhibitor deficiency
  - Idiopathic angioedema

- **Causes of generalised flushing**
  - Systemic mastocytosis
  - Carcinoid syndrome

- **Other causes**
  - Phaeochromocytoma

---

### Anaphylaxis

Anaphylaxis is a potentially life-threatening, systemic allergic reaction caused by the release of histamine and other vasoactive mediators from mast cells. The risk of death is increased in patients with pre-existing asthma, particularly if this is poorly controlled, and when treatment with adrenaline (epinephrine) is delayed.

### Clinical assessment

The clinical features are shown in Figure 4.10. The severity of a reaction should be assessed; the time between allergen exposure and onset of symptoms provides a guide. Enquiry should be made about potential triggers; if these are not immediately obvious, a detailed history of the previous 24 hours may be helpful. The most common triggers are foods, latex, insect venom and drugs (Box 4.18). A history of previous allergic responses to the offending agent is common. The route of allergen exposure may influence the principal clinical features of a reaction; for example, if an allergen is inhaled, the major symptom is frequently wheezing. Features of anaphylaxis may overlap with the direct toxic effects of drugs and venoms (Ch. 9). Potentiating factors, such as...
exercise or alcohol, can lower the threshold for an anaphylactic event.

A number of conditions may mimic anaphylaxis (see Fig. 4.10). Anaphylactoid reactions result from the non-specific degranulation of mast cells by drugs, chemicals or other triggers (see Box 4.18), and do not involve IgE antibodies. The clinical presentations are indistinguishable, and in the acute situation discriminating between them is unnecessary. However, this may be important in identifying precipitating factors and appropriate avoidance measures.

**Investigations**
Measurement of serum mast cell tryptase concentrations is useful to confirm the diagnosis. Specific IgE tests may be preferable to skin prick tests when investigating patients with a history of anaphylaxis.

**Management**
Anaphylaxis is an acute medical emergency (Box 4.19). Individuals who have recovered from an anaphylactic event should be referred for specialist assessment. The aim is to identify the trigger factor, to educate the patient regarding avoidance and management of subsequent episodes, and to identify whether specific treatment, such as immunotherapy, is indicated. If the trigger factor cannot be identified or cannot be avoided, recurrence is common. Patients who have previously experienced an anaphylactic event should be prescribed self-injectable adrenaline (epinephrine), and they and their families or carers should be instructed on its use (Box 4.20). The use of a MedicAlert (or similar) bracelet will increase the likelihood that adrenaline will be administered in an emergency. Issues most pertinent to serious allergy in adolescents are shown in Box 4.21.
Angioedema

Angioedema is the episodic, localised, non-pitting swelling of submucous or subcutaneous tissues. This most frequently affects the face (Fig. 4.11), extremities and genitalia. Involvement of the larynx or tongue may cause life-threatening respiratory tract obstruction, and oedema of the intestinal mucosa may cause abdominal pain and distension.

In most cases, the underlying mechanism is degranulation of mast cells. However, angioedema may occasionally be mediated by increased local bradykinin concentration (Box 4.22). Differentiating the mechanism of angioedema is important in determining appropriate investigations and treatment.

### 4.22 Types of angioedema

<table>
<thead>
<tr>
<th>Pathogenesis</th>
<th>Allergic reaction to specific trigger</th>
<th>Idiopathic angioedema</th>
<th>Hereditary angioedema</th>
<th>ACE-inhibitor associated angioedema</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgE-mediated degranulation of mast cells</td>
<td>IgE-mediated degranulation of mast cells</td>
<td>Non-IgE-mediated degranulation of mast cells</td>
<td>C1 inhibitor deficiency, with resulting increased local bradykinin concentration</td>
<td>Inhibition of breakdown of bradykinin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Key mediator</th>
<th>Histamine</th>
<th>Histamine</th>
<th>Bradykinin</th>
<th>Bradykinin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence</td>
<td>Common</td>
<td>Common</td>
<td>Rare autosomal dominant disorder</td>
<td>0.1–0.2% of patients treated with ACE inhibitors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical features</th>
<th>Usually associated with urticaria</th>
<th>Usually associated with urticaria</th>
<th>Not associated with urticaria or other features of allergy</th>
<th>Not associated with urticaria or other factors of allergy</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of other allergies common</td>
<td>May be triggered by physical stimuli, such as heat, pressure or exercise Dermatographism common Occasionally associated with infection or thyroid disease</td>
<td>Usually associated with urticaria or other features of allergy</td>
<td>Does not cause anaphylaxis May cause life-threatening respiratory tract obstruction Can cause severe abdominal pain</td>
<td>Does not cause anaphylaxis Usually affects the head and neck, and may cause life-threatening respiratory tract obstruction Can occur years after the start of treatment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investigations</th>
<th>Specific IgE tests or skin prick tests</th>
<th>Specific IgE tests and skin prick tests often negative Exclude hypothyroidism</th>
<th>Complement C4 (invariably low in acute attacks) C1 inhibitor levels</th>
<th>No specific investigations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Allergen avoidance Antihistamines</td>
<td>Antihistamines are mainstay of treatment and prophylaxis</td>
<td>Unresponsive to antihistamines Attenuated androgens C1 inhibitor concentrate or icatibant for acute attacks</td>
<td>Discontinue ACE inhibitor Avoid angiotensin II receptor blockers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible drug causes</th>
<th>Specific drug allergies, e.g. penicillin</th>
<th>NSAIDs</th>
<th>Opioids</th>
<th>Radiographic media</th>
<th>ACE inhibitors Angiotensin II receptor blockers</th>
</tr>
</thead>
</table>

ACE = angiotensin-converting enzyme; NSAIDs = non-steroidal anti-inflammatory drugs.
**Specific allergies**

### Insect venom allergy

Local non-IgE-mediated reactions to insect stings are common and may cause extensive swelling around the site lasting as long as 7 days. These usually do not require specific treatment. Toxic reactions to venom after multiple (50–100) simultaneous stings may mimic anaphylaxis. In addition, exposure to large amounts of insect venom frequently stimulates the production of IgE antibodies, and thus may be followed by allergic reactions to single stings. Allergic IgE-mediated reactions vary from mild to life-threatening. Antigen-specific immunotherapy with bee or wasp venom reduces the incidence of recurrent anaphylaxis from 50–60% to 10% but requires treatment for several years (see Box 4.17).

### Peanut allergy

Peanut allergy is the most common food-related allergy. More than 50% of patients present before the age of 3 years and some individuals react to their first known exposure to peanuts, possibly because of sensitisation by topical creams. Peanuts are ubiquitous in the Western diet, and every year up to 25% of peanut-allergic individuals will experience a reaction as a result of inadvertent exposure.

### Birch oral allergy syndrome

This syndrome is characterised by a combination of birch pollen hay fever and local angioedema after contact with fresh fruit (especially apples), vegetables and nuts. Cooked fruits and vegetables are tolerated without difficulty. It is due to shared or cross-reactive allergens which are destroyed by cooking or digestion, and can be confirmed by skin prick testing using fresh fruit. Severe allergic reactions are unusual.

### C1 inhibitor deficiency

#### Hereditary angioedema

Hereditary angioedema (HAE), also known as inherited C1 inhibitor deficiency, is an autosomal dominant disorder caused by decreased production or activity of C1 inhibitor protein. This complement regulatory protein inhibits spontaneous activation of the classical complement pathway (see Fig. 4.3, p. 75). C1 inhibitor is also a regulatory protein for the kinin cascade, activation of which increases local bradykinin levels and gives rise to local pain and swelling.

In HAE, angioedema may be spontaneous or triggered by local trauma or infection. Multiple parts of the body may be involved, especially the face, extremities, upper airway and gastrointestinal tract. Oedema of the intestinal wall causes severe abdominal pain. The most important complication is laryngeal obstruction, often associated with minor dental procedures, which can be fatal. Episodes of angioedema are self-limiting and usually resolve within 48 hours. Patients with HAE generally present in adolescence, but may go undiagnosed for many years. A family history can be identified in 80% of cases. HAE is not associated with allergic diseases and is specifically not associated with urticaria.

### Acquired C1 inhibitor deficiency

Acute episodes are always accompanied by low C4 levels and the diagnosis can be confirmed by C1 inhibitor measurement. Prevention is with modified androgens (e.g. danazol), which increase endogenous production of complement proteins. Severe acute attacks should be treated with purified C1 inhibitor or a bradykinin receptor antagonist (icatibant).

#### Autoimmune disease

- Suppressed T cell-mediated immune responses in pregnancy: autoimmune diseases often improve during pregnancy but flare immediately after delivery. However, an exception is SLE, which is prone to exacerbation in pregnancy.
- Passive transfer of maternal antibodies: can mediate autoimmune disease in the fetus and newborn, including SLE, Graves’ disease and myasthenia gravis.
- Antiphospholipid syndrome (p. 1055): an important cause of fetal loss, intrauterine growth restriction and pre-eclampsia.

### Transplantation immunology

Transplantation provides the opportunity for definitive treatment of end-stage organ disease. The major complications are graft rejection, drug toxicity and infection consequent on immunosuppression. Transplant survival continues to improve, as a result of the introduction of less toxic immunosuppressive agents and increased understanding of rejection mechanisms.

#### Transplant rejection

Solid organ transplantation inevitably stimulates an aggressive immune response by the recipient, unless the transplant is between monozygotic twins. The type and severity of the rejection response are determined by the genetic disparity between the donor and recipient, the immune status of the host and the nature of the tissue
transplanted (Box 4.24). The most important genetic determinant is the difference between donor and recipient HLA proteins (p. 75). The extensive polymorphism of these proteins means that donor HLA antigens are almost invariably recognised as foreign by the recipient immune system, unless an active attempt has been made to minimise incompatibility.

- **Acute cellular rejection** is the most common form of graft rejection. It is mediated by activated T cells and results in deterioration in graft function. It may cause fever, pain and tenderness over the graft. It is usually amenable to increased immunosuppressive therapy.

- **Hyperacute rejection** results in rapid and irreversible destruction of the graft. It is mediated by pre-existing recipient antibodies against donor HLA antigens, which arise as a result of previous exposure through transplantation, blood transfusion or pregnancy. It is rarely seen in practice, as the use of screening for anti-HLA antibodies and pre-transplant cross-matching ensure prior identification of recipient–donor incompatibility.

- **Acute vascular rejection** is mediated by antibody formed de novo after transplantation. It is more curtailed than the hyperacute response because of the use of intercurrent immunosuppression, but it is also associated with reduced graft survival. Aggressive immunosuppressive therapy is indicated, and physical removal of antibody through plasmapheresis may be effective. Not all post-transplant anti-donor antibodies cause graft damage; their consequences are determined by specificity and ability to trigger other immune components, such as the complement cascade.

- **Chronic allograft failure**, also known as chronic rejection, is a major cause of graft loss. It is associated with proliferation of transplant vascular smooth muscle, interstitial fibrosis and scarring. The pathogenesis is poorly understood but contributing factors include immunological damage caused by subacute rejection, hypertension, hyperlipidaemia and chronic drug toxicity.

**Investigations**

**Pre-transplantation testing**

HLA typing determines an individual’s HLA polymorphisms and facilitates donor-recipient matching.

**Complications of transplant immunosuppression**

The prevention of transplant rejection requires indefinite treatment with immunosuppressive agents. In general, two or more immunosuppressive drugs are used in synergistic combinations in order to minimise drug side-effects (Box 4.25). The major complications of long-term immunosuppression are infection and malignancy.

The risk of some opportunistic infections may be minimised through the use of prophylactic medication (e.g. ganciclovir for CMV prophylaxis and trimethoprim–sulfamethoxazole for *Pneumocystis* prophylaxis). Immunisation with killed vaccines is appropriate, although the immune response may be curtailed. Live vaccines should not be given.

The increased risk of malignancy arises because T-cell suppression results in failure to control viral infections. Virus-associated tumours include lymphoma (associated with EBV), Kaposi’s sarcoma (associated with human herpesvirus 8) and skin tumours (associated with human papillomavirus). Immunosuppression is also associated with a small increase in the incidence of common cancers not associated with viral infection.
Immunological factors in disease

<table>
<thead>
<tr>
<th>Drug</th>
<th>Mechanism of action</th>
<th>Major adverse effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-proliferative agents</td>
<td>Inhibit lymphocyte proliferation by blocking DNA synthesis</td>
<td>Increased susceptibility to infection, Leucopenia, Hepatotoxicity</td>
</tr>
<tr>
<td>e.g. azathioprine, mycophenolate mofetil</td>
<td>May be directly cytotoxic at high doses</td>
<td></td>
</tr>
<tr>
<td>Calcineurin inhibitors</td>
<td>Inhibit T-cell signalling; prevent lymphocyte activation and block cytokine transcription</td>
<td>Increased susceptibility to infection, Hypertension, Nephrotoxicity, Diabetogenic (especially tacrolimus), Gingival hypertrophy, hirsutism (ciclosporin)</td>
</tr>
<tr>
<td>e.g. ciclosporin, tacrolimus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corticosteroids</td>
<td>Decrease phagocytosis and release of proteolytic enzymes; decrease lymphocyte activation and proliferation; decrease cytokine production; decrease antibody production</td>
<td>Increased susceptibility to infection, Multiple other complications (p. 776)</td>
</tr>
<tr>
<td>Anti-T-cell induction agents</td>
<td>Antibodies to cell surface proteins deplete or inhibit T cells</td>
<td>Profound non-specific immunosuppression, Increased susceptibility to infection</td>
</tr>
<tr>
<td>e.g. anti-thymocyte globulin (ATG)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(such as lung, breast and colon cancer), reflecting the importance of T cells in anti-cancer surveillance.

Organ donation

The major problem in transplantation is the shortage of organ donors. Cadaveric organ donors are usually previously healthy individuals who experience brainstem death (p. 1161), frequently as a result of road traffic accidents or cerebrovascular events. However, even if organs were obtained from all potential cadaveric donors, their numbers would be insufficient to meet current needs. An alternative is the use of living donors. Altruistic living donation, usually from close relatives, is widely used in renal transplantation. Living organ donation is inevitably associated with some risk to the donor, and the process is highly regulated to ensure appreciation of the dangers involved. Because of concerns about coercion and exploitation, non-altruistic organ donation (the sale of organs) is illegal in most countries.

Further information and acknowledgements

Websites

- [www.anaphylaxis.org.uk](http://www.anaphylaxis.org.uk) Provides information and support for patients with severe allergies.
- [www.immunopaedia.org](http://www.immunopaedia.org) A South African site designed for health-care providers requiring a general understanding of immunology, providing clinical case studies, articles, links and news, with a particular focus on HIV immunology.
- [www.info4pi.org](http://www.info4pi.org) A US site managed by the non-profit Jeffrey Modell Foundation, which provides extensive information about primary immunodeficiency diseases.

Figure acknowledgements

Fig. 4.11 Helbert M. Flesh and bones of immunology. Edinburgh: Churchill Livingstone; 2006; copyright Elsevier.
Principles and investigation of environmental factors in disease 98

Environmental effects on health 98
Investigations in environmental health 99
Preventive medicine 100

Environmental diseases 100
Alcohol 100
Smoking 100
Obesity 101
Poverty and affluence 101
Atmospheric pollution 102
Radiation exposure 102
Extremes of temperature 103
High altitude 106
Under water 108

Nutritional factors and disease 110
Physiology of nutrition 110
Clinical assessment and investigation of nutritional status 114

Disorders of altered energy balance 115
Obesity 115
Under-nutrition 120

Micronutrients, minerals and their diseases 124
Vitamins 124
Inorganic micronutrients 130
Health emerges from a highly complex interaction between factors intrinsic to the patient and his or her environment. Many factors within the environment influence health, including aspects of the physical environment, biological environment (bacteria, viruses), built environment and social environment, but these also encompass more distant influences such as the global ecosystem (Fig. 5.1). Environmental changes affect many physiological systems and do not respect boundaries between medical specialties. The specialty of ‘public health’ in the UK is concerned with the investigation and management of health in communities and populations, but the principles apply in all specialties.

Exposure to infectious agents is a major environmental determinant of health and is described in detail in Chapter 6. This chapter describes the approach to other common environmental factors that influence health.

The hierarchy of systems – from molecules to ecologies

When assessing a patient, a clinician subconsciously considers many levels at which problems may be occurring, including molecular, cellular, tissue, organ and body systems. When the environment’s influence on health is being considered, this ‘hierarchy of systems’ extends beyond the individual to include the family, community, population and ecology. Box 5.1 shows an example of the utility of this concept in describing determinants of coronary heart disease operating at each level of a hierarchy.

Interactions between people and their environment

The hierarchy of systems demonstrates that the clinician should not focus too quickly on the disease process without considering the context. Health is an emergent quality of a complex interaction between many determinants, including genetic inheritance, the physical circumstances in which people live (e.g. housing, air quality, working environment), the social environment (e.g. levels of friendship, support and trust), personal behaviour (smoking, diet, exercise), and access to money and the other resources that give people control over their lives. Health care is not the only determinant – and is usually not the major determinant – of health status in the population.

These systems do not operate in isolation in separate communities. When one group responds to ill health by manipulating its environment, the consequences may be global. For example, an Afghan farmer who starts growing opium for money in order to feed his children influences the environment of a teenager in Europe; in turn, drug misuse in Europe has fostered higher prevalence of blood-borne infectious diseases such as human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS); in turn, these have spilled out into sexually transmitted disease. This process contributes significantly to the tragedy of the epidemic of HIV/AIDS.

The life course

The determinants of health operate over the whole lifespan. Values and behaviours acquired during childhood and adolescence have a profound influence on
Influences on health can even operate before birth. Individuals with low birth weight have been shown to have a higher prevalence of conditions such as hypertension and type 2 diabetes as young adults and of cardiovascular disease in middle age. It has been suggested that under-nutrition during middle to late gestation permanently ‘programmes’ cardiovascular and metabolic responses.

This ‘life course’ perspective highlights the cumulative effect on health of exposures to episodes of illness, adverse environmental conditions and behaviours that damage health. In this way, biological and social risk factors at each stage of life link to form pathways to disease and health.

### Investigations in environmental health

#### Incidence and prevalence

The first task is to establish how common a problem is within the population. This is expressed in two ways (Box 5.2).

- If the problem is a continuing condition (e.g., enlarged spleen due to malaria), then prevalence is the appropriate measurement and is calculated by dividing the number of people with the condition at a specified time by the number of people in the population at risk at that time. Prevalence tends to be higher if the problem is common (many new cases) and/or if it is of longer duration.

- If the problem is an event that occurs at a clear point in time (e.g., fever due to malaria), then incidence is used. Incidence is a measure of the rate at which new cases occur (e.g., confirmed pyrexia with malaria parasites on a blood film) in the population at risk during a defined period of time.

#### Variability by time, person and place

The next task is to establish how the problem varies in terms of time, person and place. The incidence may fluctuate throughout the year; for example, malaria occurs in the wet season but not the dry. Observation over longer periods establishes whether a problem is becoming more or less common: malaria may re-emerge due to drug resistance. The next questions are, who are the victims? Are males or females more commonly affected? What is the age pattern? What are the occupations and social positions of those affected? In this example, symptomatic malaria is more common in poorer, rural-dwelling children. Finally, there is the question of variability by place: the prevalence of malaria is dictated by the distribution of *Anopheles* mosquitoes.

### Measuring risk

Epidemiology is also concerned with the numerical estimation of risk. This is best illustrated by a simple example. In a rural African town with a population of 5000, disease ‘d’ is under investigation. The majority of the cases of disease ‘d’ (300 out of 360) occurred among women and children who use the river, which recently had its flow of water reduced because of a new irrigation scheme. A formal experiment is established to measure risk. The 1000 women and children who use the river are followed up for 1 year and compared to a cohort with a similar age and sex distribution who use stand-pipes as their source of water.

The incidence (new cases) of disease ‘d’ in the 1000 exposed to risk ‘r’ (river water) was 300. The incidence (new cases) of disease ‘d’ in the 1000 not exposed to risk ‘r’ was 60. The relative risk is the incidence in the exposed population (300 per 1000 per year) divided by the incidence in the non-exposed population (60 per 1000 per year); 300/60 = 5, meaning that those exposed to the river water are 5 times more likely to contract the disease – their relative risk is 5. The attributable risk of exposure ‘r’ for disease ‘d’ is the incidence in the exposed population (300) minus the incidence in the non-exposed population (60), which is 240 per 1000 per year. The fraction, or proportion, of the disease in the exposed population which can be attributed to risk (r) is called the attributable fraction, in this case (300−60)/300 = 0.8. This means that 80% of the disease can be attributed to exposure to river water.

### Establishing cause and effect

Associations between a risk factor and a disease do not prove that the risk factor causes the disease. In the northern hemisphere, both multiple sclerosis and blue eyes are more common but it is implausible that having blue eyes is the cause of multiple sclerosis. Cause and effect can only be proven by more detailed investigation. In the above example, further investigation of the river water will be needed, using the criteria for causation defined in Koch’s postulates (for infectious agents, p. 134) or the more generic Bradford Hill criteria.
Preventive medicine

There are many examples of epidemiological associations defining causative factors in disease, e.g. the association between cigarette smoking and lung cancer (p. 699). However, as illustrated above, the complexity of the interactions between physical, social and economic determinants of health means that successful prevention is often difficult. Moreover, the life course perspective illustrates that it may be necessary to intervene early in life or even before birth, to prevent important disease in later life. Successful prevention is likely to require many interventions across the life course and at several levels in the hierarchy of systems. The examples below illustrate this principle.

ENVIRONMENTAL DISEASES

The term ‘homeostasis’ describes the capacity to maintain the internal milieu by adapting to increases or decreases in a given environmental factor. However, there are limits to the coping abilities of any system, at which ‘too much’ or ‘too little’ of a given environmental factor will result in ill health. Too many calories lead to obesity, while too few lead to malnutrition. Either involuntarily or deliberately, we expose ourselves to many poisons and hazards. Examples discussed elsewhere include industrial/occupational hazards, such as asbestos (p. 718) and other carcinogens (p. 266). ‘Social’ poisons, such as tobacco, alcohol and drugs of misuse, also need to be considered (p. 240).

Alcohol

The World Health Organization (WHO) estimates that the harmful use of alcohol results in the death of 2.5 million people annually. Rates of alcohol-related harm vary by place and time but have risen dramatically in the UK, with Scotland showing the highest rates. (Fig. 5.2 demonstrates the climbing rates during the 1990s, since when rates have stabilised at very high levels.) Why did Scotland experience this dramatic increase in alcohol deaths? The most likely explanation is that the environment changed. The price of alcohol fell in real terms and availability increased (more supermarkets sold alcohol and the opening times of public houses were extended). Also, the culture changed in a way that fostered higher levels of consumption and more binge drinking. These changes have caused a trebling of male and a doubling of female deaths due to alcohol. Public, professional and governmental concern has now led to a minimum price being charged for a unit of alcohol, tightening of licensing regulations and curtailment of some promotional activity (e.g. two-for-one offers in bars). Many experts judge that even more aggressive public health measures will be needed to reverse the levels of harm in the community. The approach for individual patients suffering adverse effects of alcohol is described on pages 240 and 252.

Smoking

Smoking tobacco dramatically increases the risk of developing many diseases. It is responsible for a substantial majority of cases of lung cancer and chronic obstructive pulmonary disease, and most smokers die either from these respiratory diseases or from ischaemic heart disease. Smoking also causes cancers of the upper respiratory and gastrointestinal tracts, pancreas, bladder and kidney, and increases risks of peripheral vascular disease, stroke and peptic ulceration. Maternal smoking is an important cause of fetal growth retardation. Moreover, there is increasing evidence that passive (or ‘secondhand’) smoking has adverse effects on cardiovascular and respiratory health.

When the ill-health effects of smoking were first discovered, doctors imagined that warning people about the dangers of smoking would result in them giving up. However, it also took increased taxation of tobacco, banning of advertising and support for smoking cessation to maintain a decline in smoking rates. In several European countries (including the UK), this has culminated in a complete ban on smoking in all public places – legislation that only became possible as the public became convinced of the dangers of secondhand smoke. However, smoking rates remain high in many poorer areas and are increasing amongst young women. In many developing countries, tobacco companies have found new markets and rates are rising. Worldwide, there are approximately 1 billion smokers, and it is estimated by WHO that 6 million die prematurely each year as a result of their habit.

In reality, there is a complex hierarchy of systems that interact to cause smokers to initiate and maintain their habit. At the molecular and cellular levels, nicotine acts on the nervous system to create dependence, so that smokers experience unpleasant effects when they attempt to quit. So, even if they know it is harmful, the role of addiction in maintaining the habit is important. Influences at the personal and social level are just as important. Many individuals bolster their denial of the harmful effects of smoking by focusing on someone they knew personally who smoked until he or she was very old and died peacefully in bed. Such strong counterexamples help smokers to maintain internal beliefs that comfort them when presented with statistical evidence.

Fig. 5.2 Alcohol-related deaths in Scotland by year and sex (1990–2003). Principal (‘underlying’) and secondary (‘contributing’) causes of death. (Source: General Register Office (Scotland))
Young female smokers are often motivated more by the desire to ‘stay thin’ or ‘look cool’ than to avoid an illness in middle life.

Even if a smoker decides to quit, there are a variety of influences in the wider environment that reduce the chances of sustained success, including peer pressure, cigarette advertising, and finding oneself in circumstances where one previously smoked. The tobacco industry works very hard to maintain and expand the smoking habit, and its advertising budget is much greater than that available to health promoters.

Strategies to help individuals quit smoking are outlined in Boxes 5.3 and 5.4. Although the success rates are modest, these interventions are cost-effective and form an important part of the overall anti-tobacco strategy.

### Poverty and affluence

The adverse health and social consequences of poverty are well documented: high birth rates, high death rates and short life expectancy (Box 5.5). Typically, with malnourished and underweight. It would, however, be wrong to focus only on those who are obese because, in countries like the USA and the UK, fat deposition is affecting almost the entire population. The weight distribution of almost the whole population is shifting upwards – the slim are becoming less slim while the fat are getting fatter. In the UK, this translates into a 1-kilogram increase in weight per adult per year (on average over the adult population). It is now widely accepted that we cannot blame the current obesity epidemic on individual behaviour and poor choice, although many current approaches still focus on individuals. The best way, therefore, to understand the current obesity epidemic is to consider humans as ‘obesogenic organisms’ who, for the first time in their history, find themselves in an obesogenic environment – that is, one where people’s circumstances encourage them to eat more and exercise less. This includes the availability of cheap and heavily marketed energy-rich foods, the increase in labour-saving devices (e.g. lifts and remote controls) and the increase in passive transport (cars as opposed to walking, cycling, or walking to public transport hubs). Our physiology was formed a long time ago when food was scarce and we needed large amounts of energy in order to find food and stay alive. We are stuck with the metabolic and behavioural legacy of our evolutionary history – we are organisms that are programmed to eat when we can and preserve energy whenever possible. It is not surprising that we have problems coping with an environment that exerts constant pressure to increase energy intake and to decrease energy expenditure. The rise in obesity suggests that the effects of our obesogenic environment are overriding the biological regulatory mechanisms in more and more people. To combat the health impact of obesity, therefore, we need to help those who are already obese but also develop strategies that impact on the whole population and reverse the obesogenic environment.

### Obesity

Obesity is a condition characterised by an excess of body fat. In its simplest terms, obesity can be considered to result from an imbalance between the amount of energy consumed in the diet and the amount of energy expended through exercise and bodily functions. People who are obese are more likely to develop a range of chronic conditions. In 2006, the number of obese and overweight people in the world overtook the numbers who are

---

**5.3 Methods for smoking cessation**

<table>
<thead>
<tr>
<th>Smokers who are not motivated to try to stop smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Record smoking status at regular intervals</td>
</tr>
<tr>
<td>• Anti-smoking advice</td>
</tr>
<tr>
<td>• Encourage change in attitude towards smoking to improve motivation</td>
</tr>
</tbody>
</table>

**Motivated light smokers (< 10/day)**

- Anti-smoking advice
- Anti-smoking support programme

**Motivated heavy smokers (10–15/day)**

- As above plus nicotine replacement therapy (NRT) (minimum 8 weeks)

**Motivated heavy smokers (> 15/day)**

- As above plus bupropion if NRT and behavioural support are unsuccessful and patient remains motivated


**5.4 Smoking cessation**

‘Placebo or will-power alone has a ~2% chance of abstinence for ≥ 6 months. This can be increased by the percentage shown by:

- written self-help materials: 1%
- opportunistic advice from doctor: 2%
- face-to-face behavioural support from specialist: 4–7%
- proactive telephone counselling: 2%
- NRT with limited or intensive behavioural support: 5–12%
- bupropion with intensive behavioural support: 9%.’


---

**5.5 Examples of effects of financial resources on health**

<table>
<thead>
<tr>
<th>Poverty</th>
<th>Affluence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory disease exacerbated or caused by air pollution</td>
<td>Physical inactivity</td>
</tr>
<tr>
<td>Exposure to unnecessary hazards in the workplace or living environment</td>
<td>Alcohol and drug consumption</td>
</tr>
<tr>
<td>Poor hygiene causing diarrhoeal diseases and debilitating intestinal parasitic infections</td>
<td>High rates of suicide, depression, anxiety and stress</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>Sexually transmitted infection</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Obesity</td>
</tr>
</tbody>
</table>
industrialisation, the pattern changes: low birth rates, low death rates and longer life expectancy (Box 5.6). Instead of infections, chronic conditions such as heart disease dominate in an older population. Adverse health consequences of excessive affluence are also becoming apparent. Despite experiencing sustained economic growth for the last 50 years, people in many industrialised countries are not growing any happier and the litany of socioeconomic problems – crime, congestion, inequality – persists. Living in societies that give pride of place to economic growth means that there is constant pressure to contribute by performing ever harder at work and by consuming as much as – or more than – we can afford. As a result, people become stressed and may adopt unhealthy strategies to mitigate their discomfort; they overeat, overshop, or use sex or drugs (legal and illegal) as ‘pain-killers’. These behaviours often lead to the problems listed in Box 5.5.

Many countries are now experiencing a ‘double burden’. They have large populations still living in poverty who are suffering from problems such as diarrhoea and malnutrition, alongside affluent populations (often in cities) who suffer from chronic illness such as diabetes and heart disease. Recent research suggests that uneven distribution of wealth is a more important determinant of health than the absolute level of wealth; countries with a more even distribution of wealth enjoy longer life expectancies than countries with similar or higher gross domestic products (GDPs) but wider distributions of wealth.

Atmospheric pollution

Emissions from industry, power plants and motor vehicles of sulphur oxides, nitrogen oxides, respirable particles and metals are severely polluting cities and towns in Asia, Africa, Latin America and Eastern Europe. Increased death rates from respiratory and cardiovascular disease occur in vulnerable adults, such as those with established respiratory disease and the elderly, while children experience an increase in bronchitic symptoms. In nations like the UK that have reduced their primary emissions, the new issue of greenhouse gases has emerged. Developing countries also suffer high rates of respiratory disease as a result of indoor pollution caused mainly by heating and cooking combustion.

Carbon dioxide and global warming

Climate change is arguably the world’s most important environmental health issue. A combination of increased production of carbon dioxide and habitat destruction, both caused primarily by human activity, seems to be the main cause. The temperature of the globe is rising, climate is being affected, and if the trend continues, sea levels will rise and rainfall patterns will be altered so that both droughts and floods will become more common. These have already claimed millions of lives during the past 20 years and have adversely affected the lives of many more. The economic costs of property damage and the impact on agriculture, food supplies and prosperity have also been substantial. The health impacts of global warming will also include changes in the geographical range of some vector-borne infectious diseases.

Currently, politicians cannot agree on an effective framework of actions to tackle the problem. Meanwhile, the industrialised world continues with lifestyles and levels of waste that are beyond the planet’s ability to sustain. Rapidly growing economies in the world’s two most populous states, India and China, are going to be a vital part of the unfolding problem or solution.

Radiation exposure

Radiation includes ionising (Box 5.7) and non-ionising radiations (ultraviolet (UV), visible light, laser, infrared and microwave). Whilst global industrialisation and the generation of fluorocarbons have raised concerns about loss of the ozone layer, leading to an increased exposure to UV rays, and disasters such as the Chernobyl and Fukushima nuclear power station explosions have demonstrated the harm of ionising radiation, it is important to remember that it can be harnessed for medical benefit. Ionising radiation is used in X-rays, computed tomography (CT), radionuclide scans and radiotherapy, and non-ionising UV for therapy in skin diseases and laser therapy for diabetic retinopathy.

Types of ionising radiation

These include charged subatomic alpha and beta particles, uncharged neutrons or high-energy electromagnetic

\[\text{\begin{tabular}{|c|c|c|}
\hline
\textbf{Range in air} & \textbf{Range in tissue} & \textbf{Protection} \\
\hline
\text{Alpha particles} & Few centimetres & No penetration & Paper \\
\hline
\text{Beta particles} & Few metres & Few millimetres & Aluminium sheet \\
\hline
\text{X-rays/ gamma rays} & Kilometres & Passes through & Lead \\
\hline
\text{Neutrons} & Kilometres & Passes through & Concrete or thick polythene \\
\hline
\end{tabular}}\]
radiations such as X-rays and gamma rays. When they interact with atoms, energy is released and the resulting ionisation can lead to molecular damage. The clinical effects of different forms of radiation depend upon their range in air and tissue penetration (see Box 5.7).

**Dosage and exposure**

The dose of radiation is based upon the energy absorbed by a unit mass of tissue and is measured in grays (Gy), with 1 Gy representing 1 J/kg. To take account of different types of radiation and variations in the sensitivity of various tissues, weighting factors are used to produce a unit of effective dose, measured in sieverts (Sv). This value reflects the absorbed dose weighted for the damaging effects of a particular form of radiation and is most valuable in evaluating the long-term effects of exposure.

‘Background radiation’ refers to our exposure to naturally occurring radioactivity (e.g. radon gas and cosmic radiation). This produces an average annual individual dose of approximately 2.6 mSv per year, although this varies according to local geology.

**Effects of radiation exposure**

Effects on the individual are classified as either deterministic or stochastic.

**Deterministic effects**

Deterministic (threshold) effects occur with increasing severity as the dose of radiation rises above a threshold level. Tissues with actively dividing cells, such as bone marrow and gastrointestinal mucosa, are particularly sensitive to ionising radiation. Lymphocyte depletion is the most sensitive marker of bone marrow injury, and after exposure to a fatal dose, marrow aplasia is a common cause of death. However, gastrointestinal mucosal toxicity may cause earlier death due to profound diarrhoea, vomiting, dehydration and sepsis. The gonads are highly radiosensitive and radiation may result in temporary or permanent sterility. Eye exposure can lead to cataracts and the skin is susceptible to radiation burns. Irradiation of the lung and central nervous system may induce acute inflammatory reactions, pulmonary fibrosis and permanent neurological deficit respectively. Bone necrosis and lymphatic fibrosis are characteristic following regional irradiation, particularly for breast cancer. The thyroid gland is not inherently sensitive but its ability to concentrate iodine makes it susceptible to damage after exposure to relatively low doses of radioactive iodine isotopes, such as were released from Chernobyl.

**Stochastic effects**

Stochastic (chance) effects occur with increasing probability as the dose of radiation increases. Carcinogenesis represents a stochastic effect. With acute exposures, leukemias may arise after an interval of around 2–5 years and solid tumours after an interval of about 10–20 years. Thereafter the incidence rises with time. An individual’s risk of developing cancer depends on the dose received, the time to accumulate the total dose and the interval following exposure.

**Management of radiation exposure**

The principal problems after large-dose exposures are maintenance of adequate hydration, control of sepsis and the management of marrow aplasia. Associated injuries such as thermal burns need specialist management within 48 hours of active resuscitation. Internal exposure to radioisotopes should be treated with chelating agents (such as Prussian blue used to chelate ¹³¹I caesium after ingestion). White cell colony stimulation and haematopoietic stem cell transplantation may need to be considered for marrow aplasia.

**Extremes of temperature**

**Thermoregulation**

Body heat is generated by basal metabolic activity and muscle movement, and lost by conduction (which is more effective in water than in air), convection, evaporation and radiation (most important at lower temperatures when other mechanisms conserve heat) (Box 5.8). Body temperature is controlled in the hypothalamus, which is directly sensitive to changes in core temperature and indirectly responds to temperature-sensitive neurons in the skin. The normal ‘set-point’ of core

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Hot environment</th>
<th>Cold environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heat production</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basal metabolic</td>
<td>→</td>
<td>↓ in hypothermia</td>
</tr>
<tr>
<td>Muscle activity</td>
<td>↓ by lethargy</td>
<td>↑ by shivering</td>
</tr>
<tr>
<td></td>
<td></td>
<td>↓ in severe hypothermia</td>
</tr>
<tr>
<td>Heat loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduction*</td>
<td>↑ by vasodilatation</td>
<td>↓ by vasoconstriction</td>
</tr>
<tr>
<td></td>
<td>↑↑ by sweating</td>
<td>↑↑ in water &lt; 31°C</td>
</tr>
<tr>
<td>Convection*</td>
<td>↑ by vasodilatation</td>
<td>↑ by vasoconstriction</td>
</tr>
<tr>
<td></td>
<td>↓ by lethargy</td>
<td>↑ by wind and movement</td>
</tr>
<tr>
<td>Evaporation*</td>
<td>↑↑ by sweating</td>
<td>↑ by hyperventilation</td>
</tr>
<tr>
<td></td>
<td>↓ by high humidity</td>
<td></td>
</tr>
<tr>
<td>Radiation</td>
<td>↑ by vasodilatation</td>
<td>↓ by vasoconstriction (but is the major heat loss in dry cold)</td>
</tr>
</tbody>
</table>

*These losses are dependent on the relative ambient and skin temperatures.
temperature is tightly regulated within 37 ± 0.5°C, which is necessary to preserve the normal function of many enzymes and other metabolic processes. The temperature set-point is increased in response to infection (p. 296).

In a cold environment, protective mechanisms include cutaneous vasoconstriction and shivering; however, any muscle activity that involves movement may promote heat loss by increasing convective loss from the skin, and respiratory heat loss by stimulating ventilation. In a hot environment, sweating is the main mechanism for increasing heat loss. This usually occurs when the ambient temperature rises above 32.5°C or during exercise.

**Hypothermia**

Hypothermia exists when the body’s normal thermal regulatory mechanisms are unable to maintain heat in a cold environment and core temperature falls below 35°C (Fig. 5.3).

Whilst infants are susceptible to hypothermia because of their poor thermoregulation and high body surface area to weight ratio, it is the elderly who are at highest risk (Box 5.9). Hypothyroidism is often a contributory factor in old age, while alcohol and other drugs (e.g. phenothiazines) commonly impede the thermoregulatory response in younger people. More rarely, hypothermia is secondary to glucocorticoid insufficiency, stroke, hepatic failure or hypoglycaemia.

Hypothermia also occurs in healthy individuals whose thermoregulatory mechanisms are intact but insufficient to cope with the intensity of the thermal stress. Typical examples include immersion in cold water, when core temperature may fall rapidly (acute hypothermia), exposure to extreme climates such as during hill walking (subacute hypothermia), and slow-onset hypothermia, as develops in an immobilised older individual (subchronic hypothermia). This classification is important, as it determines the method of rewarming.

**Clinical features**

Diagnosis is dependent on recognition of the environmental circumstances and measurement of core (rectal) body temperature. Clinical features depend on the degree of hypothermia (see Fig. 5.3).

In a cold patient, it is very difficult to diagnose death reliably by clinical means. It has been suggested that, in extreme environmental conditions, irreversible hypothermia is probably present if there is asystole (no carotid pulse for 1 minute), the chest and abdomen are rigid, the core temperature is below 13°C and serum potassium is > 12 mmol/L. However, in general, resuscitative measures should continue until the core temperature is normal and only then should a diagnosis of brain death be considered (p. 1161).

**Investigations**

Blood gases, a full blood count, electrolytes, chest X-ray and electrocardiogram (ECG) are all essential investigations. Haemoconcentration and metabolic acidosis are common, and the ECG may show characteristic J waves, which occur at the junction of the QRS complex and the ST segment (Fig. 5.4). Cardiac dysrhythmias, including ventricular fibrillation, may occur. Although the arterial oxygen tension may be normal when measured at room
temperature, the arterial PO₂ in the blood falls by 7% for each 1°C fall in core temperature. Serum aspartate aminotransferase and creatine kinase may be elevated secondary to muscle damage and the serum amylase is often high due to subclinical pancreatitis. If the cause of hypothermia is not obvious, additional investigations for thyroid and pituitary dysfunction (p. 737), hypoglycaemia (p. 807) and the possibility of drug intoxication (p. 209) should be performed.

**Management**

Following resuscitation, the objectives of management are to rewarm the patient in a controlled manner while treating associated hypoxia (by oxygenation and ventilation if necessary), fluid and electrolyte disturbance, and cardiovascular abnormalities, particularly dysrhythmias. Careful handling is essential to avoid precipitating the latter. The method of rewarming is dependent not on the absolute core temperature, but on haemodynamic stability and the presence or absence of an effective cardiac output.

**Mild hypothermia**

Outdoors, continued heat loss is prevented by sheltering the patient from the cold, replacing wet clothing, covering the head and insulating him or her from the ground. Once in hospital, even in the presence of profound hypothermia, if there is an effective cardiac output then forced-air rewarming, heat packs placed in axilla, groin and around the abdomen, inhaled warmed air and correction of fluid and electrolyte disturbances are usually sufficient. Rewarming rates of 1–2°C per hour are effective in leading to a gradual and safe return to physiological normality. Underlying conditions should be treated promptly (e.g. hypothyroidism with triiodothyronine 10 µg IV 3 times daily; p. 743).

**Severe hypothermia**

In the case of severe hypothermia with cardiopulmonary arrest (non-perfusing rhythm), the aim is to restore perfusion, and rapid rewarming at a rate greater than 2°C per hour is required. This is best achieved by cardiopulmonary bypass or extracorporeal membrane oxygenation. If these are unavailable, then veno-veno haemofiltration, and pleural, peritoneal, thoracic or bladder lavage with warmed fluids are alternatives. Monitoring of cardiac rhythm and arterial blood gases, including H⁺ (pH) is essential. Significant acidosis may require correction (p. 445).

**Cold injury**

**Freezing cold injury (frostbite)**

This represents the direct freezing of body tissues and usually affects the extremities: in particular, the fingers, toes, ears and face. Risk factors include smoking, peripheral vascular disease, dehydration and alcohol consumption. The tissues may become anaesthetised before freezing and, as a result, the injury often goes unrecognised at first. Frostbitten tissue is initially pale and doughy to the touch and insensitive to pain (Fig. 5.5). Once frozen, the tissue is hard.

Rewarming should not occur until it can be achieved rapidly in a water bath. Give oxygen and aspirin 300 mg as soon as possible. Frostbitten extremities should be rewarmed in warm water at 37–39°C, with antiseptic added. Adequate analgesia is necessary, as rewarming is very painful. Vasodilators such as pentoxifylline (a phosphodiesterase inhibitor) have been shown to improve tissue survival. Once it has thawed, the injured part must not be re-exposed to the cold, and should be dressed and rested. Whilst wound débridement may be necessary, amputations should be delayed for 60–90 days, as good recovery may occur over an extended period.

**Non-freezing cold injury (trench or immersion foot)**

This results from prolonged exposure to cold, damp conditions. The limb (usually the foot) appears cold, ischaemic and numb, but there is no freezing of the tissue. On rewarming, the limb appears mottled and thereafter becomes hyperaemic, swollen and painful. Recovery may take many months, during which there may be chronic pain and sensitivity to cold. The pathology remains uncertain but probably involves endothelial injury. Gradual rewarming is associated with less pain than rapid rewarming. The pain and associated paraesthesia are difficult to control with conventional analgesia and may require amitriptyline (50 mg nocte), best instituted early. The patient is at risk of further damage on subsequent exposure to the cold.

**Chilblains**

Chilblains are tender, red or purplish skin lesions that occur in the cold and wet. They are often seen in horse riders, cyclists and swimmers, and are more common in women than men. They are short-lived, and although painful, not usually serious.

**Heat-related illness**

When generation of heat exceeds the body’s capacity for heat loss, core temperature rises. Non-exertional heat illness (NEHI) occurs with high environmental temperature in those with attenuated thermoregulatory control mechanisms: the elderly, the young, those with comorbidity or those taking drugs that affect thermoregulation (particularly phenothiazines, diuretics and alcohol). Exertional heat illness (EHI), on the other hand, typically develops in athletes when heat production exceeds the body’s ability to dissipate it.
Acclimatisation mechanisms to environmental heat include stimulation of the sweat mechanism with increased sweat volume, reduced sweat sodium content and secondary hyperaldosteronism to maintain body sodium balance. The risk of heat-related illness falls as acclimatisation occurs. Heat illness can be prevented to a large extent by adequate replacement of salt and water, although excessive water intake alone should be avoided because of the risk of dilutional hyponatraemia (p. 437).

A spectrum of illnesses occurs in the heat (see Fig. 5.3). The cause is usually obvious but the differential diagnosis should be considered (Box 5.10).

**Heat cramps**

These painful muscle contractions occur following vigorous exercise and profuse sweating in hot weather. There is no elevation of core temperature. The mechanism is considered to be extracellular sodium depletion as a result of persistent sweating, exacerbated by replacement of water but not salt. Symptoms usually respond rapidly to rehydration with oral rehydration salts or intravenous saline.

**Heat syncope**

This is similar to a vasovagal faint (p. 555) and is related to peripheral vasodilatation in hot weather.

**Heat exhaustion**

Heat exhaustion occurs with prolonged exertion in hot and humid weather, profuse sweating and inadequate salt and water replacement. There is an elevation in core (rectal) temperature to between 37°C and 40°C, leading to the clinical features shown in Figure 5.3. Blood analyses may show evidence of dehydration with mild elevation of the blood urea, sodium and haematocrit. Treatment involves removal of the patient from the heat, and active evaporative cooling using tepid sprays and fanning (strip–spray–fan). Fluid losses are replaced with either oral rehydration mixtures or intravenous isotonic saline. Up to 5 L positive fluid balance may be required in the first 24 hours. Untreated, heat exhaustion may progress to heat stroke.

**Heat stroke**

Heat stroke occurs when the core body temperature rises above 40°C and is a life-threatening condition. The symptoms of heat exhaustion progress to include headache, nausea and vomiting. Neurological manifestations include a coarse muscle tremor and confusion, aggression or loss of consciousness. The patient’s skin feels very hot, and sweating is often absent due to failure of thermoregulatory mechanisms. Complications include hypovolaemic shock, lactic acidosis, disseminated intravascular coagulation, rhabdomyolysis, hepatic and renal failure, and pulmonary and cerebral oedema.

The patient should be resuscitated with rapid cooling by spraying with water, fanning and ice packs in the axillae and groins. Cold crystalloid intravenous fluids are given but solutions containing potassium should be avoided. Over-aggressive fluid replacement must be avoided, as it may precipitate pulmonary oedema or further metabolic disturbance. Appropriate monitoring of fluid balance, including central venous pressure, is important. Investigations for complications include routine haematology and biochemistry, coagulation screen, hepatic transaminases (aspartate aminotransferase and alanine aminotransferase), creatine kinase and chest X-ray. Once emergency treatment is established, heat stroke patients are best managed in intensive care.

With appropriate treatment, recovery from heat stroke can be rapid (within 1–2 hours) but patients who have had core temperatures higher than 40°C should be monitored carefully for later onset of rhabdomyolysis, renal damage and other complications before discharge from hospital. Clear advice to avoid heat and heavy exercise during recovery is important.

**High altitude**

The physiological effects of high altitude are significant. On Everest, the barometric pressure of the atmosphere falls from sea level by approximately 50% at base camp (5400 m) and approximately 70% at the summit (8848 m). The proportions of oxygen, nitrogen and carbon dioxide in air do not change with the fall in pressure but their partial pressure falls in proportion to barometric pressure (Fig. 5.6). Oxygen tension within the pulmonary alveoli is further reduced at altitude because the partial pressure of water vapour is related to body temperature and not barometric pressure, and so is proportionately greater at altitude, accounting for only 6% of barometric pressure at sea level, but 19% at 8848 m.
Physiological effects of high altitude

Reduction in oxygen tension results in a fall in arterial oxygen saturation (see Fig. 5.6). This varies widely between individuals, depending on the shape of the sigmoid oxygen-haemoglobin dissociation curve (see Fig. 8.3, p. 183) and the ventilatory response. Acclimatisation to hypoxia at high altitude involves a shift in this dissociation curve (dependent on 2,3-diphosphoglycerate (DPG)), erythropoiesis, haemocencentration, and hyperventilation resulting from hypoxic drive (which is then sustained despite hypocapnia by restoration of cerebrospinal fluid pH to normal in prolonged hypoxia). This process takes several days, so travellers need to plan accordingly.

Illnesses at high altitude

Ascent to altitudes up to 2500 m or travel in a pressurised aircraft cabin is harmless to healthy people. Above 2500 m high-altitude illnesses may occur in previously healthy people, and above 3500 m these become common. Sudden ascent to altitudes above 6000 m, as experienced by aviators, balloonists and astronauts, may result in decompression illness with the same clinical features as seen in divers (see below), or even loss of consciousness. However, most altitude illness occurs in travellers and mountaineers.

Acute mountain sickness

Acute mountain sickness (AMS) is a syndrome comprised principally of headache, together with fatigue, anorexia, nausea and vomiting, difficulty sleeping or dizziness. Ataxia and peripheral oedema may be present. Its aetiology is not fully understood but it is thought that hypoxaemia increases cerebral blood flow and hence intracranial pressure. Symptoms occur within 6–12 hours of an ascent and vary in severity from trivial to completely incapacitating. The incidence in travellers to 3000 m may be 40–50%, depending on the rate of ascent.

Treatment of mild cases consists of rest and simple analgesia; symptoms usually resolve after 1–3 days at a stable altitude, but may recur with further ascent. Occasionally there is progression to cerebral oedema. Persistent symptoms indicate the need to descend but may respond to acetazolamide, a carbonic anhydrase inhibitor that induces a metabolic acidosis and stimulates ventilation; acetazolamide may also be used as prophylaxis if a rapid ascent is planned.

High-altitude cerebral oedema

The cardinal symptoms of high-altitude cerebral oedema (HACE) are ataxia and altered consciousness. This is rare, life-threatening and usually preceded by AMS. In addition to features of AMS, the patient suffers confusion, disorientation, visual disturbance, lethargy and ultimately loss of consciousness. Papilloedema and retinal haemorrhages are common and focal neurological signs may be found.

Treatment is directed at improving oxygenation. Descent is essential and dexamethasone (8 mg immediately and 4 mg 4 times daily) should be given. If descent is impossible, oxygen therapy in a portable pressurised bag may be helpful.

High-altitude pulmonary oedema

High-altitude pulmonary oedema (HAPE) is a life-threatening condition that usually occurs in the first 4 days after ascent above 2500 m. Unlike HACE, HAPE may occur de novo without the preceding signs of AMS. Presentation is with symptoms of dry cough, exertional dyspnoea and extreme fatigue. Later, the cough becomes wet and sputum may be blood-stained. Tachycardia and tachypnoea occur at rest and crepitations may often be heard in both lung fields. There may be profound hypoxaemia, pulmonary hypertension and radiological evidence of diffuse alveolar oedema. It is not known whether the alveolar oedema is a result of mechanical stress on the pulmonary capillaries associated with the high pulmonary arterial pressure, or an effect of hypoxia on capillary permeability. Reduced arterial oxygen saturation is not diagnostic but is a marker for disease progression.

Treatment is directed at reversal of hypoxia with immediate descent and oxygen administration. Nifedipine (20 mg 4 times daily) should be given to reduce pulmonary arterial pressure, and oxygen therapy in a portable pressurised bag should be used if descent is delayed.

Chronic mountain sickness (Monge’s disease)

This occurs in over 30% of trekkers at 5000 m. The haemorrhages are usually asymptomatic and resolve spontaneously. Visual defects can occur with haemorrhage involving the macula, but there is no specific treatment.

Venous thrombosis

This has been reported at altitudes over 6000 m. Risk factors include dehydration, inactivity and the cold. The use of the oral contraceptive pill at high altitude should be considered carefully, as this is an additional risk factor.

Refractory cough

A cough at high altitude is common and usually benign. It may be due to breathing dry, cold air and increased mouth breathing, with consequent dry oral mucosa. This may be indistinguishable from the early signs of HAPE.

Air travel

Commercial aircraft usually cruise at 10000–12000 m, with the cabin pressurised to an equivalent of around 2400 m. At this altitude, the partial pressure of oxygen is 16 kPa (120 mmHg), leading to a PaO₂ in healthy people of 7.0–8.5 kPa (53–64 mmHg). Oxygen saturation is also reduced, but to a lesser degree (see Fig. 5.6). Although well tolerated by healthy people, this degree of hypoxia may be dangerous in patients with respiratory disease.
Advice for patients with respiratory disease

The British Thoracic Society has published guidance on the management of patients with respiratory disease who want to fly. Specialist pre-flight assessment is advised for all patients who have hypoxemia (oxygen saturation < 95%) at sea level, and includes spirometry and a hypoxic challenge test with 15% oxygen (performed in hospital). Air travel may have to be avoided or undertaken only with inspired oxygen therapy during the flight. Asthmatic patients should be advised to carry their inhalers in their hand baggage. Following pneumothorax, flying should be avoided while air remains in the pleural cavity, but can be considered after proven resolution or definitive (surgical) treatment.

Advice for other patients

Other circumstances in which patients are more susceptible to hypoxia require individual assessment. These include cardiac dysrhythmia, sickle-cell disease and ischaemic heart disease. Most airlines decline to carry pregnant women after the 36th week of gestation. In complicated pregnancies it may be advisable to avoid air travel at an earlier stage. Patients who have had recent abdominal surgery, including laparoscopy, should avoid flying until all intraperitoneal gas is reabsorbed. Divers should not fly for 24 hours after a dive requiring decompression stops.

Ear and sinus pain due to changes in gas volume are common but usually mild, although patients with chronic sinusitis and otitis media may need specialist assessment. A healthy mobile tympanic membrane visualised during a Valsalva manœuvre usually suggests a patent Eustachian tube.

On long-haul flights, patients with diabetes mellitus may need to adjust their insulin or oral hypoglycaemic dosing according to the timing of in-flight and subsequent meals (p. 825). Advice is available from Diabetes UK and other websites. Patients should be able to provide documentary evidence of the need to carry needles and insulin.

Deep venous thrombosis

Air travellers have an increased risk of venous thrombosis (p. 1008), due to a combination of factors, including loss of venous emptying because of prolonged immobilisation (lack of muscular activity) and reduced barometric pressure on the tissues, together with haemodilution as a result of oedema and perhaps a degree of hypoxia-induced diuresis.

Venous thrombosis can probably be prevented by avoiding dehydration and excess alcohol, and exercising muscles during the flight. Without a clear cost-benefit analysis, prophylaxis with aspirin or heparin cannot be recommended routinely, but may be considered in high-risk cases.

Under water

Drowning and near-drowning

Drowning is defined as death due to asphyxiation following immersion in a fluid, whilst near-drowning is defined as survival for longer than 24 hours after suffocation by immersion. Drowning remains a common cause of accidental death throughout the world and is particularly common in young children (Box 5.11). In about 10% of cases, no water enters the lungs and death follows intense laryngospasm (‘dry’ drowning). Prolonged immersion in cold water, with or without water inhalation, results in a rapid fall in core body temperature and hypothermia (p. 104).

Following inhalation of water, there is a rapid onset of ventilation-perfusion imbalance with hypoxaemia, and the development of diffuse pulmonary oedema. Fresh water is hypotonic and, although rapidly absorbed across alveolar membranes, impairs surfactant function, which leads to alveolar collapse and right-to-left shunting of unoxygenated blood. Absorption of large amounts of hypotonic fluid can result in haemolysis. Salt water is hypertonic and inhalation provokes alveolar oedema, but the overall clinical effect is similar to that of fresh-water drowning.

Clinical features

Those rescued alive (near-drowning) are often unconscious and not breathing. Hypoxaemia and metabolic acidosis are inevitable features. Acute lung injury usually resolves rapidly over 48–72 hours, unless infection occurs (Fig. 5.7). Complications include dehydrogenation, hypotension, haemoptysis, rhabdomyolysis, renal failure and cardiac dysrhythmias. A small number of patients, mainly the more severely ill, progress to develop the acute respiratory distress syndrome (ARDS; p. 192).

Survival is possible after immersion for up to 30 minutes in very cold water, as the rapid development of hypothermia after immersion may be protective, particularly in children. Long-term outcome depends on the severity of the cerebral hypoxic injury and is predicted by the duration of immersion, delay in resuscitation, intensity of acidosis and the presence of cardiac arrest.

Management

Initial management requires cardiopulmonary resuscitation with administration of oxygen and maintenance of the circulation (p. 558). It is important to clear the airway of foreign bodies and protect the cervical spine. Continuous positive airways pressure (CPAP; p. 193) should be considered for spontaneously breathing patients with oxygen saturations below 94%. Observation is required for a minimum of 24 hours. Prophylactic antibiotics are only required if exposure was to obviously contaminated water.
Diving-related illness

The underwater environment is extremely hostile. Other than drowning, most diving illness is related to changes in barometric pressure and its effect on gas behaviour.

Ambient pressure underwater increases by 101 kPa (1 atmosphere) for every 10 metres of seawater (msw) depth. As divers descend, the partial pressures of the gases they are breathing increase (Box 5.12), and the blood and tissue concentrations of dissolved gases rise accordingly. Nitrogen is a weak anaesthetic agent, and if the inspiratory pressure of nitrogen is allowed to increase above ~320 kPa (i.e. a depth of approximately 30 msw), it produces 'narcosis', resulting in impairment of cognitive function and manual dexterity, not unlike alcohol intoxication. For this reason, compressed air can only be used for shallow diving. Oxygen is also toxic at inspired pressures above approximately 40 kPa (inducing apprehension, muscle twitching, euphoria, sweating, tinnitus, nausea and vertigo), so 100% oxygen cannot be used as an alternative. For dives deeper than approximately 30 msw, mixtures of oxygen with nitrogen and/or helium are used.

Whilst drowning remains the most common diving-related cause of death, another important group of disorders usually present once the diver returns to the surface: decompression illness (DCI) and barotrauma.

Clinical features

Decompression illness

This includes decompression sickness (DCS) and arterial gas embolism (AGE). Whilst the vast majority of symptoms of decompression illness present within 6 hours of a dive, they can also be provoked by flying and thus patients may present to medical services at sites far removed from the dive.

Exposure of individuals to increased partial pressures of nitrogen results in additional nitrogen being dissolved in body tissues; the amount dissolved depends on the depth/pressure and on the duration of the dive. On ascent, the tissues become supersaturated with nitrogen, and this places the diver at risk of producing a critical quantity of gas (bubbles) in tissues if the ascent is too fast. The gas so formed may cause symptoms locally, by bubbles passing through the pulmonary vascular bed (Box 5.13) or by embolisation elsewhere. Arterial embolisation may occur if the gas load in the venous system exceeds the lungs’ abilities to excrete nitrogen, or when bubbles pass through a patent foramen ovale (present asymptomatically in 25–30% of adults; p. 528).

Although DCS and AGE can be indistinguishable, their early treatment is the same.

Barotrauma

During the ascent phase of a dive, the gas in the diver’s lungs expands due to the decreasing pressure. The diver must therefore ascend slowly and breathe regularly; if ascent is rapid or the diver holds his/her breath, the expanding gas may cause lung rupture (pulmonary barotrauma). This can result in pneumomediastinum, pneumothorax or AGE due to gas passing directly into the pulmonary venous system. Other air-filled body cavities may be subject to barotrauma, including the ear and sinuses.
Management

The patient is nursed horizontally, and airway, breathing and circulation are assessed. Treatment includes the following:

- **High-flow oxygen** is given by a tight-fitting mask using a rebreathing bag. This assists in the washout of excess inert gas (nitrogen) and may reduce the extent of local tissue hypoxia resulting from focal embolic injury.
- **Fluid replacement** (oral or intravenous) corrects the intravascular fluid loss from endothelial bubble injury and the dehydration associated with immersion. Maintenance of an adequate peripheral circulation is important for the excretion of excess dissolved gas.
- **Recompression** is the definitive therapy. Transfer to a recompression chamber facility may be by surface or air, provided that the altitude remains low (< 300 m) and the patient continues to breathe 100% oxygen. Recompression reduces the volume of gas within tissues and puts nitrogen back into solution.

The majority of patients make a complete recovery with treatment, although a small but significant proportion are left with neurological disability.

Nutrients in the diet can be classified into ‘macronutrients’, which are eaten in relatively large amounts to provide fuel for energy, and ‘micronutrients’ (e.g. vitamins and minerals), which do not contribute to energy balance but are required in small amounts because they are not synthesised in the body.

Energy balance

The laws of thermodynamics dictate that energy balance is achieved when energy intake = energy expenditure (Fig. 5.9).

Energy expenditure has several components. The basal metabolic rate (BMR) describes the obligatory energy expenditure required to maintain metabolic functions in tissues and hence sustain life. It is most closely predicted by fat-free mass (i.e. total body mass minus fat mass), which is lower in females and older people (Fig. 5.9B). Extra metabolic energy is consumed during growth, pregnancy and lactation, and when febrile. Metabolic energy is also required for thermal regulation, and expenditure is higher in cold or hot environments. The energy required for digestion of food (diet-induced thermogenesis (DIT); Fig. 5.9D) accounts for approximately 10% of total energy expenditure, with protein requiring more energy than other macronutrients. Another component of energy expenditure is governed by the level of muscular activity, which can vary considerably with occupation and lifestyle (Fig. 5.9C). Physical activity levels are usually defined as multiples of BMR.

Energy intake is determined by the ‘macronutrient’ content of food. Carbohydrates, fat, protein and alcohol provide fuel for oxidation in the mitochondria to generate energy (as adenosine triphosphate (ATP); p. 45). The energy provided by each of these elements differs:

- carbohydrates (16 kJ/g)
- fat (37 kJ/g)
- protein (17 kJ/g)
- alcohol (29 kJ/g).

Regulation of energy balance

Energy intake and expenditure are highly regulated (Fig. 5.10). A link with reproductive function ensures that pregnancy is most likely to occur during times of nutritional plenty when both mother and baby have a better chance of survival. Improved nutrition is thought to be the reason for the increasingly early onset of puberty in many societies. At the other extreme, anorexia nervosa and excessive exercise can lead to amenorrhoea (p. 255).

Regulation of energy balance is coordinated in the hypothalamus, which receives afferent signals that indicate nutritional status in the short term (e.g. the stomach...
hormone ghrelin, which falls immediately after eating and rises gradually thereafter, to suppress satiety and signal that it is time for the next meal) and the long term (e.g. the adipose hormone leptin, which increases with growing fat mass and may also link fat mass to reproductive function). The hypothalamus responds with changes in many local neurotransmitters that alter activity in a number of pathways that influence energy balance (see Fig. 5.10), including hormones acting on the pituitary gland (see Fig. 20.2, p. 737), and neural control circuits that connect with the cerebral cortex and autonomic nervous system.

**Responses to under- and over-nutrition**

These complex regulatory pathways allow adaptation to variations in nutrition. In response to starvation, reproductive function is suppressed, BMR is reduced, and there are profound psychological effects, including energy conservation through lethargy. These adjustments can ‘defend’ body weight within certain limits.
However, in the low-insulin state of starvation (see Fig. 21.2, p. 801), fuels are liberated from stores initially in glycogen (in liver and muscle), then in triglyceride (lipolysis in adipose tissue, with excess free fatty acid supply to the liver leading to ketosis) and finally in protein (proteolysis in muscle).

In response to over-nutrition, BMR is increased, and extra energy is consumed in the work of carrying increased fat stores, so that body weight is again ‘defended’ within certain limits. In the high-insulin state of over-nutrition, excess energy is invested in fatty acids and stored as triglycerides; these are deposited principally in adipose tissue but they may also accumulate in the liver (non-alcoholic fatty liver disease; p. 959) and skeletal muscle. In the absence of hypothalamic function (e.g. in those with craniopharyngioma; see Fig. 20.30, p. 794) or in rare patients with mutations in relevant genes (e.g. in leptin or melanocortin-4 receptors), loss of response to satiety signals, together with loss of adaptive changes in energy expenditure, result in relentless weight gain.

**Macronutrients**  
*(energy-yielding nutrients)*

**Carbohydrates**

Types of carbohydrate and their dietary sources are listed in Box 5.14. The ‘available’ carbohydrates (starches and sugars) are broken down to monosaccharides before absorption from the gut (p. 842), and supply over half the energy in a normal, well-balanced diet (see Fig. 5.9A). No individual carbohydrate is an essential nutrient, as carbohydrates can be synthesised de novo from glycerol or protein. However, if the available carbohydrate intake is less than 100 g per day, increased lipolysis leads to ketosis (see Fig. 21.5, p. 804).

Dietary guidelines do not restrict the intake of intrinsic sugars in fruit and vegetables or the sugars in milk. However, intake of non-milk extrinsic sugars (sucrose, maltose, fructose), which increase the risk of dental caries and diabetes mellitus, should be limited. Individuals who do not produce lactase (‘lactose-intolerant’) are advised to avoid or limit dairy products and foods with added lactose. Starches in cereal foods, root foods and legumes provide the largest proportion of energy in most diets around the world. All starches are polymers of glucose, linked by the same 1–4 glycosidic linkages. However, some starches are digested promptly by salivary and then pancreatic amylase, producing rapid delivery of glucose to the blood. Other starches are digested more slowly, either because they are protected in the structure of the food, because of their crystal structure, or because the molecule is unbranched (amylose). These differences are the basis for the ‘glycaemic index’ of foods. This is the area under the curve of the rise in blood glucose concentration in the 2 hours following ingestion of 50 g carbohydrate, expressed as a percentage of the response to 50 g anhydrous glucose. There is emerging evidence linking high glycaemic index foods with obesity and type 2 diabetes (p. 806).
5.11 The principal polyunsaturated fatty acid (PUFA) in circulation. Fatty acid structures are shown in Figure 5.11, allowing access of complex molecules into the absorbed in chylomicrons (pp. 450 and 841; see Fig. 22.5,ous cause of obesity (see Fig. 5.9A). Free fatty acids are concentration.

They reduce bile salt absorption and hence plasma cholesterol reduce gastric emptying, contribute to satiety, and are important in the upper gastrointestinal tract, where indigestible polysaccharides like pectin and guar gum reduce the risk of cancer of the colon. Other viscous, constipation, appear to prevent diverticulosis and may pass unchanged into the large intestine. Dietary fibre can be broken down by the resident bacteria in the colon to produce short-chain fatty acids. This is essential fuel for the enterocytes and contributes to bowel health. The extent of flatus formed is dependent on the food source.

Some types of NSP, notably the hemicellulose of wheat, increase the water-holding capacity of colonic contents and the bulk of faeces. They relieve simple constipation, appear to prevent diverticulosis and may reduce the risk of cancer of the colon. Other viscous, indigestible polysaccharides like pectin and guar gum are important in the upper gastrointestinal tract, where they slow gastric emptying, contribute to satiety, and reduce bile salt absorption and hence plasma cholesterol concentration.

**Fats**

Fat has the highest energy density of the macronutrients (37 kJ/g) and excessive consumption may be insidious cause of obesity (see Fig. 5.9A). Free fatty acids are absorbed in chylomicrons (pp. 450 and 841; see Fig. 22.5, p. 842), allowing access of complex molecules into the circulation. Fatty acid structures are shown in Figure 5.11. The principal polyunsaturated fatty acid (PUFA) in plant seed oils is linoleic acid (18:2 \( \omega 6 \)). This and alphalinolenic acid (18:3 \( \omega 3 \)) are the ‘essential’ fatty acids, which humans cannot synthesise de novo. They undergo further desaturation and elongation, to produce, for example, \( \gamma \)-linolenic acid (18:3 \( \omega 6 \)) and arachidonic acid (20:4 \( \omega 6 \)). These are precursors of prostaglandins and eicosanoids, and form part of the structure of lipid membranes in all cells. Fish oils are rich in \( \omega 3 \) PUFA (e.g. eicosapentaenoic (20:5 \( \omega 3 \)) and docosahexaenoic (22:6 \( \omega 3 \)), which promote the anti-inflammatory cascade of prostaglandin production and occur in the lipids of the human brain and retina. They inhibit thrombosis by competitively antagonising thromboxane \( A_2 \) formation. Substituting saturated fat (i.e. from animal sources: butter, ghee or lard) with PUFA in the diet can lower the concentration of circulating low-density lipoprotein (LDL) cholesterol and may help prevent coronary heart disease. High intakes of \( \text{trans} \) fatty acids (TFA) (isomers of the natural \( \text{cis} \) fatty acids) reflect the use of oils that have been partially hydrogenated in the food industry. It is recommended that TFAs are limited to <2% of the dietary fat intake, as they are associated with cardiovascular disease. Changes in industrial practice in the UK and US have meant that TFA intake is now below 1%, with the residual amounts coming from milk as a result of ruminant digestion.

Cholesterol is also absorbed directly from food in chylomicrons and is an important substrate for steroid and sterol synthesis, but not an important source of energy.

<table>
<thead>
<tr>
<th>5.14 Dietary carbohydrates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class</strong></td>
</tr>
<tr>
<td>Free sugars</td>
</tr>
<tr>
<td>Short-chain carbohydrates</td>
</tr>
<tr>
<td>Starch polysaccharides</td>
</tr>
<tr>
<td>Non-starch polysaccharides (NSP, dietary fibre)</td>
</tr>
<tr>
<td>Sugar alcohols</td>
</tr>
</tbody>
</table>

Sugar alcohols (e.g. sorbitol) that are used as replacement sweeteners can cause diarrhoea if eaten in large amounts.

**Dietary fibre**

Dietary fibre is plant food that is not digested by human enzymes in the gastrointestinal tract. Most dietary fibre is known as the ‘non-starch polysaccharides’ (NSP) (see Box 5.14). A small percentage of ‘resistant’ dietary starch may also pass unchanged into the large intestine. Dietary fibre can be broken down by the resident bacteria in the colon to produce short-chain fatty acids. This is essential fuel for the enterocytes and contributes to bowel health. The extent of flatus formed is dependent on the food source.

Some types of NSP, notably the hemicellulose of wheat, increase the water-holding capacity of colonic contents and the bulk of faeces. They relieve simple constipation, appear to prevent diverticulosis and may reduce the risk of cancer of the colon. Other viscous, indigestible polysaccharides like pectin and guar gum are important in the upper gastrointestinal tract, where they slow gastric emptying, contribute to satiety, and reduce bile salt absorption and hence plasma cholesterol concentration.

**Fats**

Fat has the highest energy density of the macronutrients (37 kJ/g) and excessive consumption may be insidious cause of obesity (see Fig. 5.9A). Free fatty acids are absorbed in chylomicrons (pp. 450 and 841; see Fig. 22.5, p. 842), allowing access of complex molecules into the circulation. Fatty acid structures are shown in Figure 5.11. The principal polyunsaturated fatty acid (PUFA) in plant seed oils is linoleic acid (18:2 \( \omega 6 \)). This and alphalinolenic acid (18:3 \( \omega 3 \)) are the ‘essential’ fatty acids, which humans cannot synthesise de novo. They undergo further desaturation and elongation, to produce, for example, \( \gamma \)-linolenic acid (18:3 \( \omega 6 \)) and arachidonic acid (20:4 \( \omega 6 \)). These are precursors of prostaglandins and eicosanoids, and form part of the structure of lipid membranes in all cells. Fish oils are rich in \( \omega 3 \) PUFA (e.g. eicosapentaenoic (20:5 \( \omega 3 \)) and docosahexaenoic (22:6 \( \omega 3 \)), which promote the anti-inflammatory cascade of prostaglandin production and occur in the lipids of the human brain and retina. They inhibit thrombosis by competitively antagonising thromboxane \( A_2 \) formation. Substituting saturated fat (i.e. from animal sources: butter, ghee or lard) with PUFA in the diet can lower the concentration of circulating low-density lipoprotein (LDL) cholesterol and may help prevent coronary heart disease. High intakes of \( \text{trans} \) fatty acids (TFA) (isomers of the natural \( \text{cis} \) fatty acids) reflect the use of oils that have been partially hydrogenated in the food industry. It is recommended that TFAs are limited to <2% of the dietary fat intake, as they are associated with cardiovascular disease. Changes in industrial practice in the UK and US have meant that TFA intake is now below 1%, with the residual amounts coming from milk as a result of ruminant digestion.

Cholesterol is also absorbed directly from food in chylomicrons and is an important substrate for steroid and sterol synthesis, but not an important source of energy.
Proteins

Proteins are made up of some 20 different amino acids, of which nine are ‘essential’ (Box 5.15), i.e. they cannot be synthesised in humans but are required for synthesis of important proteins. Another group of five amino acids are termed ‘conditionally essential’, meaning that they can be synthesised from other amino acids, provided there is an adequate dietary supply. The remaining amino acids can be synthesised in the body by transamination, provided there is a sufficient supply of amino groups.

The nutritive or ‘biological’ value of different proteins depends on the relative proportions of essential amino acids they contain. Proteins of animal origin, particularly from eggs, milk and meat, are generally of higher biological value than proteins of vegetable origin, which are low in one or more of the essential amino acids. However, when two different vegetable proteins are eaten together (e.g. a cereal and a legume), their amino acid contents are complementary and produce an adequate mix, an important principle in vegan diets.

Dietary recommendations for macronutrients

Recommendations for energy intake (Box 5.16) and proportions of macronutrients (Box 5.17) have been calculated to provide a balance of essential nutrients and minimise the risks of excessive refined sugar (dental caries, high glycaemic index/diabetes mellitus), saturated fat or trans fat (obesity, coronary heart disease). Recommended dietary fibre intake is based on avoiding risks of colonic disease. The usual recommended protein intake for a healthy man doing light work is 65–100 g/day. The minimum requirement is around 40 g of protein with a high proportion of essential amino acids or a high biological value.

Clinical assessment and investigation of nutritional status

The diverse manifestations of inadequate nutrition dictate that its clinical assessment and investigation involve many systems. Energy balance is reflected in body composition, which is most readily assessed by clinical anthropometric measurements. It can also be tested non-invasively by the measurement of body fat by bio-impedance or dual energy X-ray absorptiometry (DEXA) scanning. Abnormal micronutrient status is commonly manifest in clinical signs in the skin and mucous membranes, or in other systems.

A dietary history provides useful information, especially when obtained by a dietitian. A weighed food diary is considered to be the gold standard dietary assessment but is rarely conducted in clinical practice.

Anthropometric measurements

Body mass index (BMI) is useful for categorising under-and over-nutrition. It is the weight in kilograms divided by the height in metres, squared. For example, an adult weighing 70 kg with a height of 1.75 m has a BMI of 70/1.75² = 22.9 kg/m². If height cannot be determined (e.g. in older people with kyphosis or in those who cannot stand), a surrogate measure is:

- the demispan: measured from the sternal notch to the middle finger; height = 0.73 × (2 × demispan) + 0.43
- knee height:

females (60–80 years): height (cm) = (knee height (cm) × 1.91) – (age (years) × 0.17) + 75.00

- These are based on a healthy target body mass index (BMI) of 22.5 kg/m². For a female, height is 162 cm and weight 59.0 kg; for a male, height is 175 cm and weight 68.8 kg. Previous average recommendations of 8.1 MJ (1950 kcal, usually rounded up to 2000 kcal) for females and 10.7 MJ (2500 kcal) for males should continue to be used, as these fall within experimental error.
males (60–80 years): height (cm) = (knee height (cm) × 2.05) + 59.01.

BMI does not discriminate between fat mass and lean body mass and can be increased by muscle mass (e.g. in athletes). Moreover, there are ethnic differences in body fat content; at the same BMI, Asians have more body fat than Europeans. For optimal health, the BMI should be 18.5–24.9 kg/m².

An indication of the degree of abdominal obesity is the waist circumference, measured at the level of the umbilicus. Hip circumference can be measured at the level of the greater trochanters; waist:hip ratios show whether the distribution of fat is android or gynoid (see below). Skinfold measurements can be used to calculate body fat content, whereas relative loss of muscle and subcutaneous fat can be estimated by measuring mid-arm circumference (at the middle of the humerus) and skinfold thickness over the triceps (using special callipers); muscle mass is estimated by subtracting triceps skinfold thickness from mid-arm circumference.

**DISORDERS OF ALTERED ENERGY BALANCE**

**Obesity**

Obesity is widely regarded as a pandemic, with potentially disastrous consequences for human health. Over one-quarter of adults in the UK were obese (i.e. BMI ≥ 30 kg/m²) in 2010, compared with 7% prevalence in 1980 and 16% in 1995. Moreover, almost two-thirds of the UK adult population are overweight (BMI ≥ 25 kg/m²), although there is considerable regional and age group variation. In developing countries, average national rates of obesity are low, but these figures may disguise high rates of obesity in urban communities; for example, nearly one-quarter of women in urban India are overweight.

There is increasing public awareness of the health implications of obesity. Many patients will seek medical help for their obesity, others will present with one of the complications of obesity, and increasing numbers are being identified during health screening examinations.

**Complications of obesity**

Obesity has adverse effects on both mortality and morbidity (Box 5.18). Changes in mortality are difficult to analyse due to the confounding effects of lower body weight in cigarette smokers and those with other illnesses (such as cancer). However, it is clear that the lowest mortality rates are seen in Europeans in the BMI range 18.5–24 kg/m² (and at lower BMI in Asians). It is suggested that obesity at age 40 years can reduce life expectancy by up to 7 years for non-smokers and by 13 years for smokers. Coronary heart disease (Fig. 5.12) is the major cause of death but cancer rates are also increased in the overweight, especially colorectal cancer in males and cancer of the gallbladder, biliary tract, breast, endometrium and cervix in females. Obesity has little effect on life expectancy above 70 years of age, but the obese do spend a greater proportion of their active life disabled. Epidemic obesity has been accompanied by an epidemic of type 2 diabetes (p. 806) and osteoarthritis, particularly of the knee. Although an increased body size results in greater bone density through increased mechanical stress, it is not certain whether this translates to a lower incidence of osteoporotic fractures (p. 1120). Obesity may have profound psychological consequences, compounded by stigmatisation of the obese in many societies.

**Body fat distribution**

For some complications of obesity, the distribution rather than the absolute amount of excess adipose tissue appears to be important. Increased intra-abdominal fat causes ‘central’ (‘abdominal’, ‘visceral’, ‘android’ or ‘apple-shaped’) obesity, which contrasts with subcutaneous fat accumulation causing ‘generalised’ (‘gynoid’ or ‘pear-shaped’) obesity; the former is more common in men and is more closely associated with type 2 diabetes, the metabolic syndrome and cardiovascular disease (see Box 5.18). The key difference between these depots of fat may lie in their vascular anatomy, with intra-abdominal fat draining into the portal vein and thence directly to the liver. Thus many factors that are released from adipose tissue (including free fatty acids; ‘adipokines’, such as tumour necrosis factor-α, adiponectin and resistin; and steroid hormones) may be at higher concentration in the liver and hence induce insulin resistance and promote type 2 diabetes (p. 805). Recent research has also highlighted the importance of fat deposition within specific organs, especially the

---

**5.18 Complications of obesity**

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Metabolic syndrome’</td>
<td>Coronary heart disease</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Stroke</td>
</tr>
<tr>
<td>Hypertension</td>
<td>Diabetes complications</td>
</tr>
<tr>
<td>Hyperlipidaemia</td>
<td></td>
</tr>
<tr>
<td>Liver fat accumulation</td>
<td>Non-alcoholic steatohepatitis</td>
</tr>
<tr>
<td>Restricted ventilation</td>
<td>Cirrhosis</td>
</tr>
<tr>
<td>Mechanical effects of weight</td>
<td>Exertional dyspnoea</td>
</tr>
<tr>
<td></td>
<td>Obstructive sleep apnoea</td>
</tr>
<tr>
<td></td>
<td>Obesity hypoventilation syndrome (Pickwickian syndrome)</td>
</tr>
<tr>
<td>Increased peripheral steroid interconversion in adipose tissue</td>
<td>Hormone-dependent cancers (breast, uterus)</td>
</tr>
<tr>
<td></td>
<td>Polycystic ovarian syndrome (infertility, hirsutism; p. 764)</td>
</tr>
<tr>
<td>Others</td>
<td>Psychological morbidity (low self-esteem, depression)</td>
</tr>
<tr>
<td></td>
<td>Socioeconomic disadvantage (lower income, less likely to be promoted)</td>
</tr>
<tr>
<td></td>
<td>Gallstones</td>
</tr>
<tr>
<td></td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td></td>
<td>Skin infections (groin and submammary candidiasis; hidradenitis)</td>
</tr>
</tbody>
</table>
overweight and obese women. The obese liver, as an important determinant of metabolic risk in the obese.

Aetiology

Accumulation of fat results from a discrepancy between energy consumption and energy expenditure that is too large to be defended by the hypothalamic regulation of BMR. A continuous small daily positive energy balance (of only 0.2–0.8 MJ (50–200 kcal; < 10% of intake) would lead to weight gain of 2–20 kg over a period of 4–10 years. Given the cumulative effects of subtle energy excess, body fat content shows ‘tracking’ with age such that obese children usually become obese adults. Weight tends to increase throughout adult life, as BMR and physical activity decrease (see Fig. 5.9).

The pandemic of obesity reflects changes in both energy intake and energy expenditure (Box 5.19), although both are difficult to measure reliably. The estimated average global daily supply of food energy per person increased from approximately 9.8 MJ (2350 kcal) in the 1960s to approximately 11.7 MJ (2800 kcal) in the 1990s, but its delivery is unequal. For example, in India it is estimated that 5% of the population receives 40% of the available food energy, leading to obesity in the urban population in parallel with persisting under-nutrition in some rural communities. In affluent societies, a significant proportion of this food supply is discarded. In the US, the average daily energy intake of men reportedly rose from 10.2 MJ (2450 kcal) in 1971 to 11.0 MJ (2618 kcal) in 2000. Portion sizes, particularly of energy-dense foods such as drinks with highly refined sugar content and salty snacks, have increased. However, data in the UK suggest that energy intakes have declined (which may in part be due to deliberate restriction or ‘dieting’), but this is apparently insufficient to compensate for the decrease in physical activity levels in recent years. Obesity is correlated positively with the number of hours spent watching television, and inversely with levels of physical activity (e.g. stair climbing). It is suggested that minor activities such as fidgeting and chewing gum may contribute to energy expenditure and protect against obesity.

Susceptibility to obesity

Susceptibility to obesity and its adverse consequences undoubtedly varies between individuals. It is not true that obese subjects have a ‘slow metabolism’, since their BMR is higher than that of lean subjects. Twin and adoption studies confirm a genetic influence on obesity. The pattern of inheritance suggests a polygenic disorder, with small contributions from a number of different genes, together accounting for 25–70% of variation in weight. Recent results from ‘genome-wide’ association studies of polymorphisms in large numbers of people (p. 53) have identified a handful of genes that influence obesity, some of which encode proteins known to be involved in the control of appetite or metabolism and some of which have unknown function. However, these genes account for less than 5% of the variation in body weight.

A few rare single-gene disorders have been identified that lead to severe childhood obesity. These include mutations of the melanocortin-4 receptor (MC4R), which account for approximately 5% of severe early-onset obesity; defects in the enzymes processing propiomelanocortin (POMC, the precursor for adrenocortico-trophic hormone (ACTH)) in the hypothalamus; and mutations in the leptin gene (see Fig. 5.9). The latter can be treated by leptin injections. Additional genetic conditions in which obesity is a feature include the Prader–Willi (see Box 3.3, p. 54) and Lawrence–Moon–Biedl syndromes.
Reversible causes of obesity and weight gain

In a small minority of patients presenting with obesity, specific causal factors can be identified and treated (Box 5.20). These patients are distinguished from those with idiopathic obesity by their short history, with a recent marked change in the trajectory of their adult weight gain.

Clinical assessment and investigations

In assessing an individual presenting with obesity, the aims are to:
- quantify the problem
- exclude an underlying cause
- identify complications
- reach a management plan.

Severity of obesity can be quantified using the BMI (Box 5.21). A waist circumference of > 102 cm in men or > 88 cm in women indicates that the risk of metabolic and cardiovascular complications of obesity is high.

A dietary history may be helpful in guiding dietary advice, but is notoriously susceptible to under-reporting of food consumption. It is important to consider ‘pathological’ eating behaviour (such as binge eating, nocturnal eating or bulimia; p. 255), which may be the most important issue to address in some patients. Alcohol is an important source of energy intake and should be considered in detail.

The history of weight gain may help diagnose underlying causes. A patient who has recently gained substantial weight or has gained weight at a faster rate than previously, and is not taking relevant drugs (see Box 5.20), is more likely to have an underlying disorder such as hypothyroidism (p. 743) or Cushing’s syndrome (p. 773). All obese patients should have thyroid function tests performed on one occasion, and an overnight dexamethasone suppression test or 24-hour urine free cortisol if Cushing’s syndrome is suspected. Mono- genic and ‘syndromic’ causes of obesity are usually only relevant in children presenting with severe obesity.

Assessment of the diverse complications of obesity (see Box 5.18) requires a thorough history, examination and screening investigations. The impact of obesity on the patient’s life and work is a major consideration. Assessment of other cardiovascular risk factors is important. Blood pressure should be measured with a large cuff, if required (p. 608). Associated type 2 diabetes and dyslipidaemia are detected by measuring blood glucose or HbA1c and a serum lipid profile, ideally in a fasting morning sample. Elevated serum transaminases occur in patients with non-alcoholic fatty liver disease (p. 959).

Management

The health risks of obesity are largely reversible. Interventions proven to reduce weight in obese patients also ameliorate cardiovascular risk factors. Lifestyle advice that lowers body weight and increases physical exercise reduces the incidence of type 2 diabetes (p. 820). Given the high prevalence of obesity and the large magnitude of its risks, population strategies to prevent and reverse obesity are high on the public health priority list for many countries. Initiatives include promoting healthy eating in schools, enhancing walking and cycling options for commuters, and liaising with the food industry to reduce energy and fat content and to label foods appropriately. Unfortunately, ‘low-fat’ foods are often still energy-dense, and current lifestyles with labour-saving devices, sedentary work and passive leisure activities have much lower energy requirements than the manual labour and household duties of previous generations.

Most patients seeking assistance with obesity are motivated to lose weight but have attempted to do so previously without long-term success. Often weight will have oscillated between periods of successful weight loss and then regain of weight (‘recidivism’). These patients may hold misconceptions that they have an underlying disease, inaccurate perceptions of their energy intake and expenditure, and an unrealistic view of the target weight that they would regard as a ‘success’. An empathetic explanation of energy balance, which recognises that some individuals are more susceptible to obesity than others and may find it more difficult to lose and sustain body weight loss, is important. Exclusion of underlying ‘hormone imbalance’ with simple tests is reassuring and shifts the focus on to consideration of energy balance. Appropriate goals for weight loss should be agreed, recognising that the slope of the relationship between obesity and many of its complications becomes steeper with increasing BMI, so that a given amount of weight loss achieves greater risk reduction at higher levels of BMI. A reasonable goal for most patients is to lose 5–10% of body weight.

<table>
<thead>
<tr>
<th>Classification*</th>
<th>Risk of obesity comorbidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.5–24.9</td>
<td>Reference range</td>
</tr>
<tr>
<td>25.0–29.9</td>
<td>Overweight</td>
</tr>
<tr>
<td>&gt; 30.0</td>
<td>Obese</td>
</tr>
<tr>
<td>30.0–34.9</td>
<td>Class I</td>
</tr>
<tr>
<td>35.0–39.9</td>
<td>Class II</td>
</tr>
<tr>
<td>&gt; 40.0</td>
<td>Class III</td>
</tr>
</tbody>
</table>

*Classification of the WHO and International Obesity Task Force. The Western Pacific Region Office of WHO recommends that, amongst Asians, BMI > 23.0 is overweight and > 25.0 is obese.
ENVIRONMENTAL AND NUTRITIONAL FACTORS IN DISEASE

The management plan will vary according to the severity of the obesity (see Box 5.21) and the associated risk factors and complications. It will also be influenced by availability of resources; health-care providers and regulators have generally been careful not to recommend expensive interventions (especially long-term drug therapy and surgery) for everyone who is overweight. Instead, most guidelines focus resources on short-term interventions in those who have high health risks and comorbidities associated with their obesity, and who have demonstrated their capacity to alter their lifestyle to achieve weight loss (Fig. 5.13).

Lifestyle advice

Behavioural modification to avoid some of the effects of the ‘obesogenic’ environment (see Box 5.19) is the cornerstone of long-term control of weight. Regular eating patterns and maximising physical activity are advised, with reference to the modest extra activity required to increase physical activity level (PAL) ratios (see Fig. 5.9C, p. 111). Where possible, this should be incorporated in the daily routine (e.g. walking rather than driving to work), since this is more likely to be sustained. Alternative exercise (e.g. swimming) may be considered if musculoskeletal complications prevent walking. Changes in eating behaviour (including food selection, portion size control, avoidance of snacking, regular meals to encourage satiety, and substitution of sugar with artificial sweeteners) should be discussed. Regular support from a dietitian or attendance at a weight loss group may be helpful.

Weight loss diets

In overweight people, adherence to the lifestyle advice given above may gradually induce weight loss. In obese patients, more active intervention is usually required to lose weight before conversion to the ‘weight maintenance’ advice given above. A significant industry has developed in marketing diets for weight loss. These vary substantially in their balance of macronutrients (Box 5.22), but there is little evidence that they vary in their medium-term (1-year) efficacy. Most involve recommending a reduction of daily total energy intake of \(-2.5\, \text{MJ} (600\, \text{kcal})\) from the patient’s normal consumption. Modelling data that take into account the reduced energy expenditure as weight is lost suggest that a reduction of energy intake of 100 KJ per day will lead to an eventual bodyweight change of about 1 kg, with half of the weight change being achieved in about 1 year and 95% of the weight change in about 3 years. Weight loss is highly variable, with patient compliance being the major determinant of success. There is some evidence that weight loss diets are most effective in their early weeks, and that compliance is improved by novelty of the diet; this provides some justification for switching to a different dietary regimen when weight loss slows on the first diet. Vitamin supplementation is wise in those diets in which macronutrient balance is markedly disturbed.

In some patients, more rapid weight loss is required, e.g. in preparation for surgery. There is no role for starvation diets, which risk profound loss of muscle mass and the development of arrhythmias (and even sudden death) secondary to elevated free fatty acids, ketosis and deranged electrolytes. Very-low-calorie diets (VLCDs) are recommended for short-term rapid weight loss, producing losses of 1.5–2.5 kg/week, compared to 0.5 kg/week on conventional regimens, but require the supervision of an experienced physician and nutritionist. The composition of the diet should ensure a minimum of 50 g of protein each day for men and 40 g for women to minimise muscle degradation. Energy content should be a minimum of 1.65 MJ (400 kcal) for women of height < 1.73 m, and 2.1 MJ (500 kcal) for all men and for women taller than 1.73 m. Side-effects are a problem in

---

### 5.22 Low-calorie diet therapy for obesity

<table>
<thead>
<tr>
<th>Diet</th>
<th>% carbohydrate</th>
<th>% fat</th>
<th>% protein</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (typical developed country)</td>
<td>50</td>
<td>30</td>
<td>15</td>
<td>Maintains balance in macronutrients and micronutrients while reducing energy-dense fats</td>
</tr>
<tr>
<td>Moderate fat (e.g. Weight Watchers)</td>
<td>60</td>
<td>25</td>
<td>15</td>
<td>Induction of ketosis may suppress hunger</td>
</tr>
<tr>
<td>Low carbohydrate (e.g. Atkins)</td>
<td>10</td>
<td>60</td>
<td>30</td>
<td>Protein has greater satiety effect than other macronutrients</td>
</tr>
<tr>
<td>High protein (e.g. Zone)</td>
<td>43</td>
<td>30</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Low fat (e.g. Ornish)</td>
<td>70</td>
<td>13</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

---

Fig. 5.13 Therapeutic options for obesity. Relevant comorbidities include type 2 diabetes, hypertension, cardiovascular disease, sleep apnoea, and waist circumference > 102 cm in men or 88 cm in women. This is an approximate consensus of the numerous national guidelines, which vary slightly in their recommendations and are revised every few years.
the early stages and include orthostatic hypotension, headache, diarrhoea and nausea.

**Drugs**

A huge investment has been made by the pharmaceutical industry in finding drugs for obesity. The side-effect profile has limited the use of many agents, with notable withdrawals from clinical use of sibutramine (increased cardiovascular events) and rimonabant (psychiatric side-effects) in recent years; only one drug, orlistat, is currently licensed for long-term use. A number of other agents are in development, so the situation could change rapidly over the next few years. There is no role for diuretics, or for thyroxine therapy without biochemical evidence of hypothyroidism.

Orlistat inhibits pancreatic and gastric lipases and thereby decreases the hydrolysis of ingested triglycerides, reducing dietary fat absorption by approximately 30%. The drug is not absorbed and adverse side-effects relate to the effect of the resultant fat malabsorption on the gut: namely, loose stools, oily spotting, faecal urgency, flatus and the potential for malabsorption of fat-soluble vitamins. Orlistat is taken with each of the three main meals of the day and the dose can be adjusted (60–120 mg) to minimise side-effects. Its efficacy is shown in Figure 5.14; these effects may be explained because patients taking orlistat adhere better to low-fat diets in order to avoid unpleasant gastrointestinal side-effects.

Drug therapy is usually reserved for patients with high risk of complications from obesity (see Fig. 5.13), and its optimum timing and duration are controversial. Although life-long therapy is advocated for many drugs that reduce risk on the basis of relatively short-term research trials (e.g. drugs for hypertension and osteoporosis), some patients who continue to take anti-obesity drugs tend to regain weight with time; this may partly reflect age-related weight gain, but significant weight gain should prompt reinforcement of lifestyle advice and, if this is unsuccessful, drug therapy should be discontinued (see Fig. 5.14).

**Surgery**

‘Bariatric’ surgery is by far the most effective long-term treatment for obesity (see Fig. 5.14 and Box 5.23) and is described in this section.
the only anti-obesity intervention that has been associated with reduced mortality. Bariatric surgery should be contemplated in motivated patients who have very high risks of complications of obesity (see Fig. 5.13), in whom extensive dietary and drug therapy has been insufficiently effective. It is usually reserved for those with severe obesity (BMI > 40 kg/m²), or those with a BMI > 35 kg/m² and significant complications, such as type 2 diabetes or obstructive sleep apnoea. Only experienced specialist surgeons should undertake these procedures, in collaboration with a multidisciplinary team. Several approaches are used (Fig. 5.15) and all can be performed laparoscopically. The mechanism of weight loss may not simply relate to limiting the stomach or absorptive capacity, but rather in disrupting the release of ghrelin from the stomach or promoting the release of other peptides from the small bowel, thereby enhancing satiety signalling in the hypothalamus. Diabetes may improve rapidly after surgery, particularly after gastric bypass, and although this may be attributed to severe energy restriction in the perioperative period, it is possible that increased release of incretin hormones such as glucagon-like peptide (GLP)–1 may contribute to the improvement in glucose control. Complications depend upon the approach. Mortality is low in experienced centres, but post-operative respiratory problems, wound infection and dehiscence, staple leaks, stomal stenosis, marginal ulcers and venous thrombosis may occur. Additional problems may arise at a later stage, such as pouch and distal oesophageal dilatation, persistent vomiting, ‘dumping’ (p. 875) and micronutrient deficiencies, particularly of folate, vitamin B₁₂ and iron, which are of concern especially to women contemplating pregnancy.

Cosmetic surgical procedures may be considered in obese patients after successful weight loss. Apronec-tomy is usually advocated to remove an overhang of abdominal skin, especially if infected or ulcerated. This operation is of no value for long-term weight reduction if food intake remains unrestricted.

Treatment of additional risk factors

Obesity must not be treated in isolation and other risk factors must be addressed, including smoking, excess alcohol consumption, diabetes mellitus, hyperlipidaemia, hypertension and obstructive sleep apnoea. Treatment of these is discussed in the relevant chapters.

Under-nutrition

Starvation and famine

There remain regions of the world, particularly rural Africa, where under-nutrition due to famine is endemic, the prevalence of BMI < 18.5 kg/m² (Box 5.24) in adults is as high as 20%, and growth retardation due to under-nutrition affects 50% of children.

WHO reports that chronic under-nutrition is responsible for more than half of all childhood deaths worldwide. Starvation is manifest as marasmus (malnutrition with marked muscle-wasting), or, when additional complicating mechanisms, such as oxidative stress, come into play, malnourished children can develop kwashiorkor (malnutrition with oedema). Growth retardation is due to deficiency of key nutrients, e.g. protein, zinc, potassium, phosphorus and sulphur. Treatment of these

<table>
<thead>
<tr>
<th>BMI (kg/m²)</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 20</td>
<td>Adequate nutrition</td>
</tr>
<tr>
<td>18.5–20</td>
<td>Marginal</td>
</tr>
<tr>
<td>&lt; 18.5</td>
<td>Under-nutrition</td>
</tr>
<tr>
<td>17–18.4</td>
<td>Mild</td>
</tr>
<tr>
<td>16–17</td>
<td>Moderate</td>
</tr>
<tr>
<td>&lt; 16</td>
<td>Severe</td>
</tr>
</tbody>
</table>

5.24 Classification of under-nutrition in adults by body mass index (weight/height²)
5.25 Causes of under-nutrition and weight loss in adults

**Decreased energy intake**
- Famine
- Persistent regurgitation or vomiting
- Anorexia, including depression and anorexia nervosa
- Malabsorption (e.g. small intestinal disease)
- Maldigestion (e.g. pancreatic exocrine insufficiency)

**Increased energy expenditure**
- Increased BMR (thyrotoxicosis, trauma, fever, cancer, cachexia)
- Excessive physical activity (e.g. marathon runners)
- Energy loss (e.g. glycosuria in diabetes)
- Impaired energy storage (e.g. Addison’s disease, phaeochromocytoma)

Childhood conditions is not discussed in this adult medicine textbook.

In adults, starvation is the result of chronic under-nutrition, i.e. sustained negative energy (calorie) balance. Causes are shown in Box 5.25. Causes of weight loss are considered further on page 859.

**Clinical assessment**

In starvation, the severity of malnutrition can be assessed by anthropometric measurements, such as BMI (see Box 5.24). Demispan and mid-arm circumference measurements (p. 114) are most useful in monitoring progress during treatment. The clinical features of severe under-nutrition in adults include:
- weight loss
- thirst, craving for food, weakness and feeling cold
- nocturia, amenorrhoea or impotence
- lax, pale, dry skin with loss of turgor and, occasionally, pigmented patches
- cold and cyanosed extremities, pressure sores
- hair thinning or loss (except in adolescents)
- muscle-wasting, best demonstrated by the loss of the temporalis and periscapular muscles and reduced mid-arm circumference
- loss of subcutaneous fat, reflected in reduced skinfold thickness and mid-arm circumference
- hypothermia, bradycardia, hypotension and small heart
- oedema, which may be present without hypoalbuminaemia (‘famine oedema’)
- distended abdomen with diarrhoea
- diminished tendon jerks
- apathy, loss of initiative, depression, introversion, aggression if food is nearby
- susceptibility to infections (Box 5.26).

Under-nutrition often leads to vitamin deficiencies, especially of thiamin, folate and vitamin C (see below).

Diarrhoea can lead to depletion of sodium, potassium and magnesium. The high mortality rate in famine situations is often due to outbreaks of infection, e.g. typhus or cholera, but the usual signs of infection may not be apparent. In advanced starvation, patients become completely inactive and may assume a flexed, fetal position. In the last stage of starvation, death comes quietly and often quite suddenly. The very old are most vulnerable. All organs are atrophied at necropsy, except the brain, which tends to maintain its weight.

**Investigations**

In a famine, laboratory investigations may be impractical, but will show that plasma free fatty acids are increased and there is ketosis and a mild metabolic acidosis. Plasma glucose is low but albumin concentration is often maintained because the liver still functions normally. Insulin secretion is diminished, glucagon and cortisol tend to increase, and reverse T3 replaces normal triiodothyronine (p. 738). The resting metabolic rate falls, partly because of reduced lean body mass and partly because of hypothalamic compensation (see Fig. 5.9, p. 111). The urine has a fixed specific gravity and creatinine excretion becomes low. There may be mild anaemia, leucopenia and thrombocytopenia. The erythrocyte sedimentation rate is normal unless there is infection. Tests of delayed skin hypersensitivity, e.g. to tuberculin, are falsely negative. The electrocardiogram shows sinus bradycardia and low voltage.

**Management**

Whether in a famine or in wasting secondary to disease, the severity of under-nutrition is graded according to BMI (see Box 5.24). People with mild starvation are in no danger; those with moderate starvation need extra feeding; those who are severely underweight need hospital care.

In severe starvation, there is atrophy of the intestinal epithelium and of the exocrine pancreas, and the bile is dilute. It is critical that the condition is managed by experts. When food becomes available, it should be given by mouth in small, frequent amounts at first, using a suitable formula preparation (Box 5.27). Individual energy requirements can vary by 30%. During rehabilitation, more concentrated formula can be given with additional food that is palatable and similar to the usual staple meal. Salt should be restricted and micronutrient supplements may be essential (e.g. potassium, magnesium, zinc and multivitamins). Between 6.3 and 8.4 MJ/day (1500–2000 kcal/day) will arrest progressive under-nutrition, but additional energy may be required for regain of weight. During refeeding, a weight gain of 5% body weight per month indicates satisfactory progress. Other care is supportive, and includes attention to the skin, adequate hydration, treatment of infections, and careful monitoring of body temperature since thermoregulation may be impaired.

Circumstances and resources are different in every famine, but many problems are non-medical and concern organisation, infrastructure, liaison, politics, procurement, security and ensuring that food is distributed on the basis of need. Lastly, plans must be made for the future for prevention and/or earlier intervention if similar circumstances prevail.
Under-nutrition in hospital

Under-nutrition is a common problem in the hospital setting. In the UK, approximately one-third of patients are affected by moderate or severe under-nutrition on admission. The elderly are particularly at risk (Box 5.28). Once in hospital, many patients lose weight due to factors such as poor appetite, poor dental health, concurrent illness and even being kept ‘nil by mouth’ for investigations. Under-nutrition is poorly recognised in hospital and has serious consequences. Physical effects include impaired immunity and muscle weakness, which in turn affect cardiac and respiratory function, and delayed wound healing after surgery with increased risks of post-operative infection. The undernourished patient is often apathetic and withdrawn, which may be mistaken for a depressive illness and can affect cooperation with treatment and rehabilitation.

This can be averted with proper monitoring and involvement of an appropriate multidisciplinary team. As a minimum standard, all patients should be weighed on admission to hospital and at least weekly until discharge. A scoring system for identifying patients at nutritional risk is shown in Figure 5.16.

Nutritional support of the hospital patient

Normal diet

As a first step, patients should be encouraged to eat a normal and adequate diet. This is often neglected and there is evidence of substantial wastage in hospital food. In patients at risk of under-nutrition (see Fig. 5.16), quantities eaten should be recorded on a food chart. Hospital staff must identify and overcome barriers to adequate food intake, such as unpalatability of food, cultural and religious factors influencing acceptability of

---

**Nutritional Risk Scoring System**

<table>
<thead>
<tr>
<th>BMI score</th>
<th>Weight loss score</th>
<th>Acute disease score</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 18.5</td>
<td>Unplanned loss in 6 months</td>
<td>Acute illness with no nutritional intake for 5 days</td>
</tr>
<tr>
<td>&lt; 18.5</td>
<td>&lt; 5% = 0</td>
<td>= 2</td>
</tr>
<tr>
<td></td>
<td>5 – 10% = 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt; 10% = 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total = 0</th>
<th>Low risk</th>
<th>Routine clinical care</th>
<th>Repeat screen weekly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total = 1</td>
<td>Medium risk</td>
<td>Document dietary intake for 3 days</td>
<td>Repeat screen weekly</td>
</tr>
<tr>
<td>Total ≥ 2</td>
<td>High risk</td>
<td>Refer to dietitian/nutrition support team</td>
<td>Review plan weekly</td>
</tr>
</tbody>
</table>

**Fig. 5.16** Screening hospitalised patients for risk of malnutrition. Acute illnesses include decompensated liver disease, cancer cachexia or being kept ‘nil by mouth’. Adapted from the British Association of Parenteral and Enteral Nutrition Malnutrition Universal Screening Tool (www.bapen.org.uk).
food, difficulty with hand dexterity (arthritis, stroke), immobility in bed, or poor oral health. Hospital catering departments have an important role in providing acceptable and adequate meals.

**Dietary supplements**

If sufficient nutritional intake cannot be achieved from normal diet alone, then dietary supplements should be used. These are drinks with high energy and protein content, and are available in cartons as manufactured, flavoured products or are made in the hospital kitchen from milk products and egg. They should be prescribed, and administered by nursing staff, to ensure that they are taken regularly. Dietary supplements do not significantly affect the patient’s consumption of normal food.

**Enteral tube feeding**

Patients who are unable to swallow may require artificial nutritional support: for example, after acute stroke or throat surgery, or when there are long-term neurological problems such as motor neuron disease and multiple sclerosis. The enteral route should always be used if possible, since feeding via the gastrointestinal tract preserves the integrity of the mucosal barrier. This prevents bacteraemia and, in intensive care patients, reduces the risk of multi-organ failure (p. 198).

If the need for artificial nutritional support is thought to be short-term, then feeding is instituted using a fine-bore nasogastric tube. The position of the tube in the stomach must be confirmed before any fluid is administered, as severe respiratory complications can occur if fluid is inadvertently infused into a bronchus (Box 22.48, p. 879). Thereafter, specially prepared liquid feeds are administered either by continuous infusion or using a bolus technique. If the patient fails to absorb the administered feed or vomits it, this may indicate gastric outlet obstruction or gastric stasis, which can be overcome by placing a nasojejunal tube.

If long-term artificial enteral feeding is needed, a percutaneous endoscopic gastrostomy (PEG) should be sited (Fig. 5.17). A PEG tube is more comfortable for the patient, since there is no irritation to the nasal mucosa. The tube is less likely to become displaced or to be pulled out, so the feed can be given more reliably. However, inserting a gastrostomy is an invasive procedure, especially in frail patients with significant comorbidities. It may be complicated by local infection (30%) and inadvertent puncture of other intra-abdominal organs, causing peritonitis and bleeding, so the indication for placement must be carefully considered. It takes approximately 10 days for a fibrous tract to form around the PEG tube. If the PEG is displaced or removed during that time, there is a high risk of peritonitis. If a problem occurs with food absorption, a jejunal extension can be placed through the PEG tube and liquid feed administered directly into the small bowel.

**Parenteral nutrition**

Intravenous feeding should only be used when enteral feeding is impossible. Parenteral feeding is expensive and carries higher risks of complications. There is little benefit if parenteral feeding is required for less than 1 week.

There are a number of possible routes for parenteral nutrition:

- **Peripheral venous cannula.** This can only be used for low-osmolality solutions due to the development of thrombophlebitis, and is unsuitable for patients with high nutritional requirements.
- **Peripherally inserted cannula (PIC).** A 20 cm cannula is placed in a mid-arm vein. Once again, hyperosmolar solutions cannot be used.

![Fig. 5.17 Percutaneous endoscopic gastrostomy (PEG) placement. A] Finger pressure on the anterior abdominal wall is noted by the endoscopist. B] Following insertion of a cannula through the anterior abdominal wall into the stomach, a guidewire is threaded through the cannula and grasped by the endoscopic forceps or snare. C] The endoscope is withdrawn with the guidewire. The gastrostomy tube is then attached to the guidewire. D] The guidewire and tube are pulled back through the mouth, oesophagus and stomach to exit on the anterior abdominal wall, and the endoscope is repassed to confirm the site of placement of the retention device. The latter closely abuts the gastric mucosa; its position is maintained by an external fixation device (see inset). It is also possible to place PEG tubes using fluoroscopic guidance in patients in whom endoscopy is difficult (radiologically inserted gastrostomy (RIG)).
• **Peripheral inserted central catheter (PICC).** A 60 cm cannula is inserted into a vein in the antecubital fossa. The distal end lies in a central vein, allowing hyperosmololar solutions to be used.
• **Central line.** The subclavian route is preferred to the internal jugular vein, due to lower infection rates. Hyperosmololar solutions can be used without difficulty. Lines need to be handled with strict aseptic technique, and a single-lumen tube is preferred, to prevent infection.

If access has been gained to a central vein, nutritional support is usually given as an ‘all-in-one’ mixture. The main energy source is provided by carbohydrate, usually as glucose. The solution also contains amino acids, lipid emulsion, electrolytes, trace elements and vitamins. These are mixed as a large bag in a sterile environment, with the constituents adjusted according to the results of regular blood monitoring. Relevant tests include:

- **daily:** urea and electrolytes, glucose
- **twice weekly:** liver function tests, calcium, phosphate, magnesium
- **weekly:** full blood count, zinc, triglycerides
- **monthly:** copper, selenium, manganese

If the patient develops fever or other features of septicaemia, it should be assumed to be due to a line infection (p. 200). Blood cultures should be taken, the existing line removed, the tip sent for bacteriological analysis, and a new line inserted.

**Refeeding syndrome**

When nutritional support is given to an under-nourished patient, there is a rapid conversion from a catabolic to an anabolic state. Administration of carbohydrates stimulates release of insulin, leading to cellular uptake of phosphate, potassium and magnesium, which may provoke significant falls in serum levels. The resulting electrolyte imbalance can have serious consequences, such as cardiac arrhythmias, so careful monitoring is essential. In patients who are thiamin-deficient, Wernicke’s encephalopathy can be precipitated by refeeding with carbohydrates (p. 253); this is prevented by administering thiamin before starting nutritional support.

**Legal and ethical aspects of artificial nutritional support**

The ability to intervene with artificial nutritional support raises many legal and ethical dilemmas (pp. 9 and 291). Starvation will inevitably lead to death, but inability to eat may be part of the terminal stages of a disease process. Difficult decisions are raised by situations such as strokes, which affect swallowing. The instigation of feeding may speed recovery and lead to better functional outcome; on the other hand, feeding might prolong the process of dying in severe stroke. There will be different approaches to these decisions, depending on the local availability of resources as well as legal, cultural and religious influences. Some guidelines are given in Box 5.29.

**Cachexia**

Cachexia is the weight loss and muscle-wasting associated with chronic illness, which is characteristic of chronic infections such as HIV/AIDS, end-stage organ failure and certain cancers (especially of the lung and upper gastrointestinal tract). Although there is decreased energy intake with loss of appetite, the main cause is thought to be increased metabolic rate through the production of key cytokines and other proteolytic factors.

---

**5.29 Ethical and legal considerations in the management of artificial nutritional support**

- **Care of the sick involves the duty of providing adequate fluid and nutrients**
- **Food and fluid should not be withheld from a patient who expresses a desire to eat and drink, unless there is a medical contraindication (e.g. risk of aspiration)**
- **A treatment plan should include consideration of nutritional issues and should be agreed by all members of the health-care team**
- **In the situation of palliative care, tube feeding should only be instituted if it is needed to relieve symptoms**
- **Tube feeding is usually regarded in law as a medical treatment. Like other treatments, the need for such support should be reviewed on a regular basis and changes made in the light of clinical circumstances**
- **A competent adult patient must give consent for any invasive procedures, including the passage of a nasogastric tube or the insertion of a central venous cannula**
- **If a patient is unable to give consent, the health-care team should act in that person’s best interests, taking into account any wishes previously expressed by the patient and the views of family**
- **Under certain specified circumstances (e.g. anorexia nervosa), it will be appropriate to provide artificial nutritional support to the unwilling patient**

*Based on British Association for Parenteral and Enteral Nutrition guidelines (www.bapen.org.uk).

**MICRONUTRIENTS, MINERALS AND THEIR DISEASES**

**Vitamins**

Vitamins are organic substances with key roles in certain metabolic pathways, and are categorised into those that are fat-soluble (vitamins A, D, E and K) and those that are water-soluble (vitamins of the B complex group and vitamin C).

Recommended daily intakes of micronutrients (Box 5.30) vary between countries and the nomenclature has become potentially confusing. In the UK, the ‘reference nutrient intake’ (RNI) has been calculated as the mean plus two standard deviations (SD) of daily intake in the population, which therefore describes normal intake for 97.5% of the population. The lower reference nutrient intake (LRNI) is the mean minus 2 SD, below which would be considered deficient in most of the population. These dietary reference values (DRV) have superseded the terms RDI (recommended daily intakes) and RDA (recommended daily amounts). Other countries use different terminology. Additional amounts of some micronutrients may be required in pregnancy and lactation (Box 5.31).

Vitamin deficiency diseases are most prevalent in developing countries but still occur in developed
### 5.30 Summary of clinically important vitamins

<table>
<thead>
<tr>
<th>Vitamin</th>
<th>Rich</th>
<th>Important</th>
<th>Reference nutrient intake (RNI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fat-soluble</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (retinol)</td>
<td>Liver</td>
<td>Milk and milk products, eggs, fish oils</td>
<td>700 µg men</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>600 µg women</td>
</tr>
<tr>
<td>D (cholecalciferol)</td>
<td>Fish oils</td>
<td>UV exposure to skin</td>
<td>10 µg if &gt; 65 yrs or no sunlight exposure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Egg yolks, margarine, fortified cereals</td>
<td></td>
</tr>
<tr>
<td>E (tocopherol)</td>
<td>Sunflower oil</td>
<td>Vegetables, nuts, seed oils</td>
<td>No RNI. Safe intake:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 mg men</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 mg women</td>
</tr>
<tr>
<td>K (phylloquinone, menaquinone)</td>
<td>Green vegetables</td>
<td>Soya oil, menaquinones produced by intestinal bacteria</td>
<td>No RNI. Safe intake: 1 µg/kg</td>
</tr>
<tr>
<td><strong>Water-soluble</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1 (thiamin)</td>
<td>Pork</td>
<td>Cereals, grains, beans</td>
<td>0.8 mg per 9.68 MJ (2000 kcal) energy intake</td>
</tr>
<tr>
<td>B2 (riboflavin)</td>
<td>Milk</td>
<td>Milk and milk products, breakfast cereals, bread</td>
<td>1.3 mg men</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.1 mg women</td>
</tr>
<tr>
<td>B3 (niacin, nicotinic acid, nicotinamide)</td>
<td>Meat, cereals</td>
<td></td>
<td>17 mg men</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13 mg women</td>
</tr>
<tr>
<td>B6 (pyridoxine)</td>
<td>Meat, fish, potatoes, bananas</td>
<td>Vegetables, intestinal microflora synthesis</td>
<td>1.4 mg men</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.2 mg women</td>
</tr>
<tr>
<td>Folate</td>
<td>Liver</td>
<td>Green leafy vegetables, fortified breakfast cereals</td>
<td>200 µg</td>
</tr>
<tr>
<td>B12 (cobalamin)</td>
<td>Animal products</td>
<td>Bacterial colonisation</td>
<td>1.5 µg</td>
</tr>
<tr>
<td>Biotin</td>
<td>Egg yolk</td>
<td>Intestinal flora</td>
<td>No RNI. Safe intake: 10–200 µg</td>
</tr>
<tr>
<td>C (ascorbic acid)</td>
<td>Citrus fruit</td>
<td>Fresh fruit, fresh and frozen vegetables</td>
<td>40 mg</td>
</tr>
</tbody>
</table>

*Rich sources contain the nutrient in high concentration but are not generally eaten in large amounts; important sources contain less but contribute most because larger amounts are eaten.

### 5.31 Nutrition in pregnancy and lactation

- **Energy requirements**: increased in both the mother and fetus, but can be met through reduced maternal energy expenditure.
- **Micronutrient requirements**: adaptive mechanisms ensure increased uptake of minerals in pregnancy, but extra increments of some are required during lactation (see Box 5.33). Additional increments of some vitamins are recommended during pregnancy and lactation:
  - Vitamin A: for growth and maintenance of the fetus, and to provide some reserve (important in some countries to prevent blindness associated with vitamin A deficiency). Teratogenic in excessive amounts.
  - Vitamin D: to ensure bone and dental development in the infant. Higher incidences of hypocalcaemia, hypoparathyroidism and defective dental enamel have been seen in infants of women not taking vitamin D supplements at > 50° latitude.
  - Folate: to avoid neural tube defects (see Box 5.32).
  - Vitamin B12: in lactation only.
  - Thiamin: to meet increased fetal energy demands.
  - Riboflavin: to meet extra demands.
  - Niacin: in lactation only.
  - Vitamin C: for the last trimester to maintain maternal stores as fetal demands increase.
  - Iodine: in countries with high consumption of staple foods (e.g. brassicas, maize, bamboo shoots) that contain goitrogens (thiocyanates or perchlorates) that interfere with iodine uptake, supplements prevent infants being born with cretinism.

### EBM 5.32 Periconceptual folate supplementation and neural tube defects

‘Folate supplementation in advance of conception and during the first trimester reduces the incidence of neural tube defects by ~70%.


For further information: [www.cochrane.org/cochrane-reviews](http://www.cochrane.org/cochrane-reviews)

countries. Older people (and alcoholics) are at risk of deficiencies in B vitamins and in vitamins D and C. Nutritional deficiencies in pregnancy can affect either the mother or the developing fetus, and extra increments of vitamins are recommended in the UK (see Boxes 5.31 and 5.32). Darker-skinned individuals living at higher latitude, and those who cover up or do not go outside are at increased risk of vitamin D deficiency due to inadequate sunlight exposure. Dietary supplements are recommended for these ‘at-risk’ groups. Some nutrient deficiencies are induced by diseases or drugs.
Deficiencies of fat-soluble vitamins are seen in conditions of fat malabsorption (e.g. biliary obstruction).

Some vitamins also have pharmacological actions when given at supraphysiological doses, e.g. the use of vitamin A (p. 1282) for acne. Taking vitamin supplements is fashionable in many countries, although there is no evidence of benefit. Toxic effects are most serious with high dosages of vitamins A, B₆ and D.

Investigation of suspected vitamin deficiency or excess may involve biochemical assessment of body stores (Box 5.33). However, measurements in blood should be interpreted carefully in conjunction with the clinical presentation.

### Fat-soluble vitamins

#### Vitamin A (retinol)

Pre-formed retinol is found only in foods of animal origin. Vitamin A can also be derived from carotenes, which are present in green and coloured vegetables and some fruits. Carotenes provide most of the total vitamin A in the UK, and constitute the only supply in vegans. Retinol is converted to several other important molecules:

- **11-cis retinaldehyde** is part of the photoreceptor complex in rods of the retina.
- **Retinoic acid** induces differentiation of epithelial cells by binding to specific nuclear receptors, which induce responsive genes. In vitamin A deficiency, mucus-secreting cells are replaced by keratin-producing cells.
- **Retinoids** are necessary for normal growth, fetal development, fertility, haematopoiesis and immune function.

Globally, the most important consequence of vitamin A deficiency is irreversible blindness in young children. Asia is most notably affected and the problem is being addressed through widespread vitamin A supplementation.

---

**Fig. 5.18** Eye signs of vitamin A deficiency. **A** Bitot’s spots in xerophthalmia, showing the white triangular plaques (arrows). **B** Keratomalacia in a 14-month-old child. There is liquefactive necrosis affecting the greater part of the cornea, with typical sparing of the superior aspect. (B) From WHO 1976 – see p. 132.
supplementation programmes. Adults are not usually at risk because liver stores can supply vitamin A when foods containing vitamin A are unavailable.

Early deficiency causes impaired adaptation to the dark (night blindness). Keratinisation of the cornea (xerophthalmia) gives rise to characteristic Bitot’s spots, and progresses to keratomalacia, with corneal ulceration, scarring and irreversible blindness (Fig. 5.18). In countries where vitamin A deficiency is endemic, pregnant women should be advised to eat dark-green, leafy vegetables and yellow fruits (to build up stores of retinol in the fetal liver), and infants should be fed the same. WHO is according high priority to prevention in communities where xerophthalmia occurs, giving single prophylactic oral doses of 60 mg retinyl palmitate (providing 200 000 U retinol) to pre-school children. This also reduces mortality from gastroenteritis and respiratory infections.

Repeated moderate or high doses of retinol can cause liver damage, hyperostrosis and teratogenicity. Women in countries where deficiency is not endemic are therefore advised not to take vitamin A supplements in pregnancy. Retinol intake may also be restricted in those at risk of osteoporosis. Acute overdose leads to nausea and headache, increased intracranial pressure and skin desquamation. Excessive intake of carotene can cause pigmentation of the skin (hypercarotenosis); this gradually fades when intake is reduced.

**Vitamin D**

The natural form of vitamin D, cholecalciferol or vitamin D₃, is formed in the skin by the action of UV light on 7-dehydrocholesterol, a metabolite of cholesterol. Few foods contain vitamin D naturally and skin exposure to sunlight is the main source. Moving away from the equator, the intensity of UV light decreases, so that at a latitude above 50° (including northern Europe), vitamin D is not synthesised in winter, and even above 30° there is seasonal variation. The body store accumulated during the summer is consumed during the winter. Vitamin D is converted in the liver to 25(OH)D, which is further hydroxylated in the kidneys to 1,25-dihydroxy-vitamin D (1,25(OH)₂D), the active form of the vitamin (see Fig. 25.55, p. 1127). 1,25(OH)₂D activates specific intracellular receptors which influence calcium metabolism, bone mineralisation and tissue differentiation. The synthetic form, ergocalciferol, or vitamin D₂, is considered to be less potent than the endogenous D₃.

Recommended dietary intakes aim to prevent rickets and osteomalacia. There is increasing evidence that vitamin D is important for immune and muscle function, and osteomalacia. There is increasing evidence that isolated metabolites cause hypercalcaemia (p. 767).

**Vitamin E**

There are eight related fat-soluble substances with vitamin E activity. The most important dietary form is α-tocopherol. Vitamin E has many direct metabolic actions:

- It prevents oxidation of polyunsaturated fatty acids in cell membranes by free radicals.
- It helps maintain cell membrane structure.
- It affects DNA synthesis and cell signalling.
- It is involved in the anti-inflammatory and immune systems.

Human deficiency is rare and has only been described in premature infants and in malabsorption. It can cause a mild haemolytic anaemia, ataxia and visual scotomas. Vitamin E intakes are considered safe up to 3200 mg/day (1000-fold greater than recommended intakes). Diets rich in vitamin E are consumed in countries with lower rates of coronary heart disease. However, randomised controlled trials have not demonstrated cardioprotective effects of vitamin E or other antioxidants.

**Vitamin K**

Vitamin K is supplied in the diet mainly as vitamin K₁ (phylloquinone) in the UK, or as vitamin K₂ (menaquinone) from fermented products in parts of Asia. Vitamin K₂ is also synthesised by bacteria in the colon. Vitamin K is a co-factor for carboxylation reactions: in particular, the production of γ-carboxyglutamate (gla). Gla residues are found in four of the coagulation factor proteins (II, VII, IX and X; p. 997), conferring their capacity to bind to phospholipid surfaces in the presence of calcium. Other important gla proteins are osteocalcin and matrix gla protein, which are important in bone mineralisation.

Vitamin K deficiency leads to delayed coagulation and bleeding. In obstructive jaundice, dietary vitamin K is not absorbed and it is essential to administer the vitamin in parenteral form before surgery. Warfarin and related anticoagulants (p. 1019) act by antagonising vitamin K. Vitamin K is given routinely to newborn babies to prevent haemorrhagic disease. Symptoms of excess have been reported only in infants, with synthetic preparations linked to haemolysis and liver damage.

**Water-soluble vitamins**

**Thiamin (vitamin B₁)**

Thiamin is widely distributed in foods of both vegetable and animal origin. Thiamin pyrophosphate (TPP) is a co-factor for enzyme reactions involved in the
metabolism of macronutrients (carbohydrate, fat and alcohol), including:

- decarboxylation of pyruvate to acetyl-coenzyme A, which bridges between glycolysis and the tricarboxylic acid (Krebs) cycle
- transketolase activity in the hexose monophosphate shunt pathway
- decarboxylation of α-ketoglutarate to succinate in the Krebs cycle.

In thiamin deficiency, cells cannot metabolise glucose aerobically to generate energy as ATP. Neuronal cells are most vulnerable, since they depend almost exclusively on glucose for energy requirements. Impaired glucose oxidation also causes an accumulation of pyruvic and lactic acids, which produce vasodilatation and increased cardiac output.

**Deficiency – beri-beri**

In the developed world, thiamin deficiency is mainly encountered in chronic alcoholics. Poor diet, impaired absorption, storage and phosphorylation of thiamin in the liver, and the increased requirements for thiamin to metabolise ethanol all contribute. In the developing world, deficiency usually arises as a consequence of a diet based on polished rice. The body has very limited stores of thiamin, so deficiency is manifest after only 1 month on a thiamin-free diet. There are two forms of the disease in adults:

- **Dry** (or neurological) beri-beri manifests with chronic peripheral neuropathy and with wrist and/or foot drop, and may cause Korsakoff’s psychosis and Wernicke’s encephalopathy (p. 253).
- **Wet** (or cardiac) beri-beri causes generalised oedema due to biventricular heart failure with pulmonary congestion.

In dry beri-beri, response to thiamin administration is not uniformly good. However, multivitamin therapy seems to produce some improvement, suggesting that other vitamin deficiencies may be involved. Wernicke’s encephalopathy and wet beri-beri should be treated without delay with intravenous vitamin B and C mixture (‘Pabrinex’, p. 253). Korsakoff’s psychosis is irreversible and does not respond to thiamin treatment.

**Riboflavin (vitamin B₂)**

Riboflavin is required for the flavin co-factors involved in oxidation–reduction reactions. It is widely distributed in animal and vegetable foods. Levels are low in staple cereals but germination increases its content. It is destroyed under alkaline conditions by heat and by exposure to sunlight.

Deficiency is rare in developed countries. It mainly affects the tongue and lips and manifests as glossitis, angular stomatitis and cheilosis. The genitals may be affected, as well as the skin areas rich in sebaceous glands, causing nasolabial or facial dyssebacea. Rapid recovery usually follows administration of riboflavin 10 mg daily by mouth.

**Niacin (vitamin B₃)**

Niacin encompasses nicotinic acid and nicotinamide. Nicotinamide is an essential part of the two pyridine nucleotides, nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), which play a key role as hydrogen acceptors and donors for many enzymes. Niacin can be synthesised in the body in limited amounts from the amino acid tryptophan.

**Deficiency – pellagra**

Pellagra was formerly endemic among poor people who subsisted chiefly on maize, which contains niacin, a form of niacin that the body is unable to utilise. Pellagra can develop in only 8 weeks in individuals eating diets that are very deficient in niacin and tryptophan. It remains a problem in parts of Africa, and is occasionally seen in alcoholics and in patients with chronic small intestinal disease in developed countries. Pellagra can occur in Hartnup’s disease, a genetic disorder characterised by impaired absorption of several amino acids, including tryptophan. It is also seen occasionally in carcinoid syndrome (p. 784), when tryptophan is consumed in the excessive production of 5-hydroxytryptamine (5-HT). Pellagra has been called the disease of the three Ds:

- **Dermatitis.** Characteristically, there is erythema resembling severe sunburn, appearing symmetrically over the parts of the body exposed to sunlight, particularly the limbs and especially on the neck, but not the face (Casal’s necklace, Fig. 5.19). The skin lesions may progress to vesication, cracking, exudation and secondary infection.
- **Diarrhoea.** This is often associated with anorexia, nausea, glossitis and dysphagia, reflecting the presence of a non-infective inflammation that extends throughout the gastrointestinal tract.
- **Dementia.** In severe deficiency, delirium occurs acutely and dementia develops in chronic cases.

![Fig. 5.19 Dermatitis due to pellagra (niacin deficiency). The lesions appear on those parts of the body exposed to sunlight. The classic ‘Casal’s necklace’ can be seen around the neck and upper chest. From Karthikeyan and Thappa 2002 – see p. 132.](image_url)
Treatment is with nicotinamide, given in a dose of 100 mg 3 times daily orally or parenterally. The response is usually rapid. Within 24 hours, the erythema diminishes, the diarrhoea ceases and a striking improvement occurs in the patient’s mental state.

Toxicity

Excessive intakes of niacin may lead to reversible hepatotoxicity. Nicotinic acid is a lipid-lowering agent, but at doses above 200 mg a day gives rise to vasodilatory symptoms (‘flushing’ and/or hypotension).

Pyridoxine (vitamin B₆)

Pyridoxine, pyridoxal and pyridoxamine are different forms of vitamin B₆ that undergo phosphorylation to produce pyridoxal 5-phosphate (PLP). PLP is the co-factor for a large number of enzymes involved in the metabolism of amino acids. Vitamin B₆ is available in most foods.

Deficiency is rare, although certain drugs, such as isoniazid and penicillamine, act as chemical antagonists to pyridoxine. Pyridoxine administration is effective in isoniazid-induced peripheral neuropathy and some cases of sideroblastic anemia. Large doses of vitamin B₆ have an antiemetic effect in radiotherapy-induced nausea. Although vitamin B₆ supplements have become popular in the treatment of nausea in pregnancy, carpal tunnel syndrome and premenstrual syndrome, there is no convincing evidence of benefit. Very high doses of vitamin B₆ taken for several months can cause a sensory polyneuropathy.

Biotin

Biotin is a co-enzyme in the synthesis of fatty acids, isoleucine and valine and is also involved in gluconeogenesis. Deficiency results from consuming very large quantities of raw egg whites (> 30% energy intake) because the avidin they contain binds to and inactivates biotin in the intestine. It may also be seen after long periods of total parenteral nutrition. The clinical features of deficiency include scaly dermatitis, alopecia and paraesthesia.

Folate (folic acid)

Folates exist in many forms. The main circulating form is 5-methyltetrahydrofolate. The natural forms are prone to oxidation. Folic acid is the stable synthetic form. Folate works as a methyl donor for cellular methylation and protein synthesis. It is directly involved in DNA and RNA synthesis, and requirements increase during embryonic development.

Folate deficiency may cause three major birth defects (spina bifida, anencephaly and encephalocele) resulting from imperfect closure of the neural tube, which takes place 3–4 weeks after conception. The UK Department of Health advises that women who have experienced a pregnancy affected by a neural tube defect should take 5 mg of folic acid daily from before conception and throughout the first trimester (see Box 5.32, p. 125). All women planning a pregnancy are advised to include good sources of folate in their diet, and to take folate supplements throughout the first trimester. Liver is the richest source of folate but an alternative source (e.g. leafy vegetables) is advised in early pregnancy because of the high vitamin A content of liver (p. 126). Folate deficiency has also been associated with heart disease, dementia and cancer. There is mandatory fortification of flour with folic acid in the US and voluntary fortification of many foods across Europe. There are now concerns that this may contribute to the increased incidence of colon cancer through promotion of the growth of polyps.

Hydroxycobalamin (vitamin B₁₂)

Vitamin B₁₂ is a co-factor in folate co-enzyme recycling and nerve myelination. Vitamin B₁₂ and folate are particularly important in DNA synthesis in red blood cells (p. 1024). The haematological disorders (macrocytic or megaloblastic anaemias) due to their deficiency are discussed on pages 1024–1026. Vitamin B₁₂, but not folate, is needed for the integrity of myelin, so that vitamin B₁₂ deficiency is also associated with neurological disease (see Box 24.35, p. 1024).

Neurological consequences of vitamin B₁₂ deficiency

In older people and chronic alcoholics, vitamin B₁₂ deficiency arises from insufficient intake and/or from malabsorption. Several drugs, including neomycin, can render vitamin B₁₂ inactive. Adequate intake of folate maintains erythropoiesis and there is a concern that fortification of foods with folate may mask underlying vitamin B₁₂ deficiency. In severe deficiency there is insidious, diffuse and uneven demyelination. It may be clinically manifest as peripheral neuropathy or spinal cord degeneration affecting both posterior and lateral columns (‘subacute combined degeneration of the spinal cord’; p. 1222), or there may be cerebral manifestations (resembling dementia) or optic atrophy. Vitamin B₁₂ therapy improves symptoms in most cases.

Vitamin C (ascorbic acid)

Ascorbic acid is the most active reducing agent in the aqueous phase of living tissues and is involved in intracellular electron transfer. It takes part in the hydroxylation of proline and lysine in protocollagen to hydroxyproline and hydroxylsine in mature collagen. It is very easily destroyed by heat, increased pH and light, and is very soluble in water; hence many traditional cooking methods reduce or eliminate it. Claims that high-dose vitamin C improves immune function (including resistance to the common cold) and cholesterol turnover remain unsubstantiated.

Deficiency – scurvy

Vitamin C deficiency causes defective formation of collagen with impaired healing of wounds, capillary haemorrhage and reduced platelet adhesiveness (normal platelets are rich in ascorbate) (Fig. 5.20). Precipitants and clinical features of scurvy are shown in Box 5.35. A dose of 250 mg vitamin C 3 times daily by mouth should saturate the tissues quickly. The deficiencies of the patient’s diet also need to be corrected and other vitamin supplements given if necessary. Daily intakes of more than 1 g/day have been reported to cause diarrhoea and the formation of renal oxalate stones.

Other dietary organic compounds

There are a number of non-essential organic compounds with purported health benefits such as reducing risk of heart disease or cancer. Groups of compounds such as the flavonoids and phytoestrogens show bioactivity
through their respective antioxidant and oestrogenic or anti-oestrogenic activities. Flavonoids (of which there are a number of different classes of compound) are found in fruit and vegetables, tea and wine; phytoestrogens are found in soy products (with higher intakes in parts of Asia compared to Europe and the US) and pulses. Caffeine from tea and coffee and carbonated beverages affects the nervous system and can improve mental performance in the short term, with adverse effects seen at higher intakes. Intake of non-carbonic organic acids (which are not metabolised to carbon dioxide), e.g. oxalates, may be restricted in individuals prone to kidney stones.

**Calcium and phosphorus**

Calcium is the most abundant cation in the body and powerful homeostatic mechanisms control circulating ionised calcium levels (pp. 766 and 1126). WHO’s dietary guidelines for calcium differ between countries, with higher intakes usually recommended in places with higher fracture prevalence. Between 20 and 30% of calcium in the diet is absorbed, depending on vitamin D status and food source. Calcium requirements depend on phosphorus intakes, with an optimum molar ratio (Ca:P) of 1:1. Excessive phosphorus intakes (e.g. 1–1.5 g/day) with a Ca:P of 1:3 have been shown to cause hypocalcaemia and secondary hyperparathyroidism (p. 768).

Calcium absorption may be impaired in vitamin D deficiency (pp. 766 and 1126) and in malabsorption secondary to small intestinal disease. Calcium deficiency causes impaired bone mineralisation and can lead to osteomalacia in adults. The potential benefits of high calcium intake in osteoporosis are discussed on page 1124. Too much calcium can lead to constipation and toxicity has been observed in ‘milk-alkali syndrome’ (p. 767).

Dietary deficiency of phosphorus is rare (except in older people with limited diets) since it is present in nearly all foods and phosphates are added to a number of processed foods. Phosphate deficiency in adults occurs:

- in patients with renal tubular phosphate loss (p. 448)
- due to prolonged high dosage of aluminium hydroxide (p. 488)
- sometimes when alcoholics are fed with high-carbohydrate foods
- in patients receiving parenteral nutrition if inadequate phosphate is provided.

Deficiency causes hypophosphataemia (p. 448) and muscle weakness secondary to ATP deficiency.

**Iron**

Iron is involved in the synthesis of haemoglobin, and is required for the transport of electrons within cells and in a number of enzyme reactions. Non-haem iron in cereals and vegetables is poorly absorbed but makes the greater contribution to overall intake, compared to the well-absorbed haem iron from animal products. Fruits and vegetables containing vitamin C enhance iron absorption, while the tannins in tea reduce it. Dietary calcium reduces iron uptake from the same meal, which

---

**Inorganic micronutrients**

A number of inorganic elements are essential dietary constituents for humans (Box 5.36). Deficiency is seen when there is inadequate dietary intake of minerals or excessive loss from the body. Toxic effects have also been observed from self-medication and disordered absorption or excretion. Examples of clinical toxicity include excess of iron (haemochromatosis or haemosiderosis), fluoride (fluorosis; p. 223), copper (Wilson’s disease) and selenium (selenosis, seen in parts of China).

For most minerals, the available biochemical markers do not accurately reflect dietary intake and dietary assessment is required.

---

**Fig. 5.20 Scurvy.**

may precipitate iron deficiency in those with borderline iron stores. There is no physiological mechanism for excretion of iron, so homeostasis depends on the regulation of iron absorption (see Fig. 24.20, p. 1022). This is regulated at the level of duodenal enterocytes by hepcidin. The expression of hepcidin (a peptide secreted by hepatocytes in the duodenum) is suppressed when body iron is low, leading to enhanced efflux of iron into the circulation. The normal daily loss of iron is 1 mg, arising from desquamated surface cells and intestinal losses. A regular loss of only 2 mL of blood per day doubles the iron requirement. On average, an additional 20 mg of iron is lost during menstruation, so pre-menopausal women require about twice as much iron as men (and more if menstrual losses are heavy).

The major consequence of iron deficiency is anaemia (p. 1021). This is one of the most important nutritional causes of ill health in all parts of the world. In the UK, it is estimated that 10% women are iron-deficient. Dietary iron overload is occasionally observed and results in iron accumulation in the liver and, rarely, cirrhosis. Haemochromatosis results from an inherited increase in iron absorption (p. 972).

**Iodine**

Iodine is required for synthesis of thyroid hormones (p. 738). It is present in sea fish, seaweed and most plant foods grown near the sea. The amount of iodine in soil and water influences the iodine content of most foods. Iodine is lacking in the highest mountainous areas of the world (e.g. the Alps and the Himalayas) and in the soil of frequently flooded plains (e.g. Bangladesh).

About a billion people in the world are estimated to have an inadequate iodine intake and hence are at risk of iodine deficiency disorder. Goitre is the most common manifestation, affecting about 200 million people (p. 752).

In those areas where most women have endemic goitre, 1% or more of babies are born with cretinism (characterised by mental and physical retardation). There is a higher than usual prevalence of deafness, slowed reflexes and poor learning in the remaining population. The best way of preventing neonatal cretinism is to ensure adequate levels of iodine during pregnancy. This can be achieved by intramuscular injections with 1–2 mL of iodised poppy seed oil (475–950 mg iodine) to women of child-bearing age every 3–5 years, by administration of iodised oil orally at 6-monthly or yearly intervals to adults and children, or by providing iodised salt for cooking.

**Zinc**

Zinc is present in most foods of vegetable and animal origin. It is an essential component of many enzymes, including carbonic anhydrase, alcohol dehydrogenase and alkaline phosphatase.

Acute zinc deficiency has been reported in patients receiving prolonged zinc-free parenteral nutrition and causes diarrhoea, mental apathy, a moist, eczematoid dermatitis, especially around the mouth, and loss of hair. Chronic zinc deficiency occurs in dietary deficiency, malabsorption syndromes, alcoholism and its associated hepatic cirrhosis. It causes the clinical features seen in the very rare congenital disorder known as acrodermatitis enteropathica (growth retardation, hair loss and chronic diarrhoea). Zinc deficiency is thought to be responsible for one-third of the world’s population not reaching their optimal height. In the Middle East, chronic deficiency has been associated with dwarfism and hypogonadism. In starvation, zinc deficiency

**5.36 Summary of clinically important minerals**

<table>
<thead>
<tr>
<th>Mineral</th>
<th>Rich</th>
<th>Important</th>
<th>Reference nutrient intake (RNI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>Milk and milk products, tofu</td>
<td>Milk, boned fish, green vegetables, beans</td>
<td>700 mg&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>Most foods contain phosphorus</td>
<td>Milk, cereal products, bread and meat</td>
<td>550 mg&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Magnesium</td>
<td>Whole grains, nuts</td>
<td>Unprocessed and wholegrain foods</td>
<td>300 mg men &lt;br&gt; 270 mg women&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Iron</td>
<td>Liver, red meat (haem iron)</td>
<td>Non-haem iron from vegetables, wholemeal bread</td>
<td>8.7 mg &lt;br&gt; 14.8 mg women &lt; 50 yrs</td>
</tr>
<tr>
<td>Zinc</td>
<td>Red meat, seafood</td>
<td>Dairy produce, wholemeal bread</td>
<td>9.5 mg men &lt;br&gt; 7 mg women&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Iodine</td>
<td>Edible seaweeds</td>
<td>Milk and dairy products</td>
<td>140 µg</td>
</tr>
<tr>
<td>Selenium</td>
<td>Fish, wheat grown in selenium-rich soils</td>
<td>Fish</td>
<td>75 µg men &lt;br&gt; 60 µg women&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Copper</td>
<td>Shellfish, liver</td>
<td>Bread, cereal products, vegetables</td>
<td>1.2 mg&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fluoride</td>
<td>Drinking water, tea</td>
<td>No RNI. Safe intake: 0.5 mg/kg</td>
<td></td>
</tr>
<tr>
<td>Potassium</td>
<td>Dried fruit, potatoes, coffee</td>
<td>Fresh fruit, vegetables, milk</td>
<td>3500 mg</td>
</tr>
<tr>
<td>Sodium</td>
<td>Table salt, anchovies</td>
<td>Processed foods, bread, bacon</td>
<td>1600 mg</td>
</tr>
</tbody>
</table>

<sup>1</sup>Rich sources contain the nutrient in high concentration but are not generally eaten in large amounts; important sources contain less but contribute most because larger amounts are eaten. <sup>2</sup>Increased amounts are required in women during lactation.
causes thymic atrophy, and zinc supplements may accelerate the healing of skin lesions, promote general well-being, improve appetite and reduce the morbidity associated with the under-nourished state, and lower the mortality associated with diarrhoea and pneumonia in children.

**Selenium**

The family of seleno-enzymes includes glutathione peroxidase, which helps prevent free radical damage to cells, and monodeiodinase, which converts thyroxine to triiodothyronine. North American soil has a higher selenium content than European and Asian soil, and the decreasing reliance of Europe on imported American food in recent decades has resulted in a decline in dietary selenium intake.

Selenium deficiency can cause hypothyroidism, cardiomyopathy in children (Keshan’s disease) and myopathy in adults. Excess selenium can cause heart disease.

**Fluoride**

Fluoride helps prevent dental caries, since it increases the resistance of the enamel to acid attack. It is a component of bone mineral and some studies have shown anti-fracture effects at low doses, but excessive intakes may compromise bone structure.

If the local water supply contains more than 1 part per million (ppm) of fluoride, the incidence of dental caries is low. Soft waters usually contain no fluoride, whilst very hard waters may contain over 10 ppm. The benefit of fluoride is greatest when it is taken before the permanent teeth erupt, while their enamel is being laid down. The addition of traces of fluoride (at 1 ppm) to public water supplies is now a widespread practice. Chronic fluoride poisoning is occasionally seen where the water supply contains >10 ppm fluoride. It can also occur in workers handling cryolite (aluminium sodium fluoride), used in smelting aluminium. Fluoride poisoning is described on page 223. Pitting of teeth is a result of too much fluoride as a child.

**Sodium, potassium and magnesium**

Western diets are high in sodium due to the sodium chloride (salt) that is added to processed food. In the UK, it is suggested that daily salt intakes are kept well below 6 g. The roles of sodium, potassium and magnesium, along with the disease states associated with abnormal intakes or disordered metabolism, are discussed in Chapter 16.

**Other essential inorganic nutrients**

These include chloride (a counter-ion to sodium and potassium), cobalt (required for vitamin B12), sulphur (a constituent of methionine and cysteine), manganese (needed for or activates many enzymes) and chromium (necessary for insulin action). Deficiency of chromium presents as hyperglycaemia and has been reported in adults as a rare complication of prolonged parenteral nutrition.

Copper metabolism is abnormal in Wilson’s disease (p. 973). Deficiency occasionally occurs but only in young children, causing microcytic hypochromic anaemia, neutropenia, retarded growth, skeletal rarefaction and dermatosis.

---

**Further information and acknowledgements**

**Websites**

- [www.diversalertnetwork.org](http://www.diversalertnetwork.org) Advice on the clinical management of diving illness and emergency assistance services.
- [www.hpa.org.uk/radiation](http://www.hpa.org.uk/radiation) The Health Protection Agency provides information and links on all forms of radiation for patients and professionals.
- [www.nice.org.uk](http://www.nice.org.uk) NICE guidelines for nutritional support and obesity.
- [www.who.int/nutrition](http://www.who.int/nutrition) WHO recommendations and intervention programmes for macronutrient- and micronutrient-related diseases.

**Telephone numbers**

- In the UK, two organisations provide advice on the clinical management of diving illness and the availability of the nearest recompression facility:
  - Aberdeen Royal Infirmary +44 (0)845 48 6008
  - Royal Navy +44 (0)7831 151523.
- Outside the UK, contact the Divers Alert Network (DAN; see above).

**Figure acknowledgements**

- Fig. 5.8 Crown copyright. Department of Health in association with the Welsh Government, the Scottish Government and the Food Standards Agency in North Ireland.
- Fig. 5.18B WHO. Report of a joint WHO/USAID meeting, vitamin A deficiency and xerophthalmia (WHO technical report series no. 5 W); 1976.
- Fig. 5.19 Karthikeyan K, Thappa DM. Pellagra and skin. Int J Dermatol 2002; 41:476–481.
- Fig. 5.20AB Ho V, Prinsloo P, Ombiga J. Persistent anaemia due to scurvy. Journal of the New Zealand Medical Association 2007; 120(1262):62. Reproduced with permission.
Principles of infectious disease

Infectious agents 134
Normal flora 136
Host–pathogen interactions 137
Investigation of infection 138
Direct detection 139
Culture 140
Specific immunological tests 141
Antimicrobial susceptibility testing 143
Epidemiology of infection 143
Infection prevention and control 145
Health care-acquired infection 145
Outbreaks of infection 147
Immunisation 148

Treatment of infectious diseases 149
Principles of antimicrobial therapy 149
Beta-lactam antibiotics 154
Macrolide and lincosamide antibiotics 156
Ketolides 156
Aminoglycosides 156
Quinolones and fluoroquinolones 157
Glycopeptides 157
Folate antagonists 158
Tetracyclines and glycylcyclines 158
Nitroimidazoles 158
Other antibacterial agents 158
Antifungal agents 159
Antiviral agents 160
Antiparasitic agents 162
Infection is the establishment of foreign organisms, or ‘infectious agents’, in or on a human host. This may result in colonisation, if the microorganism exists at an anatomical site without causing harm, or infectious disease, when the interaction between the host and microorganism (pathogen) results in illness. In clinical practice, the term ‘infection’ is often used interchangeably with ‘infectious disease’. Most pathogens are microorganisms, although some are multicellular organisms.

The host-pathogen interaction is dynamic and complex. Whilst it is rarely in the microorganism’s interest to kill the host (on which it relies for nutrition and protection), the manifestations of disease may aid its dissemination (e.g. diarrhoea, sneezing). Conversely, it is in the host’s interests to kill microorganisms likely to cause disease, whilst preserving colonising organisms, which may be beneficial.

Communicable diseases are caused by organisms transmitted between hosts, whereas endogenous diseases are caused by organisms already colonising the host. Cross-infection with colonising organisms (e.g. meticillin-resistant *Staphylococcus aureus*, MRSA) is both communicable and endogenous. Opportunistic infections may be communicable or endogenous and arise only in individuals with impaired host defence. The chain of infection (Fig. 6.1) describes six essential elements for communicable disease transmission.

Despite dramatic advances in hygiene, immunisation and antimicrobial therapy, infectious diseases are still a major cause of disease worldwide. Key challenges remain in tackling infection in resource-poor countries and in the emergence of new infectious agents and antimicrobial-resistant microorganisms. This chapter describes the biological and epidemiological principles of infectious diseases and the general approach to their prevention, diagnosis and treatment. Specific infectious diseases are described in Chapters 13–15 and many of the organ-based chapters.

### Infectious Agents

The concept of an infectious agent was established by Robert Koch in the 19th century (Box 6.1). Although fulfilment of ‘Koch’s postulates’ became the standard for the definition of an infectious agent, they do not apply to uncultivable organisms (e.g. *Mycobacterium leprae*, *Tropheryma whipplei*) or members of the normal human flora (e.g. *Escherichia coli*, *Candida spp.*). The following groups of infectious agents are now recognised.

#### Prions

Prions are unique amongst infectious agents in that they are devoid of any nucleic acid. They appear to be transmitted by acquisition of a normal mammalian protein (prion protein, PrP\(^\*)\) which is in an abnormal conformation (PrP\(^\text{sc}\), containing an excess of beta-sheet protein); the abnormal protein inhibits the 26S proteasome, which can degrade misfolded proteins, leading to accumulation of the abnormally configured PrP\(--\) protein instead of normal PrP\(--\). The result is accumulation of protein which forms amyloid in the central nervous system, causing a transmissible spongiform encephalopathy (see Box 13.40, p. 329, and p. 1211).

#### Viruses

Viruses are incapable of independent replication, instead subverting host cellular processes to ensure synthesis of their nucleic acids and proteins. A virus that infects a bacterium is a bacteriophage (phage). Viruses contain genetic material (genome), which may be single- or double-stranded DNA or RNA. Retroviruses transcribe their RNA into DNA by reverse transcription. An antigenically unique protein coat (capsid) encloses the genome, together forming the nucleocapsid. In many viruses, the nucleocapsid is packaged within a lipid envelope. Enveloped viruses are less able to survive in the environment and are spread by respiratory, sexual or blood-borne routes, including arthropod-based transmission.

![Fig. 6.1 Chain of infection.](image)

**Fig. 6.1 Chain of infection.** The infectious agent is the organism that causes the disease. The reservoir is the place where the population of an infectious agent is maintained. The portal of exit is the point from which the infectious agent leaves the reservoir. Transmission is the process by which the infectious agent is transferred from the reservoir to the human host, either directly or via a vector or fomite. The portal of entry is the body site that is first accessed by the infectious agent. Finally, in order for disease to ensue, the person to whom the infectious agent is transmitted must be a susceptible host.

**6.1 Definition of an infectious agent – Koch’s postulates**

1. The same organism must be present in every case of the disease
2. The organism must be isolated from the diseased host and grown in pure culture
3. The isolate must cause the disease, when inoculated into a healthy, susceptible animal
4. The organism must be re-isolated from the inoculated, diseased animal
Infectious agents

6

1. Adsorption
Interaction between host receptor molecule and virus ligand (determines host-specificity of the virus)

2. Penetration
Receptor-mediated endocytosis or, in some enveloped viruses, membrane fusion (shown here)

3. Uncoating
Nucleic acid is liberated from the phagosome (if endocytosed) and/or capsid by complex enzymatic and/or receptor-mediated processes

4. Synthesis
Nucleic acid and protein synthesis are mediated by host and/or viral enzymes. These take place in nucleus or cytoplasm, depending on the specific virus

5. Assembly
Assembly of virus components is mediated by host and/or viral enzymes

6. Release
Complete virus particles are released by budding of host cell membrane (shown here) or disintegration of host cell

Host cell

Lipid envelope
Capsid
Nucleic acid

Virus

Fig. 6.2 A generic virus life cycle. Life cycle components common to most viruses are host cell attachment and penetration, virus uncoating, nucleic acid and protein synthesis, virus assembly and release. Virus release is achieved either by budding, as illustrated, or by lysis of the cell membrane. Life cycles vary between viruses.

Transmission. Non-enveloped viruses survive better in the environment and are predominantly transmitted by faecal–oral or, less often, respiratory routes. A generic virus life cycle is shown in Figure 6.2.

Prokaryotes: bacteria (including mycobacteria and actinomycetes)

Prokaryotic cells are capable of synthesising their own proteins and nucleic acids, and are able to reproduce autonomously, although they lack a nucleus. The bacterial cell membrane is bounded by a peptidoglycan cell wall, which is thick (20–80 nm) in Gram-positive organisms and thin (5–10 nm) in Gram-negative ones. The Gram-negative cell wall is surrounded by an outer membrane containing lipopolysaccharide. Plasmids are rings of extra-chromosomal DNA within bacteria, which can be transferred between organisms. Bacteria may be embedded in a polysaccharide capsule, and motile bacteria are equipped with flagella. Although many prokaryotes are capable of independent existence, some (e.g. Chlamydia trachomatis, Coxiella burnetii) are obligate intracellular organisms. Bacteria that replicate in artificial culture media are classified and identified using a range of characteristics (Box 6.2), with examples in Figures 6.3 and 6.4.

Eukaryotes: fungi, protozoa and helminths

Eukaryotes contain functional organelles, including nuclei, mitochondria and Golgi apparatus. Eukaryotes involved in human infection include fungi, protozoa (unicellular eukaryotes with a flexible cell membrane, p. 353) and helminths (complex multicellular organisms including nematodes, trematodes and cestodes, p. 369).

Fungi exist as either moulds (filamentous fungi) or yeasts. Dimorphic fungi exist in either form, depending on environmental conditions (see Fig. 13.57, p. 382). The fungal plasma membrane differs from the human cell membrane in that it contains the sterol, ergosterol. Fungi have a cell wall made up of polysaccharides, chitin and manno-proteins. In most fungi, the main structural component of the cell wall is β-1,3-D-glucan, a glucose polymer.

Protozoa and helminths are often referred to as parasites. Many parasites have complex multi-stage life cycles, which involve animal and/or plant hosts in addition to humans.
Every human is host to an estimated $10^{13}$–$10^{14}$ colonising microorganisms, which constitute the normal flora. Resident flora are able to survive and replicate at a body site, whereas transient flora are present only for short periods.

Knowledge of non-sterile body sites and their normal flora is required to interpret culture results (Fig. 6.5).

The relationship between human host and normal flora is symbiotic, meaning that the organisms are in close proximity, and either mutualistic (both organisms benefit) or commensal (one organism benefits whilst the other derives neither benefit nor harm). The microbiome is the total burden of microorganisms, their genes and environmental interactions; the human microbiome is recognised increasingly as exerting a profound influence over human health and disease.

Maintenance of the normal flora is beneficial to health. For example, lower gastrointestinal tract bacteria synthesise and excrete vitamins (e.g. vitamins K and B₁₂); colonisation with normal flora confers ‘colonisation resistance’ to infection with pathogenic organisms by altering the local environment (e.g. lowering pH), producing antibacterial agents (e.g. bacteriocins, fatty acids and metabolic waste products), and inducing host antibodies which cross-react with pathogenic organisms.

Conversely, normally sterile body sites must be kept sterile. The mucociliary escalator transports environmental material deposited in the respiratory tract to the nasopharynx. The urethral sphincter prevents flow from the non-sterile urethra to the sterile bladder. Physical barriers, including the skin, lining of the gastrointestinal...
tract and mucous membranes, maintain sterility of the blood stream, peritoneal and pleural cavities, chambers of the eye, subcutaneous tissue and so on.

The normal flora contribute to endogenous disease by either excessive growth at the ‘normal’ site (overgrowth) or translocation to a sterile site. Overgrowth is exemplified by ‘blind loop’ syndrome (p. 882), dental caries and vaginal thrush, in which external factors favour overgrowth of specific components of the normal flora. Translocation results from spread along a surface or penetration of a closed barrier: for example, in urinary tract infection caused by perineal/enteric flora, and in surgical site infections, particularly of prosthetic materials, caused by skin flora such as staphylococci. Normal flora also contribute to disease by cross-infection, in which organisms that are colonising one individual cause disease when transferred to another, more susceptible, individual.

**HOST–PATHOGEN INTERACTIONS**

Pathogenicity is the capability of an organism to cause disease and virulence is the extent to which a pathogen is able to cause disease. Pathogens produce proteins and other factors, termed virulence factors, which interact with host cells to contribute to disease.

- **Primary pathogens** cause disease in a proportion of individuals to whom they are exposed, regardless of their immunological status.
- **Opportunistic pathogens** cause disease only in individuals whose host defences are compromised; for example, by genetic susceptibility or immunosuppressive disease or therapy.

**Characteristics of successful pathogens**

Successful pathogens have a number of attributes. They compete with host cells and colonising flora by various methods, including sequestration of nutrients, use of metabolic pathways not used by competing bacteria, and production of bacteriocins (small antimicrobial peptides/proteins that kill closely related bacteria). Motility enables pathogens to reach their site of infection, often in sterile sites that colonising bacteria do not reach, such as the distal airway. Many microorganisms, including viruses, use ‘adhesins’ to attach to host cells at the site of infection. Other pathogens can invade through tissues.
Pathogens may produce toxins, microbial molecules that cause adverse effects on host cells, either at the site of infection, or remotely following carriage through the blood stream. Endotoxin is the lipid A domain of Gram-negative bacterial outer membrane lipopolysaccharide. It is released when bacterial cells are damaged and has generalised inflammatory effects. Exotoxins are proteins released by living bacteria, which often have specific effects on target organs (Box 6.3).

Intracellular pathogens, including viruses, bacteria (e.g. Salmonella spp., Listeria monocytogenes and Mycobacterium tuberculosis), parasites (e.g. Leishmania spp.) and fungi (e.g. Histoplasma capsulatum), are able to survive in intracellular environments, including after phagocytosis by macrophages.

Pathogenic bacteria express different arrays of genes, depending on environmental stress (pH, iron starvation, O₂ starvation and so on) and anatomical location. In quorum sensing, bacteria communicate with one another to adapt their replication or metabolism according to local population density. Bacteria and fungi may respond to the presence of an artificial surface (e.g. prosthetic device, venous catheter) by forming a biofilm, which is a population of organisms encased in a matrix of extracellular molecules. Biofilm-associated organisms are highly resistant to antimicrobial agents.

Genetic diversity enhances the pathogenic capacity of bacteria. Some virulence factor genes are found on plasmids or in phages and are exchanged between different strains or species. The ability to acquire genes from the gene pool of all strains of the species (the ‘bacterial supragenome’) increases diversity and the potential for pathogenicity. Viruses exploit their rapid reproduction and potential to exchange nucleic acid with host cells to enhance diversity. Once a strain acquires a particularly effective combination of virulence genes, it may become an epidemic strain, accounting for a large subset of infections in a particular region. This phenomenon accounts for influenza pandemics.

### The host response

Innate and adaptive immune and inflammatory responses which humans use to control the normal flora and respond to pathogens are reviewed in Chapter 4.

---

### Pathogenesis of infectious disease

The harmful manifestations of infection are determined by a combination of the virulence factors of the organism and the host response to infection. Despite the obvious benefits of an intact host response, an excessive response is undesirable. Cytokines and antimicrobial factors contribute to tissue injury at the site of infection, and an excessive inflammatory response may lead to hypotension and organ dysfunction (p. 82). The contribution of the immune response to disease manifestations is exemplified by the immune reconstitution inflammatory syndrome (IRIS). This is seen, for example, in human immunodeficiency virus (HIV) infection, post-transplantation neutropenia or tuberculosis (which causes suppression of T-cell function): there is a paradoxical worsening of the clinical condition as the immune dysfunction is corrected, caused by an exuberant but dysregulated inflammatory response.

### The febrile response

Thermoregulation (p. 103) is altered in infectious disease. Microbial pyrogens or the endogenous pyrogens released during tissue necrosis stimulate specialised cells such as monocytes/macrophages to release cytokines, including interleukin (IL)-β, tumour necrosis factor-alpha (TNF)-α, IL-6 and interferon (IFN)-γ. Cytokine receptors in the pre-optic region of the anterior hypothalamus activate phospholipase A, releasing arachidonic acid as substrate for the cyclo-oxygenase pathway and producing prostaglandin E₂ (PGE₂), which in turn alters the responsiveness of thermosensitive neurons in the thermoregulatory centre. Rigors occur when the body inappropriately attempts to ‘reset’ core temperature to a higher level by stimulating skeletal muscle activity and shaking.

The role of the febrile response as a defence mechanism requires further study, but there are data to support the hypothesis that raised body temperature interferes with the replication and/or virulence of pathogens.
Direct detection methods provide rapid results and may be applied to organisms that cannot be grown easily on artificial culture media, such as Chlamydia spp. They do not usually provide information on antimicrobial susceptibility or the degree to which organisms are related to each other (which is important in the investigation of possible outbreaks), unless relevant specific nucleic acid sequences are detected by polymerase chain reaction (PCR).

Detection of whole organisms

Whole organisms are detected by examination of biological fluids or tissue using a microscope.

- **Bright field microscopy** (in which the test sample is interposed between the light source and the objective lens) uses stains to enhance visual contrast between the organism and its background. Examples include Gram staining of bacteria and Ziehl–Neelsen or auramine staining of acid- and alcohol-fast bacilli (AAFB) in tuberculosis. In histopathological examination of tissue samples, multiple stains are used to demonstrate not only the presence of microorganisms, but also features of disease pathology.

- **Dark field microscopy** (in which light is scattered to make organisms appear bright on a dark background) is used, for example, to examine genital chancre fluid in suspected syphilis.

- **Electron microscopy** may be used to examine stool and vesicle fluid to detect enteric and herpesviruses, respectively, but its use has largely been supplanted by nucleic acid detection (see below).

Detection of components of organisms

Components of microorganisms detected for diagnostic purposes include nucleic acids, cell wall molecules, toxins and other antigens. Commonly used examples include Legionella pneumophila serogroup 1 antigen in urine and cryptococcal polysaccharide antigen in cerebrospinal fluid (CSF). Most antigen detection methods are based on in vitro binding of specific antigen/antibody and are described below (p. 141). However, other methods may be used, such as mouse bioassay for detection of Clostridium botulinum toxin or tissue culture cytotoxicity assay for C. difficile toxin. In toxin-mediated disease, detection of toxin may be of greater relevance than identification of the organism itself (e.g. stool C. difficile toxin).

Nucleic acid amplification tests (NAAT)

Specific sequences of microbial DNA and RNA are identified using a nucleic acid primer which is amplified exponentially by enzymes to generate multiple copies of the specific sequence. The most commonly used amplification method is the polymerase chain reaction (PCR; see Fig. 3.12, p. 60). Reverse transcription (RT) PCR is used to detect RNA from RNA viruses.
PRINCIPLES OF INFECTIOUS DISEASE

(e.g. hepatitis C virus and HIV-1). The use of fluorescent-labelled primers and probes enables ‘real-time’ detection of amplified DNA; quantification is based on the principle that the time taken to reach the detection threshold is proportional to the initial number of copies of the target nucleic acid sequence. In multiplex PCR, multiple primer pairs are used to enable detection of several different organisms in a single assay.

Nucleic acid sequencing is also used to assign microorganisms to specific strains according to their genotype, which may be relevant to treatment and/or prognosis (e.g. in hepatitis C infection, p. 954). Genes that are relevant to pathogenicity (such as toxin genes) or antimicrobial resistance can also be detected. For example, detection of the mecA gene is used to screen for MRSA.

NAAT are the most sensitive direct detection methods and are particularly useful when a rapid diagnosis is required. They are used widely in virology, where the possibility of false-positive results from colonising or contaminating organisms is remote, and are applied to blood, respiratory samples, stool and urine. In bacteriology, PCR is used to examine CSF, blood, tissue and genital samples, and multiplex PCR is being developed for use in faeces. PCR is also being used increasingly in mycology and parasitology.

**Culture**

Microorganisms may be both detected and further characterised by culture from clinical samples (e.g. tissue, swabs and body fluids).

- **In vivo culture** (in a living organism) is not used in routine diagnostic microbiology.
- **Ex vivo culture** (tissue or cell culture) was widely used in the isolation of viruses, but has been largely supplanted by nucleic acid amplification techniques.

---

**Fig. 6.6** An overview of the processing of blood cultures. *In laboratories equipped with MALDI-TOF (p. 136), rapid definitive organism identification may be achieved at stage 6 and/or stage 8.*
**Investigation of infection**

**Antibody detection**

Organism-specific antibody detection is applied mainly to blood (Fig. 6.7). Results are typically expressed as titres: that is, the reciprocal of the highest dilution of the serum at which antibody is detectable (for example, detection at serum dilution of 1:64 gives a titre of 64). 'Seroconversion' is defined as either a change from negative to positive detection or a fourfold rise in titre between acute and convalescent serum samples. An acute sample is usually taken during the first week of disease and the convalescent sample 2–4 weeks later. Earlier diagnosis can be achieved by detection of IgM antibodies, which are produced early in infection (p. 77). A limitation of these tests is that antibody production requires a fully functional host immune system, so there may be false-negative results in immunocompromised patients. Also, other than in chronic infections and with IgM detection, antibody tests usually provide a retrospective diagnosis.

**Enzyme-linked immunosorbent assay**

The principles of the enzyme-linked immunosorbent assay (ELISA, EIA) are illustrated in Figure 6.8. These assays rely on linking an antibody with an enzyme which generates a colour change on exposure to a chromogenic substrate. Various configurations allow detection of antigens or specific subclasses of immunoglobulin (e.g. IgG, IgM, IgA). ELISA may also be adapted to detect PCR products, using immobilised oligonucleotide hybridisation probe and various detection systems.

**Immunoblot (Western blot)**

Microbial proteins are separated according to molecular weight by polyacrylamide gel electrophoresis (PAGE) and transferred (blotted) on to a nitrocellulose membrane, which is incubated with patient serum. Binding of specific antibody is detected with an **In vitro culture** (in artificial culture media) of bacteria and fungi is used for definitive identification, to test for antimicrobial susceptibility and to subtype the organism for epidemiological purposes.

However, culture has its limitations. Results are not immediate, even for organisms which are easy to grow, and negative culture rarely excludes infection completely. Organisms such as Mycobacterium tuberculosis are inherently slow-growing, typically taking at least 2 weeks to be detectable, even in specialised systems. Certain organisms, such as Mycobacterium leprae and Tropheryma whipplei, cannot be cultivated on artificial media, and others (e.g. Chlamydia spp. and viruses) grow only in ex vivo systems, which are slow and labour-intensive to use.

**Blood culture**

Rapid microbiological diagnosis is required for bloodstream infection (BSI; Fig. 6.6). To diagnose BSI, a liquid culture medium is inoculated with freshly drawn blood, transported to the microbiology laboratory and incubated in a system that monitors it constantly for products of microbial respiration (mainly CO₂), generally using fluorescence. If growth is detected, organisms are identified and sensitivity testing is performed. Traditionally, identification has been achieved by Gram stain and culture. However, MALDI-TOF (see Box 6.2) is being used increasingly, as it is rapid and inexpensive, and enables identification of organisms directly from the blood-culture medium.

**Specific immunological tests**

Immunological tests may be used to detect the host response to a specific microorganism, and can enable the diagnosis of infection with organisms that are difficult to detect by other methods or are no longer present in the host. The term ‘serology’ describes tests carried out on serum, and is used to include both antigen and antibody detection.

**Antibody detection**

Organism-specific antibody detection is applied mainly to blood (Fig. 6.7). Results are typically expressed as titres: that is, the reciprocal of the highest dilution of the serum at which antibody is detectable (for example, detection at serum dilution of 1:64 gives a titre of 64). ‘Seroconversion’ is defined as either a change from negative to positive detection or a fourfold rise in titre between acute and convalescent serum samples. An acute sample is usually taken during the first week of disease and the convalescent sample 2–4 weeks later. Earlier diagnosis can be achieved by detection of IgM antibodies, which are produced early in infection (p. 77). A limitation of these tests is that antibody production requires a fully functional host immune system, so there may be false-negative results in immunocompromised patients. Also, other than in chronic infections and with IgM detection, antibody tests usually provide a retrospective diagnosis.

**Enzyme-linked immunosorbent assay**

The principles of the enzyme-linked immunosorbent assay (ELISA, EIA) are illustrated in Figure 6.8. These assays rely on linking an antibody with an enzyme which generates a colour change on exposure to a chromogenic substrate. Various configurations allow detection of antigens or specific subclasses of immunoglobulin (e.g. IgG, IgM, IgA). ELISA may also be adapted to detect PCR products, using immobilised oligonucleotide hybridisation probe and various detection systems.

**Immunoblot (Western blot)**

Microbial proteins are separated according to molecular weight by polyacrylamide gel electrophoresis (PAGE) and transferred (blotted) on to a nitrocellulose membrane, which is incubated with patient serum. Binding of specific antibody is detected with an

---

*Fig. 6.7 Detection of antigen, nucleic acid and antibody in infectious disease.* The acute sample is usually taken during the first week of illness, and the convalescent sample 2–4 weeks later. Detection limits and duration of detectability vary between tests and diseases, although in most diseases immunoglobulin (Ig) M is detectable within the first 1–2 weeks.
enzyme–anti-immunoglobulin conjugate similar to that used in ELISA, and specificity is confirmed by its location on the membrane. Immunoblotting is a highly specific test, which may be used to confirm the results of less specific tests such as ELISA.

**Immunofluorescence assays**

Immunofluorescence assays (IFAs) are highly specific. In indirect immunofluorescence, a serum sample is incubated with immobilised antigen (e.g. cells known to be infected with virus on a glass slide) and antibody binding is detected using a fluorescent-labelled anti-human immunoglobulin (the ‘secondary’ antibody). This method can also detect organisms in clinical samples (usually tissue or centrifuged cells) using a specific antibody in place of patient serum. In direct immunofluorescence, clinical samples are incubated directly with fluorescent-labelled specific antibodies to detect antigen, eliminating the need for secondary antibody.

**Complement fixation test**

In a complement fixation test (CFT), patient serum is heat-treated to inactivate complement, and added to specific antigen. Any specific antibody present in the serum will complex with the antigen. Complement is then added to the reaction. If antigen–antibody complexes are present, the complement will be ‘fixed’ (consumed). Sheep erythrocytes, coated with an anti-erythrocyte antibody, are added. The degree of erythrocyte lysis reflects the remaining complement and is inversely proportional to the level of the specific antigen–antibody complexes.

**Agglutination tests**

When antigens are present on the surface of particles (e.g. cells, latex particles or microorganisms) and cross-linked with antibodies, visible clumping (or ‘agglutination’) occurs.

- **In direct agglutination**, patient serum is added to a suspension of organisms that express the test antigen. For example, in the Weil–Felix test, host antibodies to various rickettsial species cause agglutination of Proteus bacteria because they cross-react with bacterial cell surface antigens.

- **In indirect (passive) agglutination**, specific antigen is attached to the surface of carrier particles which agglutinate when incubated with patient samples that contain specific antibodies.

- **In reverse passive agglutination** (an antigen detection test), the carrier particle is coated with antibody rather than antigen.

**Other tests**

Immunodiffusion involves antibodies and antigen migrating through gels, with or without the assistance of electrophoresis, and forming insoluble complexes where they meet. The complexes are seen on staining as ‘precipitin bands’. Immunodiffusion is used in the diagnosis of endemic mycoses (p. 381) and some forms of aspergillosis (p. 697).

Immunochromatography is used to detect antigen. The system consists of a porous test strip (e.g. a nitrocellulose membrane), at one end of which there is target-specific antibody, complexed with coloured microparticles. Further specific antibody is immobilised in a transverse narrow line some distance along the strip. Test material (e.g. blood or urine) is added to the antibody–particle complexes, which then migrate along the strip by capillary action. If these are complexed with antigen, they will be immobilised by the specific antibody and visualised as a transverse line across the strip. If the test is negative, the antibody–particle complexes will bind to a line of immobilised anti-immunoglobulin antibody placed further along the strip, which acts as a negative control. Immunochromatographic tests are rapid and relatively cheap to perform, and are appropriate for point-of-care testing, e.g. in HIV 1.

**Antibody-independent specific immunological tests**

Interferon-gamma release assays (IGRA) are being used increasingly to diagnose tuberculosis (p. 692). The principle of the assay is that T lymphocytes of patients infected with *Mycobacterium tuberculosis* (MTB) release IFN-γ when they are exposed to MTB-specific peptides. The absence of these peptides in bacille Calmette–Guérin (BCG; see Box 6.14) vaccine results in IGRA tests being more specific for the diagnosis of tuberculosis.
infection than the tuberculin skin test (p. 692), because the latter may be positive as a result of previous BCG vaccination.

**Antimicrobial susceptibility testing**

If growth of microorganisms in culture is inhibited by the addition of an antimicrobial agent, the organism is considered to be susceptible. Bacteriostatic agents cause reversible inhibition of growth and bactericidal agents cause cell death; the terms fungistatic/fungicidal are equivalent for antifungal agents, and virustatic/virucidal for antiviral agents. The lowest concentration of antimicrobial agent at which growth is inhibited is the minimum inhibitory concentration (MIC), and the lowest concentration that causes cell death is the minimum bactericidal concentration (MBC). If the MIC is less than or equal to a predetermined breakpoint threshold, the organism is considered susceptible, and if the MIC is greater than the breakpoint, it is resistant.

Breakpoints are determined for each antimicrobial agent from a combination of pharmacokinetic and clinical data. The relationship between in vitro antimicrobial susceptibility and clinical response is complex, as response also depends on immune status, pharmacokinetic variability (p. 21), comorbidities that may influence pharmacokinetics or pharmacodynamics, and antibiotic dosing, as well as MIC/MBC. Thus, susceptibility testing does not guarantee therapeutic success.

Susceptibility testing is most often carried out by disc diffusion (Fig. 6.9). Antimicrobial-impregnated filter paper discs are placed on an agar plate containing bacteria. The antibiotic diffuses through the agar, resulting in a concentration gradient centred on the disc. Bacteria are unable to grow where the antibiotic concentration exceeds the MIC, which may therefore be inferred from the size of the zone of inhibition. Susceptibility testing methods using antimicrobials diluted in liquid media are generally more accurate and reproducible, and are used for generating epidemiological data.

**Epidemiology of infection**

The communicability of infectious disease means that, once a clinician has diagnosed an infectious disease, potential exposure of other patients must be considered. The patient may require treatment in isolation, or an outbreak of disease may need to be investigated in the community (Ch. 5). The approach will be specific to the microorganism involved (Chs 13-15) but the principles are outlined below.

**Geographic and temporal patterns of infection**

**Endemic disease**

Endemic disease has a constant presence within a given geographic area or population. The infectious agent may have a reservoir, vector or intermediate host that is geographically restricted, or may itself have restrictive environmental requirements (e.g. temperature range, humidity). The population affected may be geographically isolated, or the disease may be limited to unvaccinated populations. Factors that alter geographical restriction include:

- expansion of an animal reservoir (e.g. Lyme disease from deforestation)
- vector escape (e.g. airport malaria)
- extension of host range (e.g. schistosomiasis from dam construction)
- human migration (e.g. severe acute respiratory syndrome (SARS) coronavirus)
- public health service breakdown (e.g. diphtheria in unvaccinated areas)
- climate change.

**Emerging and re-emerging disease**

An emerging infectious disease is one that has newly appeared in a population, or has been known for some time but is increasing in incidence or geographic range. If the disease was previously known and thought to have been controlled or eradicated, it is considered to be re-emerging. Many emerging diseases are caused by organisms which infect animals and have undergone adaptations that enable them to infect humans. This is exemplified by HIV, which is believed to have originated in higher primates in Africa. The geographical pattern of some recent emerging and re-emerging infections is shown in Figure 6.10.

**Reservoirs of infection**

The US Centers for Disease Control (CDC) define a reservoir of infection as ‘one or more epidemiologically connected populations or environments in which a pathogen can be permanently maintained, and from which infection is transmitted to a defined target population’. Reservoirs of infection may be human, animal or environmental.

**Human reservoirs**

Colonised individuals or those with clinical infectious disease may act as reservoirs, e.g. for *Staph. aureus* (including MRSA), which is carried in the nares of 30–40% of humans, and *C. difficile*. For infected humans to act as reservoirs, the infections caused must be long-lasting and/or non-fatal, at least in a proportion of those affected, to enable onward transmission (e.g. tuberculosis, sexually transmitted infections). Humans are the only reservoir for some organisms (e.g. smallpox and measles).
Animal reservoirs
The World Health Organization (WHO) defines a zoonosis as ‘a disease or infection that is naturally transmissible from vertebrate animals to humans’. The infected animal may be asymptomatic. Zoonotic agents may be transmitted via any of the routes described below. Primary infection with zoonoses may be transmitted onward between humans, causing secondary disease (e.g. Q fever, brucellosis, Ebola).

Environmental reservoirs
Many infective pathogens are acquired from an environmental source. However, some of these are maintained in human or animal reservoirs, with the environment acting only as a conduit for infection.

Transmission of infection
Infectious agents may be transmitted by one or more of the following routes:

- **Respiratory route**: inhalation.
- **Faecal–oral route**: ingestion of infectious material originating from faecal matter.
- **Sexually transmitted infections**: direct contact between mucous membranes.
- **Blood-borne infections**: direct inoculation of infected blood or body fluids.
- **Direct contact**: very few organisms are capable of causing infection by direct contact with intact skin. Most infection by this route requires inoculation or contact with damaged skin.
- **Via a vector or fomite**: the vector/fomite bridges the gap between the infected host or reservoir and the uninfected host. Vectors are animate, and include mosquitoes in malaria and dengue, fleas in plague and humans in MRSA. Fomites are inanimate, and include items such as door handles, water taps, ultrasound probes and so on, which are particularly associated with health care-associated infection.

The likelihood of infection following transmission of an infectious agent depends on organism factors and host susceptibility. The number of organisms required to cause infection or death in 50% of the exposed population is referred to as the ID$_{50}$ (infectious dose) and LD$_{50}$ (lethal dose), respectively. The incubation period is the time between exposure and development of disease, and the period of infectivity is the period after exposure during which the patient is infectious to others. Knowledge of incubation periods and periods of infectivity is important in controlling the spread of disease, although for many diseases these estimates are imprecise (Boxes 6.6 and 6.7).

### Fig. 6.10 Geographic locations of some infectious disease outbreaks, with examples of emerging and re-emerging diseases.

(MDR-TB = multidrug-resistant tuberculosis; SARS = severe acute respiratory syndrome; vCJD = variant Creutzfeldt–Jakob disease; VRSA = vancomycin-resistant Staph. aureus) Adapted from Samaranayake 2006 – see p. 164.

### 6.6 Periods of infectivity in childhood infectious diseases

<table>
<thead>
<tr>
<th>Disease</th>
<th>Infectious period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chickenpox</td>
<td>From 4 days before$^1$ until 5 days after appearance of the rash$^2$</td>
</tr>
<tr>
<td>Measles</td>
<td>From 1–2 days before onset of rash; duration unknown$^3$</td>
</tr>
<tr>
<td>Mumps</td>
<td>Unknown$^4$</td>
</tr>
<tr>
<td>Rubella</td>
<td>Unknown, but most infectious during prodromal illness$^5$</td>
</tr>
<tr>
<td>Scarlet fever</td>
<td>Unknown$^5$</td>
</tr>
<tr>
<td>Whooping cough</td>
<td>Unknown$^5,6$</td>
</tr>
</tbody>
</table>

$^1$From Richardson M, Elliman D, Maguire H, et al. Pediatr Infect Dis J 2001; 20:380–388. These recommendations may differ from local or national guidance. $^2$Transmission before 48 hrs prior to the onset of rash is rare. $^3$Exclude from contact with non-immune and immunocompromised people for 5 days from onset of rash, onset of parotitis or start of antibiotic treatment. $^4$Exclude for 3 weeks if untreated.
### Deliberate release

The deliberate release of infectious agents with the intention of causing disease is known as biological warfare or bioterrorism, depending on the scale and context. Deliberate release incidents have included a 750-person outbreak of *Salmonella typhimurium* by contamination of salads in 1984 (Oregon, USA) and 22 cases of anthrax (five fatal) from the mailing of finely powdered (weaponised) anthrax spores in 2001 (New Jersey, USA). Diseases with high potential for deliberate release include anthrax, plague, tularemia, smallpox and botulism (through toxin release).

### INFECTION PREVENTION AND CONTROL

Infection prevention and control (IPC) describes the measures applied to populations with the aim of breaking the chain of infection (see Fig. 6.1, p. 134).

### Incubation periods of important infections

<table>
<thead>
<tr>
<th>Infection</th>
<th>Incubation period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Short incubation periods</strong></td>
<td></td>
</tr>
<tr>
<td>Anthrax, cutaneous</td>
<td>9 hrs to 2 wks</td>
</tr>
<tr>
<td>Anthrax, inhalational</td>
<td>2 days</td>
</tr>
<tr>
<td>Bacillary dysentery</td>
<td>1–6 days</td>
</tr>
<tr>
<td>Cholera</td>
<td>2 hrs to 5 days</td>
</tr>
<tr>
<td>Dengue haemorrhagic fever</td>
<td>3–14 days</td>
</tr>
<tr>
<td>Diphtheria</td>
<td>1–10 days</td>
</tr>
<tr>
<td>Gonorrohea</td>
<td>2–10 days</td>
</tr>
<tr>
<td>Influenza</td>
<td>1–3 days</td>
</tr>
<tr>
<td>Meningococcaemia</td>
<td>2–10 days</td>
</tr>
<tr>
<td>Norovirus</td>
<td>1–3 days</td>
</tr>
<tr>
<td>SARS coronavirus</td>
<td>2–7 days</td>
</tr>
<tr>
<td>Scarlet fever</td>
<td>2–4 days</td>
</tr>
<tr>
<td><strong>Intermediate incubation periods</strong></td>
<td></td>
</tr>
<tr>
<td>Amoebiasis</td>
<td>1–4 wks</td>
</tr>
<tr>
<td>Brucellosis</td>
<td>5–30 days</td>
</tr>
<tr>
<td>Chickenpox</td>
<td>11–20 days</td>
</tr>
<tr>
<td>Lassa fever</td>
<td>3–21 days</td>
</tr>
<tr>
<td>Malaria</td>
<td>10–15 days</td>
</tr>
<tr>
<td>Measles</td>
<td>6–19 days</td>
</tr>
<tr>
<td>Mumps</td>
<td>15–24 days</td>
</tr>
<tr>
<td>Poliomyelitis</td>
<td>3–35 days</td>
</tr>
<tr>
<td>Psittacosis</td>
<td>1–4 wks</td>
</tr>
<tr>
<td>Rubella</td>
<td>15–20 days</td>
</tr>
<tr>
<td>Typhoid</td>
<td>5–31 days</td>
</tr>
<tr>
<td>Whooping cough</td>
<td>5–21 days</td>
</tr>
<tr>
<td><strong>Long incubation periods</strong></td>
<td></td>
</tr>
<tr>
<td>Hepatitis A</td>
<td>3–7 wks</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>6 wks to 6 mths</td>
</tr>
<tr>
<td>Leishmaniasis, cutaneous</td>
<td>Weeks to months</td>
</tr>
<tr>
<td>Leishmaniasis, visceral</td>
<td>Months to years</td>
</tr>
<tr>
<td>Leprosy</td>
<td>5–20 yrs</td>
</tr>
<tr>
<td>Rabies</td>
<td>2–8 wks</td>
</tr>
<tr>
<td>Trypanosoma brucei gambiense</td>
<td>Months to years</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>1–12 mths</td>
</tr>
</tbody>
</table>

1 Incubation periods are approximate and may differ from local or national guidance. Longer incubation periods have been reported.

**Reference sources:**

- WHO. 1 Health Protection Agency (now Health Protection England).
- Centers for Disease Control, USA.

### Measures used in infection prevention and control (IPC)

**Institutional**

- Handling, storage and disposal of clinical waste
- Containment and safe removal of spilled blood and body fluids
- Cleanliness of environment and medical equipment
- Specialised ventilation (e.g. laminar flow, air filtration, controlled pressure gradients)
- Sterilisation and disinfection of instruments and equipment
- Food hygiene
- Laundry management

**Health-care staff**

- Education
- Hand hygiene, including hand-washing (see Fig. 6.12)
- Sharps management and disposal
- Use of personal protective equipment (masks, sterile and non-sterile gloves, gowns and aprons)
- Screening health workers for disease (e.g. tuberculosis, hepatitis B virus, MRSA)
- Immunisation and post-exposure prophylaxis

**Clinical practice**

- Antibiotic stewardship (use only when necessary; avoid drugs known to select multi-resistant organisms or predispose to other infections)
- Aseptic technique (see Box 6.10)
- Perioperative antimicrobial prophylaxis
- Screening patients for colonisation or infection (e.g. MRSA, GRE, CPE)

**Response to infections**

- Surveillance to detect alert organism (see text) outbreaks and antimicrobial resistance
- Antibiotic chemoprphyaxis to infectious disease contacts, if indicated (see Box 6.19)
- Isolation (see Box 6.9)
- Reservoir control
- Vector control

(CPE = carbapenemase-producing Enterobacteriaceae; GRE = glycopeptide-resistant enterococci; MRSA = meticillin-resistant *Staphylococcus aureus*)

---

### Health care-acquired infection

Admission to a health-care facility in the developed world carries a considerable risk of acquiring infection, estimated by the UK Department of Health as 6–10%. Factors that contribute to health care-acquired infection (HCAI, or nosocomial infection) are shown in Figure 6.11. Many nosocomial bacterial infections are caused by organisms that are resistant to numerous antibiotics (multi-resistant bacteria), including meticillin-resistant *Staph. aureus* (MRSA) (p. 330), extended-spectrum β-lactamase (ESBL)-producing Enterobacteriaceae, glycopeptide-resistant enterococci (GRE) and carbapenemase-producing Enterobacteriaceae (CPE). Other infections of particular concern in hospitals include *C. difficile* (p. 342) and norovirus (p. 327).

IPC measures are described in Box 6.8. The most important infection prevention practice is maintenance of good hand hygiene (Fig. 6.12). Hand decontamination or washing is mandatory before and after every patient contact. In most cases, decontamination with alcohol gel...
Fig. 6.11 Commonly encountered healthcare-associated infections (HCAI) and the factors that predispose to them. (ESBL = extended spectrum β-lactamases; GRE = glycopeptide-resistant enterococci; MRSA = multidrug-resistant Staph. aureus; RSV = respiratory syncytial virus)

Wash hands only when visibly soiled! Otherwise use handrub!

Duration of the entire procedure: 40–60 sec.

Fig. 6.12 Hand-washing. Good hand hygiene, whether with soap/water or alcohol handrub, includes areas that are often missed, such as fingertips, web spaces, palmar creases and the backs of hands. Adapted from WHO guidance at www.who.int – see p. 164.

is adequate. However, hand-washing (with hot water, liquid soap and complete drying) is required after any procedure that involves more than casual physical contact, or if hands are visibly soiled. In situations where the prevalence of C. difficile is high (e.g. a local outbreak), alcohol gel decontamination between patient contacts is inadequate, as it does not kill C. difficile spores, and hands must be washed with soap and water. Some infections necessitate additional measures to prevent cross-infection (Box 6.9). To avoid infection, all
### Infection prevention and control

#### 6.9 Types of isolation precaution[^1]

<table>
<thead>
<tr>
<th>Airborne transmission</th>
<th>Contact transmission</th>
<th>Droplet transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precautions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative pressure room with air exhausted externally or filtered</td>
<td>Private room preferred (otherwise, inter-patient spacing ≥ 1 m)</td>
<td>Private room preferred (otherwise, inter-patient spacing ≥ 1 m)</td>
</tr>
<tr>
<td>N95 masks or personal respirators for staff; avoid using non-immune staff</td>
<td>Gloves and gown for staff in contact with patient or contaminated areas</td>
<td>Surgical masks for staff in close contact with patient</td>
</tr>
</tbody>
</table>

#### Infections managed with these precautions

**Measles**
- Tuberculosis, pulmonary or laryngeal, confirmed or suspected

- Enteroviral infections in young children (diapered or incontinent)
- Norovirus[^2]
- C. difficile infection
- Multidrug-resistant organisms (e.g. MRSA, ESBL, GRE, VRSA, penicillin-resistant *Strep. pneumoniae*[^3]
- Parainfluenza in infants and young children
- Rotavirus
- RSV in infants, children and immunocompromised
- Viral conjunctivitis, acute

**Infections managed with multiple precautions**

- Smallpox, monkeypox, VZV (chickenpox or disseminated disease)[^4]
- SARS, viral haemorrhagic fever[^5]
- Adenovirus pneumonia

[^1]: Recommendations based on 2007 CDC guideline for isolation precautions. May differ from local or national recommendations. ^[2]: Not a CDC recommendation. ^[3]: Subject to local risk assessment. *Strap. pneumoniae* is used to refer to penicillin-resistant *Staph. aureus*.

[^4]: MRSA = meticillin-resistant *Staph. aureus*; ESBL = extended-spectrum *β*-lactamase; GRE = glycopeptide-resistant enterococci; MRSA = meticillin-resistant *Staph. aureus*; RSV = respiratory syncytial virus; SARS = severe acute respiratory syndrome; VRSA = vancomycin-resistant *Staph. aureus*; VZV = varicella zoster virus.

---


---

### Outbreaks of infection

Descriptive terms are defined in Box 6.11. Confirmation of an infectious disease outbreak usually requires evidence from typing (p. 138) that the causal organisms have identical genotypic characteristics. If this is found not to be the case, the term pseudo-outbreak is used. When an outbreak of infection is suspected, a case definition is agreed. The number of cases that meet the case definition is then assessed by case-finding, using methods ranging from administration of questionnaires to national reporting systems. Case-finding usually includes microbiological testing, at least in the early stages of an outbreak. Temporal changes in cases are noted in order to plot an outbreak curve, and demographic details are collected to identify possible sources of infection. A case control study, in which recent activities (potential exposures) of affected ‘cases’ are compared to those of unaffected ‘controls’, may be undertaken to establish the outbreak source, and measures are taken to manage the outbreak and control its spread. Good communication between relevant personnel during and after the outbreak is important to inform practice in future outbreaks.

Surveillance ensures that disease outbreaks are either pre-empted or identified early. In hospitals, staff are made aware of the isolation of alert organisms, which have the propensity to cause outbreaks, and alert conditions, which are likely to be caused by such organisms. Analogous systems are used nationally; many countries publish lists of organisms and diseases, which, if detected (or suspected), must be reported to public health authorities (reportable or notifiable diseases). Reasons for a disease to be classified as reportable are shown in Box 6.12.

### Principles of food hygiene

‘Food poisoning’ (p. 306) is largely preventable by food hygiene measures. The main principles are:

- segregation of uncooked food (which may be contaminated with pathogenic microorganisms) from cooked food
- avoidance of conditions which allow growth of pathogenic bacteria before or after cooking
- adequate bacterial killing during cooking.

Safe storage depends on the temperatures at which food bacteria are inhibited and destroyed (Fig. 6.13).
**6.11 Terminology in outbreaks of infection**

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification of related cases of infectious disease</strong></td>
<td></td>
</tr>
<tr>
<td>Cluster</td>
<td>An aggregation of cases of a disease which are closely grouped in time and place, and may or may not exceed the expected number of cases.</td>
</tr>
<tr>
<td>Epidemic</td>
<td>The occurrence of more cases of disease than expected in a given area or among a specific group of people over a particular period of time.</td>
</tr>
<tr>
<td>Outbreak</td>
<td>Synonymous with epidemic. Alternatively, a localised, as opposed to generalised, epidemic.</td>
</tr>
<tr>
<td>Pandemic</td>
<td>An epidemic occurring over a very wide area (several countries or continents) and usually affecting a large proportion of the population.</td>
</tr>
<tr>
<td><strong>Classification of affected patients (cases)</strong></td>
<td></td>
</tr>
<tr>
<td>Index case</td>
<td>The first case identified in an outbreak.</td>
</tr>
<tr>
<td>Primary cases</td>
<td>Cases acquired from a specific source of infection.</td>
</tr>
<tr>
<td>Secondary cases</td>
<td>Cases acquired from primary cases.</td>
</tr>
<tr>
<td><strong>Types of outbreak</strong></td>
<td></td>
</tr>
<tr>
<td>Common source outbreak</td>
<td>Exposure to a common source of infection (e.g. water-cooling tower, medical staff member shedding MRSA). New primary cases will arise until the source is no longer present.</td>
</tr>
<tr>
<td>Point source outbreak</td>
<td>Exposure to a single source of infection at a specific point in time (e.g. contaminated food at a party). Primary cases will develop disease synchronously.</td>
</tr>
<tr>
<td>Person-to-person spread</td>
<td>Outbreak with both primary and secondary cases. May complicate point source or common source outbreak.</td>
</tr>
</tbody>
</table>

*Adapted from www.cdc.gov.

**6.12 Reasons for including an infectious disease on a regional/national list of reportable diseases**

<table>
<thead>
<tr>
<th>Reason for inclusion</th>
<th>Common examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endemic/local disease with potential to spread and/or cause outbreaks</strong></td>
<td>Influenza, Salmonella, tuberculosis</td>
</tr>
<tr>
<td><strong>Imported disease with the propensity to spread and/or cause outbreaks</strong></td>
<td>Typhoid, cholera (depending on local epidemiology)</td>
</tr>
<tr>
<td><strong>Evidence of a possible breakdown in health protection/public health functions</strong></td>
<td>Legionella, Cryptosporidium</td>
</tr>
<tr>
<td><strong>Evidence of a possible breakdown in food safety practices</strong></td>
<td>Botulism, verotoxigenic E. coli</td>
</tr>
<tr>
<td><strong>Evidence of a possible failure of a vaccination programme</strong></td>
<td>Measles, poliomyelitis, pertussis</td>
</tr>
<tr>
<td><strong>Disease with the potential to be a novel or increasing threat to human health</strong></td>
<td>SARS, multi-resistant bacteria</td>
</tr>
<tr>
<td><strong>Evidence of expansion of the range of a reservoir/vector</strong></td>
<td>Lyme disease, rabies, West Nile encephalitis</td>
</tr>
<tr>
<td><strong>Evidence of possible deliberate release</strong></td>
<td>Anthrax, tularemia, plague, smallpox, botulism</td>
</tr>
</tbody>
</table>

*Given the different geographic ranges of individual diseases, and wide national variations in public health services, vaccination programmes and availability of resources, reporting regulations vary between regions, states and countries. Many diseases are reportable for more than one reason.

---

**Immunisation**

Immunisation may be passive or active. Passive immunisation is achieved by administering antibodies targeted against a specific pathogen. Antibodies are obtained from blood, so confer some of the risks associated with blood products (p. 1012). The protection afforded by passive immunisation is immediate but of short duration (a few weeks or months); it is used to prevent or attenuate infection before or after exposure (Box 6.13).

**Vaccination**

Active immunisation is achieved by vaccination with whole organisms or organism components (Box 6.14).

**Types of vaccine**

Whole cell vaccines consist of live or inactivated (killed) microorganisms; component vaccines contain only extracted or synthesised components of microorganisms (e.g. polysaccharides or proteins). Live vaccines contain organisms with attenuated (reduced) virulence, which induce T-lymphocyte and humoral responses (p. 77) and are therefore more immunogenic than inactivated whole cell vaccines. The use of live vaccines in immunocompromised individuals requires careful consideration.

Component vaccines consisting only of polysaccharides, such as the pneumococcal polysaccharide vaccine (PPV), are poor activators of T-lymphocytes, and produce a short-lived antibody response without long-lasting memory. Conjugation of polysaccharide to...
a protein, as in the Haemophilus influenzae type B (Hib) vaccine, activates T lymphocytes, which results in a sustained response and immunological memory. Toxoids are bacterial toxins that have been modified to reduce toxicity but maintain antigenicity. Vaccine response can be improved by co-administration with mildly pro-inflammatory adjuvants, such as aluminium hydroxide.

**Use of vaccines**

Vaccination may be applied to entire populations or to subpopulations at specific risk through travel, occupation or other activities. In ring vaccination, the population immediately surrounding a case or outbreak of infectious disease is vaccinated to curtail further spread. Vaccination is aimed mainly at preventing infectious disease. However, vaccination against human papillomavirus (HPV) was introduced to prevent cervical and other cancers which complicate HPV infection. Vaccination becomes successful once the number of susceptible hosts in a population falls below the level required to sustain continued transmission of the target organism (herd immunity). Naturally acquired smallpox was declared to have been eradicated worldwide in 1980 through mass vaccination. In 1988, the WHO resolved to eradicate poliomyelitis by vaccination; the number of cases worldwide has since fallen from 350,000 per annum to 223 in 2012. Recommended vaccination schedules vary between countries. In addition to standard vaccination schedules, catch-up schedules are specified for individuals who join vaccination programmes later than the recommended age.

**TREATMENT OF INFECTIOUS DISEASES**

The key components of treating infectious disease are:

- addressing predisposing factors, e.g. diabetes mellitus or known immune deficit (HIV, neutropenia)
- antimicrobial therapy
- adjuvant therapy, e.g. removal of an indwelling catheter (urinary or vascular), abscess drainage or débridement of an area of necrotising fasciitis
- treatment of the consequences of infection, e.g. the systemic inflammatory response syndrome (SIRS; p. 184), inflammation and pain.

For communicable disease, treatment must also take into account contacts of the infected patient, and may include infection prevention and control activities such as isolation, antimicrobial prophylaxis, vaccination and contact tracing.

**Principles of antimicrobial therapy**

When infection is diagnosed, it is important to start appropriate antimicrobial therapy promptly. The principles underlying the choice of antimicrobial agent(s) are

---

### 6.13 Indications for post-exposure prophylaxis with immunoglobulins

**Human normal immunoglobulin (pooled immunoglobulin)**

- Hepatitis A (unvaccinated contacts*)
- Measles (if exposed child has heart or lung disease)

**Human specific immunoglobulin**

- Hepatitis B (sexual partners, inoculation injuries, infants born to infected mothers)
- Tetanus (high-risk wounds or when immunisation status is incomplete or unknown)
- Rabies
- Chickenpox (immunosuppressed children and adults, pregnant women)

*Active immunisation is preferred if contact is with a patient who is within 1 wk of onset of jaundice.

### 6.14 Vaccines in current clinical use

**Live attenuated vaccines**

- Measles, mumps, rubella (MMR)
- Oral poliomyelitis (OPV, not used in UK)
- Tuberculosis (bacille Calmette–Guérin, BCG)
- Typhoid (oral typhoid vaccine)
- Varicella zoster virus
- Rotavirus

**Inactivated (killed) whole-cell vaccines**

- Cholera
- Hepatitis A
- Poliomyelitis (inactivated polio virus, IPV)
- Rabies
- Influenza
- Human papillomavirus (recombinant capsid proteins)
- Haemophilus influenzae type B (conjugated capsular polysaccharide)
- Meningococcal, quadrivalent A, C, Y, W135 (conjugated capsular polysaccharide)
- Meningococcal, serogroup C (conjugated capsular polysaccharide)
- Pneumococcal conjugate (PCV; conjugated capsular polysaccharide, 13 serotypes)
- Pneumococcal polysaccharide (PPV; purified capsular polysaccharide, 23 serotypes)
- Haemophilus influenzae type B (Hib) vaccine
- Typhoid (purified Vi capsular polysaccharide)

**Component vaccines**

- Anthrax (adsorbed extracted antigens)
- Diphtheria (adsorbed toxoid)
- Hepatitis B (adsorbed recombinant HBsAg)
- Pertussis (adsorbed extracted antigens)
- Pneumococcal conjugate (PCV; conjugated capsular polysaccharide, 13 serotypes)
- Pneumococcal polysaccharide (PPV; purified capsular polysaccharide, 23 serotypes)
- Tetanus (adsorbed toxoid)
- Typhoid (purified Vi capsular polysaccharide)

---

### 6.15 Guidelines for vaccination against infectious disease

- The principal contraindication to inactivated vaccines is an anaphylactic reaction to a previous dose or a component of the vaccine
- Live vaccines should not be given in the presence of an acute infection, to pregnant women or to the immunosuppressed, unless the immunosuppression is mild and the benefits outweigh the risks
- If two live vaccines are required, they should be given either simultaneously in opposite arms or 4 wks apart
- Live vaccines should not be given for 3 mths after an injection of human normal immunoglobulin (HNI)
- HNI should not be given for 2 wks after a live vaccine
- Hay fever, asthma, eczema, sickle-cell disease, topical corticosteroid therapy, antibiotic therapy, prematurity and chronic heart and lung diseases, including tuberculosis, are not contraindications to vaccination
discussed below. The process of selecting appropriate antimicrobial therapy has been summarised in UK guidance as ‘Start Smart – Then Focus’ (Fig. 6.14).

**Antimicrobial action and spectrum**

Antimicrobial agents kill microorganisms by inhibiting, damaging or destroying a target that is a required component of the organism. The range, or spectrum, of microorganisms that is killed by a particular antimicrobial agent must be considered in selecting therapy. The mechanisms of action of the major classes of antibacterial agent are listed in Box 6.16 and appropriate antibiotic choices for a range of common infecting organisms are shown in Box 6.17. In severe infections and/or immunocompromised patients, it is customary to use bactericidal agents in preference to bacteriostatic agents.

**Empiric versus targeted therapy**

Empiric antimicrobial therapy is selected to treat a clinical syndrome (e.g. meningitis) before a microbiological diagnosis has been made. Targeted therapy is aimed at the causal pathogen(s) of known antimicrobial sensitivity. ‘Start Smart – Then Focus’ describes the principle of using appropriate broad-spectrum agents in empiric therapy, followed by narrow-spectrum agents in targeted therapy. Optimum empiric therapy depends on the site of infection, patient characteristics and local antimicrobial resistance patterns. Hospital antibiotic policies are used to guide rational antimicrobial prescribing, maximising efficacy while minimising antimicrobial resistance and cost.

**Combination therapy**

It is sometimes appropriate to use antimicrobial agents in combination:

- to increase efficacy (e.g. enterococcal endocarditis, where a β-lactam/aminoglycoside combination
### 6.17 Antimicrobial options for common infecting bacteria

<table>
<thead>
<tr>
<th>Organism</th>
<th>Antimicrobial options*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-positive organisms</strong></td>
<td></td>
</tr>
<tr>
<td><em>Enterococcus faecalis</em></td>
<td>Ampicillin, vancomycin/teicoplanin</td>
</tr>
<tr>
<td><em>Enterococcus faecium</em></td>
<td>Vancomycin/teicoplanin, linezolid</td>
</tr>
<tr>
<td>Glycopeptide-resistant enterococci (GRE)</td>
<td>Linezolid, tigecycline, quinupristin–dalfopristin, daptomycin</td>
</tr>
<tr>
<td>MRSA</td>
<td>Clindamycin, vancomycin, rifampicin (never used as monotherapy), linezolid, daptomycin, tetracyclines, tigecycline, co-trimoxazole</td>
</tr>
<tr>
<td><em>Staph. aureus</em></td>
<td></td>
</tr>
<tr>
<td><em>Strep. pyogenes</em></td>
<td>Penicillin, clindamycin, erythromycin</td>
</tr>
<tr>
<td><em>Strep. pneumoniae</em></td>
<td>Penicillin, macrolides, cephalosporins, levofloxacin, vancomycin</td>
</tr>
<tr>
<td><strong>Gram-negative organisms</strong></td>
<td></td>
</tr>
<tr>
<td><em>E. coli, ‘coliforms’</em> (enteric Gram-negative bacilli)</td>
<td>Trimethoprim, cefuroxime, ciprofloxacin, co-amoxiclav</td>
</tr>
<tr>
<td><em>Enterobacter spp.</em>, <em>Citrobacter spp.</em></td>
<td>Ciprofloxacin, meropenem, ertapenem, aminoglycosides</td>
</tr>
<tr>
<td>ESBL-producing Enterobacteriaceae</td>
<td>Ciprofloxacin, meropenem, ertapenem (if sensitive), piperacillin–tazobactam, aminoglycosides, tigecycline</td>
</tr>
<tr>
<td>Carbapenemase-producing Enterobacteriaceae</td>
<td>Ciprofloxacin, aminoglycosides, tigecycline, colistin</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>Amoxicillin, co-amoxiclav, macrolides, cefuroxime, cefotaxime, ciprofloxacin</td>
</tr>
<tr>
<td><em>Legionella pneumophila</em></td>
<td>Azithromycin, levofloxacin, doxycycline</td>
</tr>
<tr>
<td><em>Neisseria gonorrhoeae</em></td>
<td>Ceftriaxone/cefexime, spectinomycin</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em></td>
<td>Penicillin, cefotaxime/ceftriaxone, chloramphenicol</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>Ciprofloxacin, piperacillin–tazobactam, aztreonam, meropenem, aminoglycosides, ceftazidime/cefepine</td>
</tr>
<tr>
<td><em>Salmonella typhi</em></td>
<td>Ceftriaxone, azithromycin (uncomplicated typhoid), chloramphenicol (resistance common)</td>
</tr>
<tr>
<td><strong>Strict anaerobes</strong></td>
<td></td>
</tr>
<tr>
<td><em>Bacteroides spp.</em></td>
<td>Metronidazole, clindamycin, co-amoxiclav, piperacillin–tazobactam, meropenem</td>
</tr>
<tr>
<td><em>Clostridium difficile</em></td>
<td>Metronidazole, vancomycin (oral), fidaxomycin</td>
</tr>
<tr>
<td><em>Clostridium spp.</em></td>
<td>Penicillin, metronidazole, clindamycin</td>
</tr>
<tr>
<td><em>Fusobacterium spp.</em></td>
<td>Penicillin, metronidazole, clindamycin</td>
</tr>
<tr>
<td><strong>Other organisms</strong></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydia trachomatis</em></td>
<td>Azithromycin, doxycycline</td>
</tr>
<tr>
<td><em>Treponema pallidum</em></td>
<td>Penicillin, doxycycline</td>
</tr>
</tbody>
</table>

*Antibiotic selection depends on multiple factors, including local susceptibility patterns, which vary enormously between geographic areas. There are many appropriate alternatives to those listed.*

Results in better outcomes than a β-lactam alone
- when no single agent’s spectrum covers all potential pathogens (e.g. in polymicrobial infection or empiric treatment of sepsis)
- to reduce antimicrobial resistance, as the organism would need to develop resistance to multiple agents simultaneously (e.g. *antituberculous chemotherapy* (p. 693), antiretroviral therapy (ART, p. 407)).

**Antimicrobial resistance**

Microorganisms have evolved in the presence of naturally occurring antibiotics, and have therefore developed resistance mechanisms (categorised in Fig. 6.15) to all classes of antimicrobial agent (antibiotics and their derivatives). Intrinsic resistance is an innate property of a microorganism, whereas acquired resistance arises by spontaneous mutation or horizontal transfer of genetic material from another organism in a phage or plasmid. Plasmids often encode resistance to multiple antibiotics. For some agents, e.g. penicillins, a degree of resistance occurs in vivo when the bacterial load is high and the molecular target for the antimicrobial is down-regulated (an ‘inoculum effect’).

The mecA gene encodes a low-affinity penicillin-binding protein, which confers resistance to β-lactam antibiotics in staphylococci. Extended spectrum β-lactamases (ESBL) are encoded on plasmids which are transferred relatively easily between bacteria, including Enterobacteriaceae. Plasmid-encoded carbapenemases have been detected in strains of *Klebsiella pneumoniae* (e.g. New Delhi metallo-β-lactamase 1, NDM-1). Strains of MRSA have been described that exhibit intermediate resistance to glycopeptides (GISA) through the development of a relatively impermeable cell wall.

Factors promoting antimicrobial resistance include the inappropriate use of antibiotics (e.g. in viral infections), inadequate dosage or treatment duration, and use
Fig. 6.15 Examples of mechanisms of antimicrobial resistance. (CAT = chloramphenicol acetyltransferase; ESBL = extended spectrum β-lactamases; GISA = glycopeptide-intermediate Staph. aureus; MRSA = meticillin-resistant Staph. aureus; NDM-1 = New Delhi metallo-β-lactamase 1.)

of antimicrobials as growth-promoters in agriculture. However, *any* antimicrobial use exerts a selection pressure that favours the development of resistance. Combination antimicrobial therapy may reduce the emergence of resistance. This is recommended in treatment of patients infected with HIV, which is highly prone to spontaneous mutation (p. 407). Despite use of combination therapy for *M. tuberculosis*, multidrug-resistant tuberculosis (MDR-TB, resistant to isoniazid and rifampicin) and extremely drug-resistant tuberculosis (XDR-TB, resistant to isoniazid and rifampicin, any fluoroquinolone and at least one injectable antimicrobial antituberculous agent) have been reported worldwide and are increasing in incidence (p. 693).

The term post-antibiotic era has been coined to describe a future in which the acquisition of resistance by bacteria will have been so extensive that antibiotic therapy is rendered useless. A more realistic scenario, which is currently being experienced, is a gradual but inexorable progression of resistance, necessitating the use of ever more toxic and expensive antimicrobials.

**Duration of therapy**

Treatment duration reflects the severity of infection and accessibility of the infected site to antimicrobial agents. For most infections, there is limited evidence available to support a specific duration of treatment (Box 6.18). Depending on the indication, initial intravenous therapy may be switched to oral after fever has settled for approximately 48 hours. In the absence of specific guidance, antimicrobial therapy should be stopped when there is no longer any clinical evidence of infection.

**Antimicrobial prophylaxis**

Primary prophylaxis is used when there is a risk of infection from a procedure or exposure (Box 6.19). It should be of short duration with minimal adverse effects, and may be combined with passive immunisation (see Box 6.13). Secondary prophylaxis is used in patients who have been treated successfully for an infection but remain predisposed to it. It is used in haemato-oncology patients in the context of fungal infection and in HIV-positive individuals with an opportunistic infection who do not respond to antiretroviral therapy.

**Pharmacokinetics and pharmacodynamics**

Pharmacokinetics of antimicrobial agents determine whether adequate concentrations are obtained at the sites of infection. Septic patients often have poor gastrointestinal absorption, so the preferred initial route of therapy is intravenous. Knowledge of anticipated antimicrobial drug concentrations at sites of infection is critical. For example, achieving a ‘therapeutic’ blood level of gentamicin is of little practical use in treating meningitis, as CSF penetration of the drug is poor. Knowledge of routes of antimicrobial elimination is also critical; for instance, urinary tract infection is ideally treated with a drug that is excreted unchanged in the urine.

Pharmacodynamics describes the relationship between antimicrobial concentration and microbial
Fig. 6.16  Antimicrobial pharmacodynamics. The curve represents drug concentrations after a single dose of an antimicrobial agent. Factors that determine microbial killing are $C_{\text{max}}$:MIC ratio (concentration-dependent killing), time above MIC (time-dependent killing), and AUC : MIC ratio.

### Infection Duration of therapy

<table>
<thead>
<tr>
<th>Infection</th>
<th>Duration of therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral infections</td>
<td></td>
</tr>
<tr>
<td>Herpes simplex encephalitis</td>
<td>2–3 wks</td>
</tr>
<tr>
<td>Bacterial infections</td>
<td></td>
</tr>
<tr>
<td>Gonorrhoea</td>
<td>Single dose</td>
</tr>
<tr>
<td>Infective endocarditis</td>
<td>4 wks ± gentamicin for first 1 wk</td>
</tr>
<tr>
<td>(streptococcal, native valve)</td>
<td></td>
</tr>
<tr>
<td>Infective endocarditis</td>
<td>$\geq$ 6 wks</td>
</tr>
<tr>
<td>(prosthetic valve)</td>
<td></td>
</tr>
<tr>
<td>Osteomyelitis</td>
<td>4–6 wks</td>
</tr>
<tr>
<td>Pneumonia (community-acquired, severe)</td>
<td>10 days (no organism identified), 14–21 days (Staph. aureus or Legionella spp.)</td>
</tr>
<tr>
<td>Septic arthritis</td>
<td>2–4 wks</td>
</tr>
<tr>
<td>Urinary tract infection (male)</td>
<td>2 wks</td>
</tr>
<tr>
<td>Urinary tract infection, upper (female)</td>
<td>7 days</td>
</tr>
<tr>
<td>Urinary tract infection, lower (female)</td>
<td>3 days</td>
</tr>
<tr>
<td>Mycobacterial infections</td>
<td></td>
</tr>
<tr>
<td>Tuberculosis (meningal)</td>
<td>12 mths</td>
</tr>
<tr>
<td>Tuberculosis (pulmonary)</td>
<td>6 mths</td>
</tr>
<tr>
<td>Fungal infections</td>
<td></td>
</tr>
<tr>
<td>Invasive pulmonary aspergillosis</td>
<td>Until clinical/radiological resolution and reversal of predisposition</td>
</tr>
<tr>
<td>Candidaemia (acute disseminated)</td>
<td>2 wks after last positive blood culture and resolution of signs and symptoms</td>
</tr>
</tbody>
</table>

*All recommendations are indicative. Actual duration takes into account predisposing factors, specific organisms and antimicrobial susceptibility, adjuvant therapies, current guidelines and clinical response.

### Infection prophylaxis in adults

<table>
<thead>
<tr>
<th>Infection risk</th>
<th>Recommended antimicrobial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial</td>
<td></td>
</tr>
<tr>
<td>Diphtheria (prevention of secondary cases)</td>
<td>Erythromycin</td>
</tr>
<tr>
<td>Gas gangrene (after high amputation or major trauma)</td>
<td>Penicillin or metronidazole</td>
</tr>
<tr>
<td>Lower gastrointestinal tract surgery</td>
<td>Cefuroxime + metronidazole, gentamicin + metronidazole, or co-amoxiclav (single dose only)</td>
</tr>
<tr>
<td>Meningococcal disease (prevention of secondary cases)</td>
<td>Rifampicin or ciprofloxacin</td>
</tr>
<tr>
<td>Rheumatic fever (prevention of recurrence)</td>
<td>Phenoxymethylpenicillin or sulfadiazine</td>
</tr>
<tr>
<td>Tuberculosis (prevention of secondary cases)</td>
<td>Isoniazid ± rifampicin</td>
</tr>
<tr>
<td>Whooping cough (prevention of secondary cases)</td>
<td>Erythromycin</td>
</tr>
<tr>
<td>Viral</td>
<td></td>
</tr>
<tr>
<td>HIV, occupational exposure (sharps injury)</td>
<td>Combination tenofovir/ emtricitabine and lopinavir/ ritonavir. Modified if index case’s virus known to be resistant</td>
</tr>
<tr>
<td>Influenza A (prevention of secondary cases in adults with chronic respiratory, cardiovascular or renal disease, immunosuppression or diabetes mellitus)</td>
<td>Oseltamivir</td>
</tr>
<tr>
<td>Fungal</td>
<td></td>
</tr>
<tr>
<td>Aspergillosis (in high-risk haematology patients)</td>
<td>Itraconazole, voriconazole or posaconazole</td>
</tr>
<tr>
<td>Pneumocystis pneumonia (prevention in HIV and other immunosuppressed states)</td>
<td>Co-trimoxazole, pentamidine or dapsone</td>
</tr>
<tr>
<td>Protozoal</td>
<td></td>
</tr>
<tr>
<td>Malaria (prevention of travel-associated disease)</td>
<td>Specific antimalarials depend on travel itinerary (p. 357)</td>
</tr>
</tbody>
</table>

*These are based on current UK practice. Recommendations may vary locally or nationally. Antimicrobial prophylaxis for infective endocarditis during dental procedures is not currently recommended in the UK.

killing. For many agents, antimicrobial effect can be categorised as concentration-dependent or time-dependent. The concentration of antimicrobial achieved after a single dose is illustrated in Figure 6.16. The maximum concentration achieved is $C_{\text{max}}$ and the measure of overall exposure is the area under the curve (AUC). The efficacy of antimicrobial agents whose killing is concentration-dependent (e.g. aminoglycosides) increases with the amount by which $C_{\text{max}}$ exceeds the minimum inhibitory concentration ($C_{\text{max}}$:MIC ratio). For this reason, it has become customary to administer aminoglycosides (e.g. gentamicin) infrequently at high doses (e.g. 7 mg/kg) rather than frequently at low doses. This has the added advantage of minimising toxicity by reducing the likelihood of drug accumulation. Conversely, the β-lactam antibiotics, macrolides and clindamycin exhibit time-dependent killing, and their efficacy depends on $C_{\text{max}}$.\]
6.20 Antimicrobial agents in pregnancy

Contraindicated
- Chloramphenicol: neonatal ‘grey baby’ syndrome – collapse, hypotension and cyanosis
- Fluconazole: teratogenic in high doses
- Quinolones: arthropathy in animal studies
- Sulphonamides: neonatal haemolysis and methaemoglobinaemia
- Tetracyclines, glycyclines: skeletal abnormalities in animals in 1st trimester; fetal dental discoloration and maternal hepatotoxicity with large parenteral doses in 2nd or 3rd trimesters
- Trimethoprim: teratogenic in 1st trimester

Relatively contraindicated
- Aminoglycosides: potential damage to fetal auditory and vestibular nerves in 2nd and 3rd trimesters
- Metronidazole: avoidance of high dosages is recommended

Not known to be harmful; use only when necessary
- Aciclovir
- Cephalosporins
- Clarithromycin
- Clindamycin
- Erythromycin
- Glycopeptides
- Linezolid
- Meropenem
- Penicillins

2Theoretical risk of teratogenicity, not supported by available clinical evidence.

6.21 Problems with antimicrobial therapy in old age
- *Clostridium difficile* infection: all antibiotics predispose to some extent, but second- and third-generation cephalosporins and co-amoxiclav especially so.
- Hypersensitivity reactions: rise in incidence due to increased previous exposure.
- Renal impairment: may be significant in old age, despite ‘normal’ creatinine levels (p. 467).
- Nephrotoxicity: more likely, e.g. first-generation cephalosporins, aminoglycosides.
- Accumulation of β-lactam antibiotics: may result in myoclonus, seizures or coma.
- Reduced gastric acid production: gastric pH is higher, which causes increased penicillin absorption.
- Reduced hepatic metabolism: results in a higher risk of isoniazid-related hepatotoxicity.
- Quinolones: associated with confusion and may increase the risk of seizures.

6.22 Beta-lactam antibiotics

These antibiotics have a β-lactam ring structure (Fig. 6.17) and exert a bactericidal action by inhibiting enzymes involved in cell wall synthesis (penicillin-binding proteins, PBP). They are classified in Box 6.22.

Pharmacokinetics
- Good drug levels are achieved in lung, kidney, bone, muscle and liver, and in pleural, synovial, pericardial and peritoneal fluids.
- CSF levels are low, except in the presence of inflammation.
- Activity is not inhibited in abscess (e.g. by low pH and PO₂, high protein or neutrophils).
- Beta-lactams are subject to an ‘inoculum effect’ – activity is reduced in the presence of a high organism burden (PBP expression is down-regulated by high organism density).
- Generally safe in pregnancy (except imipenem/cilastatin).

Adverse effects
Generalised allergy to penicillin occurs in 0.7–10% of cases and anaphylaxis in 0.004–0.015%. A large proportion of patients with infectious mononucleosis develop a rash if given aminopenicillins; this does not imply lasting allergy. The relationship between allergy to penicillin and allergy to cephalosporins depends on the specific cephalosporin used. Although there is significant cross-reactivity with first-generation cephalosporins, cross-reactivity to second- and third-generation cephalosporins is less common. However, avoidance of cephalosporins is recommended in patients who have a type 1 penicillin allergy (p. 89). Cross-reactivity between penicillin and carbapenems is rare (approximately 1% exceeding the MIC for a certain time (which is different for each class of agent). This is reflected in the dosing interval of benzylpenicillin, which is usually given every 4 hours in severe infection (e.g. meningococcal meningitis), and may be administered by continuous infusion. For other antimicrobial agents, the pharmacodynamic relationships are more complex and often less well understood. With some agents, bacterial inhibition persists after antimicrobial exposure (post-antibiotic and post-antibiotic sub-MIC effects).

Therapeutic drug monitoring
Therapeutic drug monitoring is used to confirm that levels of antimicrobial agents with a low therapeutic
by skin-prick testing). Although avoidance of carbapenems is recommended in penicillin-allergic patients, these drugs may be administered if there are no suitable alternatives and appropriate resuscitation facilities are available.

Gastrointestinal upset and diarrhoea are common, and a mild reversible hepatitis is recognised with many β-lactams. Leucopenia, thrombocytopenia and coagulation deficiencies, and interstitial nephritis and potentiation of aminoglycoside-mediated renal damage are also recognised (p. 502). Seizures and encephalopathy have been reported, particularly with high doses in the presence of renal insufficiency. Thrombophlebitis occurs in up to 5% of patients receiving parenteral β-lactams.

Drug interactions
Synergy occurs in combination with aminoglycosides. Ampicillin decreases the biological effect of oral contraceptives and the whole class is significantly affected by concurrent administration of probenecid, producing a 2–4-fold increase in the peak serum concentration.

Penicillins
Natural penicillins are primarily effective against Gram-positive organisms (except staphylococci, most of which produce a penicillinase) and anaerobic organisms. Strep. pyogenes has remained sensitive to natural penicillins worldwide. According to the European Antimicrobial Resistance Surveillance Network (EARS-Net), the prevalence of high-level penicillin resistance in Strep. pneumoniae in Europe in 2010 was 2.7%. However, the prevalence in individual countries was as high as 33% (Cyprus).

Penicillinase-resistant penicillins are the mainstay of treatment for infections with Staph. aureus, other than meticillin-resistant strains (MRSA). However, EARS-Net data from 2010 indicate that almost 1:5 (18.5%) Staph. aureus isolates in Europe were MRSA.

Aminopenicillins have the same spectrum of activity as the natural penicillins, with additional Gram-negative cover against Enterobacteriaceae. Amoxicillin has better oral absorption than ampicillin. Unfortunately, resistance to these agents is widespread (54% of E. coli Europe-wide in 2010, range 34–83%), so they are no longer appropriate for first-line use in Gram-negative infections. In many organisms, resistance is due to β-lactamase production, which can be overcome by the addition of β-lactamase inhibitors (clavulanic acid or sulbactam).

Carboxypenicillins (e.g. ticarcillin) and ureidopenicillins (e.g. piperacillin) are particularly active against Gram-negative organisms, especially Pseudomonas spp. which are resistant to the aminopenicillins. Beta-lactamase inhibitors may be added to extend their spectrum of activity (e.g. piperacillin–tazobactam).

Cephalosporins and cephapemycins
Cephalosporins are reliable broad-spectrum agents. Unfortunately, their use is associated with C. difficile infections.

Fig. 6.17 Beta-lactam antibiotics. With the exception of aztreonam (monobactam), the β-lactam antibiotics have bicyclic nuclei. This may explain the absence of cross-reaction to aztreonam in penicillin-allergic patients. However, aztreonam and ceftazidime have identical side-chains, so patients with a specific ceftazidime allergy should not be given aztreonam.
infection (p. 343). With the exception of ceftobiprole, the group has no activity against Enterococcus spp. Only the cephamycins have significant anti-anaerobic activity. All cephalosporins are inactivated by ESBL. Cephalosporins are arranged in ‘generations’ (Box 6.23).

- **First-generation compounds** have excellent activity against Gram-positive organisms and some activity against Gram-negatives.
- **Second-generation drugs** retain Gram-positive activity but have extended Gram-negative activity. Cephamycins (e.g. cefotaxime), included in this group, are active against anaerobic Gram-negative bacilli.
- **Third-generation agents** further improve anti-Gram-negative cover. For some (e.g. ceftriaxone), this is extended to include Pseudomonas spp. Cefotaxime and ceftriaxone have excellent Gram-negative activity and retain good activity against Strep. pneumoniae and β-haemolytic streptococci. Ceftriaxone is administered once daily, and is therefore a suitable agent for outpatient antimicrobial therapy.
- **Fourth-generation agents** have an extremely broad spectrum of activity, including Pseudomonas spp., Staph. aureus and streptococci.
- **‘Next generation’ agents** have a third- or fourth-generation spectrum enhanced to include MRSA.

### Monobactams

Aztreonam is the only available monobactam. It is excellent against Gram-negative, except ESBL-producing, organisms, but no useful activity against Gram-positive organisms or anaerobes. It is a parenteral-only agent and may be used safely in penicillin-allergic patients.

### Carbapenems

These intravenous agents have the broadest antibiotic activity of the β-lactam antibiotics, covering most clinically significant bacteria, including anaerobes.

### Macrolide and lincosamide antibiotics

Macrolides (erythromycin, clarithromycin and azithromycin) and lincosamides (lincomycin and clindamycin) are bacteriostatic agents which have related properties. Both classes bind to the same component of the ribosome, so they are potentially competitive and should not be administered together. Macrolides are used for Gram-positive infections in penicillin-allergic patients and in Mycoplasma and Chlamydia infections. Erythromycin is administered 4 times daily and clarithromycin twice daily. The long intracellular half-life of azithromycin allows single-dose/short-course therapy for genitourinary Chlamydia/Mycoplasma spp. infections. Clarithromycin and azithromycin are also used to treat legionellosis.

#### Pharmacokinetics

**Macrolides**

- Variable bioavailability.
- Short half-life (except azithromycin).
- High protein binding.
- Excellent intracellular accumulation.

**Lincosamides (e.g. clindamycin)**

- Good bioavailability.
- Food has no effect on absorption.
- Limited CSF penetration.

#### Adverse effects

- Gastrointestinal upset, especially in young adults (erythromycin 30%).
- Cholestatic jaundice with erythromycin estolate.
- Prolongation of QT interval on ECG, potential for torsades de pointes.
- Clindamycin predisposes to C. difficile infection.

### Ketolides

The ketolides were developed in response to the emergence of penicillin and macrolide resistance in respiratory pathogens. Cross-resistance with macrolides is uncommon. Telithromycin is administered orally and has useful activity against common bacterial causes of respiratory infection, as well as Mycoplasma, Chlamydia and Legionella spp.

### Aminoglycosides

Aminoglycosides are effective mainly in Gram-negative infections. They act synergistically with β-lactam antibiotics and are particularly useful where β-lactam or quinolone resistance occurs in health care-acquired infections. They cause very little local irritation at injection sites and negligible allergic responses. Oto- and nephrotoxicity must be avoided by monitoring of renal function and drug levels and by use of short treatment regimens. Aminoglycosides are not subject to an inoculum effect (p. 151) and they all exhibit a post-antibiotic effect (p. 153).

#### Pharmacokinetics

- Negligible oral absorption.
- Hydrophilic, so excellent penetration to extracellular fluid in body cavities and serosal fluids.
- Very poor intracellular penetration (except hair cells in cochlea and renal cortical cells).
- Negligible CSF and corneal penetration.
- Peak plasma levels 30 minutes after infusion.
- Monitoring of therapeutic levels required.
Gentamicin dosing

- Except in certain forms of endocarditis, pregnancy, severe burns, end-stage renal disease and paediatric patients, gentamicin is administered at 7 mg/kg body weight. The appropriate dose interval depends on drug clearance, and is determined by reference to the Hartford nomogram (Fig. 6.18).
- In streptococcal and enterococcal endocarditis, gentamicin is used with a cell wall active agent (usually a β-lactam), to provide synergy. The usual dose is 1 mg/kg/day 3 times daily for enterococcal endocarditis and 3 mg/kg once a day for most strains of viridans streptococci. Target pre- and post-dose levels are <1 mg/L and 3–5 mg/L respectively when gentamicin is dosed 3 times daily.
- When not used once daily or for endocarditis, gentamicin is administered twice or 3 times daily at 3–5 mg/kg/day. Target pre- and post-dose levels are <1 mg/L and 3–5 mg/L respectively when gentamicin is dosed 3 times daily.
- For other aminoglycosides, consult local guidance.

Adverse effects

- Renal toxicity (usually reversible) accentuated by other nephrotoxic agents.
- Cochlear toxicity (permanent) more likely in older people and those with a predisposing mitochondrial gene mutation.
- Neuromuscular blockade after rapid intravenous infusion (potentiated by calcium channel blockers, myasthenia gravis and hypomagnesaemia).

Quinolones and fluoroquinolones

These are effective and generally well-tolerated bactericidal agents. The quinolones have purely anti-Gram-negative activity, whereas the fluoroquinolones are broad-spectrum agents (Box 6.24). Ciprofloxacin has anti-pseudomonal activity but resistance emerges rapidly. In 2010, 21% of E. coli isolates were resistant to fluoroquinolones in Europe.

Pharmacokinetics

- Well absorbed after oral administration but delayed by food, antacids, ferrous sulphate and multivitamins.
- Wide volume of distribution; tissue concentrations twice those in serum.
- Good intracellular penetration, concentrating in phagocytes.

Adverse effects

- Gastrointestinal side-effects in 1–5%.
- Rare skin reactions (phototoxicity).
- Achilles tendon rupture is reported, especially in older people.
- CNS effects (confusion, tremor, dizziness and occasional seizures in 5–12%), especially in older people.
- Reduces clearance of xanthines and theophyllines, potentially inducing insomnia and increased seizure potential.
- Reports of prolongation of QT interval on ECG with newer fluoroquinolones.
- Cases of hypo- or hyperglycaemia in association with gatifloxacin, so glucose monitoring is needed in patients with diabetes or those with severe hepatic dysfunction.
- Ciprofloxacin use is associated with the acquisition of MRSA and emergence of C. difficile ribotype 027 (p. 343).

Glycopeptides

Glycopeptides (vancomycin and teicoplanin) are effective against Gram-positive organisms only, and are used against MRSA and ampicillin-resistant enterococci. Some staphylococci and enterococci demonstrate intermediate sensitivity or resistance. Vancomycin use
should be restricted to limit emergence of resistant strains. Teicoplanin is not available in all countries. Neither drug is absorbed after oral administration, but vancomycin is used orally to treat *C. difficile* infection.

**Pharmacokinetics**

**Vancomycin**
- Administered by slow intravenous infusion, good tissue distribution and short half-life.
- Enters the CSF only in the presence of inflammation.
- Therapeutic monitoring of intravenous vancomycin is recommended, to maintain pre-dose levels of $> 10 \text{ mg/L}$ (15–20 mg/L in serious staphylococcal infections).

**Teicoplanin**
- Long half-life allows once-daily dosing.

**Adverse effects**
- Histamine release due to rapid vancomycin infusion produces a ‘red man’ reaction (rare with modern preparations).
- Nephrotoxicity is rare, but may occur with concomitant aminoglycoside use, as may ototoxicity.
- Teicoplanin can cause rash, bronchospasm, eosinophilia and anaphylaxis.

**Folate antagonists**

These bacteriostatic antibiotics interfere with the bacterial synthesis of folic acid from para-aminobenzoic acid. A combination of a sulphonamide and either trimethoprim or pyrimethamine is most commonly used, which interferes with two consecutive steps in the metabolic pathway. Available combinations include trimethoprim/sulfamethoxazole (co-trimoxazole) and pyrimethamine with either sulfadoxine (used to treat malaria) or sulfadiazine (used in toxoplasmosis). Co-trimoxazole in high dosage (120 mg/kg daily in 2–4 divided doses) is the first-line drug for *Pneumocystis jirovecii* (carinii) infection. The clinical use of these agents is limited by adverse effects. Folinic acid should be given if they are used long-term or unavoidably in early pregnancy.

**Pharmacokinetics**
- Well absorbed orally.
- Sulphonamides are hydrophilic, distributing well to the extracellular fluid.
- Trimethoprim is lipophilic with high tissue concentrations.

**Adverse effects**
- Trimethoprim is generally well tolerated, with few adverse effects.
- Sulphonamides and dapsone may cause haemolysis in glucose-6-phosphate dehydrogenase deficiency (p. 1029).
- Sulphonamides and dapsone cause skin and mucocutaneous reactions, including Stevens–Johnson syndrome and ‘dapsone syndrome’ (rash, fever and lymphadenopathy).
- Dapsone causes methaemoglobinemia and peripheral neuropathy.

**Tetracyclines and glycylcyclines**

**Tetracyclines**

Of this mainly bacteriostatic class, the newer drugs doxycycline and minocycline show better absorption and distribution than older ones. Most streptococci and Gram-negative bacteria are now resistant, in part due to use in animals (which is banned in Europe). Tetracyclines are indicated for *Mycoplasma* spp., *Chlamydia* spp., *Rickettsia* spp., *Coxiella* spp., *Bartonella* spp., *Helicobacter pylori*, *Treponema pallidum* and atypical mycobacterial infections. Minocycline is occasionally used in chronic staphylococcal infections.

**Pharmacokinetics**
- Best oral absorption is in the fasting state (doxycycline is 100% absorbed unless gastric pH rises).

**Adverse effects**
- All tetracyclines except doxycycline are contraindicated in renal failure.
- Dizziness with minocycline.
- Binding to metallic ions in bones and teeth causes discoloration (avoid in children and pregnancy) and enamel hypoplasia.
- Phototoxic skin reactions.

**Glycylcyclines (tigecycline)**

Chemical modification of tetracycline has produced tigecycline, a broad-spectrum, parenteral-only antibiotic with activity against resistant Gram-positive and Gram-negative pathogens, such as MRSA and ESBL (but excluding *Pseudomonas* spp.). Re-analysis of trial data has shown that there was excess mortality following tigecycline treatment compared with comparator antibiotics, so tigecycline should be used only when there are no available alternative agents.

**Nitroimidazoles**

Nitroimidazoles are highly active against strictly anaerobic bacteria, especially *Bacteroides fragilis*, *C. difficile* and other *Clostridium* spp. They also have significant anti-protozoal activity against amoebae and *Giardia lamblia*.

**Pharmacokinetics**
- Almost completely absorbed after oral administration (60% after rectal administration).
- Well distributed, especially to brain and CSF.
- Safe in pregnancy.

**Adverse effects**
- Metallic taste (dose-dependent).
- Severe vomiting if taken with alcohol – ‘Antabuse effect’.
- Peripheral neuropathy with prolonged use.

**Other antibacterial agents**

Anti-tuberculous agents are discussed in detail on page 693.
**Chloramphenicol**

This is a potent and cheap antibiotic, still widely prescribed throughout the world despite its potential toxicity. Its use is increasingly reserved for severe and life-threatening infections where other antibiotics are either unavailable or impractical. It is bacteriostatic to most organisms but apparently bactericidal to *H. influenzae, Strep. pneumoniae* and *Neisseria meningitidis*. It has a very broad spectrum of activity against aerobic and anaerobic organisms, spirochaetes, *Rickettsia, Chlamydia* and *Mycoplasma* spp. It also has quite useful activity against anaerobes, such as *B. fragilis*. It competes with macrolides and lincosamides for ribosomal binding sites, so should not be used in combination with these agents. Significant adverse effects are ‘grey baby’ syndrome in infants (cyanosis and circulatory collapse due to inability to conjugate drug and excrete the active form in urine); reversible dose-dependent bone marrow depression in adults receiving high cumulative doses; and severe aplastic anaemia in 1 in 25000–40000 exposures (unrelated to dose, duration of therapy or route of administration).

**Daptomycin**

Daptomycin is a cyclic lipopeptide with bactericidal activity against Gram-positive organisms (including MRSA and GRE) but not Gram-negatives. It is not absorbed orally, and is used intravenously to treat resistant Gram-positive infections, e.g. soft tissue infections and infective endocarditis, if other options are not available. Treatment can be associated with increased levels of creatine kinase and patients receiving lipid-lowering statins (p. 453) should discontinue these to avoid myopathy.

**Fusidic acid**

This antibiotic, active against Gram-positive bacteria, is available in intravenous, oral or topical formulations. It is lipid-soluble and distributes well to tissues. However, its antibacterial activity is unpredictable. Fusidic acid is used in combination, typically with antistaphylococcal penicillins, or for MRSA with clindamycin or rifampicin. It interacts with coumarin derivatives and oral contraceptives.

**Nitrofurantoin**

This drug has very rapid renal elimination and is active against aerobic Gram-negative and Gram-positive bacteria, including enterococci. It is used only for treatment of urinary tract infection, being generally safe in pregnancy and childhood. However, with prolonged use, it can cause eosinophilic lung infiltrates, fever, pulmonary fibrosis, peripheral neuropathy, hepatitis and haemolytic anaemia.

**Linezolid**

Linezolid is the only currently licensed oxazolidinone antibiotic. It shows excellent oral absorption with good activity against Gram-positive organisms, including MRSA and GRE. It is competitively inhibited by co-administration of chloramphenicol, vancomycin or clindamycin. Common adverse effects include mild gastrointestinal upset and tongue discoloration. Myelodysplasia and peripheral neuropathy can occur with prolonged use. Linezolid has monoamine oxidase inhibitor (MAOI) activity (p. 244), and co-administration with other MAOIs or serotonin re-uptake inhibitors should be avoided, as this may precipitate a serotonin syndrome (neuromuscular effects, autonomic hyperactivity and altered mental status).

**Fidaxomicin**

Fidaxomicin is an inhibitor of RNA synthesis, which was introduced for the treatment of *C. difficile* infection (CDI) in 2012. In non-severe CDI, it appears to be non-inferior to oral vancomycin and is associated with a lower recurrence rate. Its effectiveness has not been assessed in severe CDI.

**Spectinomycin**

Chemically similar to the aminoglycosides and given intramuscularly, spectinomycin was developed to treat strains of *N. gonorrhoeae* resistant to β-lactam antibiotics. Unfortunately, resistance to spectinomycin is very common. Its only indication is the treatment of gonococcal urethritis in pregnancy or in patients allergic to β-lactam antibiotics.

**Streptogramins**

Quinupristin/dalfopristin (supplied as a 30:70% combination) is active against MRSA and GRE (*Enterococcus faecium* but not *E. faecalis*), and its use should be reserved for these organisms. It is available in intravenous formulation only and shows good tissue penetration, but does not cross the blood–brain barrier or the placenta. Significant phlebitis occurs at injection sites and a raised serum creatinine and eosinophilia may occur.

**Antifungal agents**

See Box 6.25.

**Azole antifungals**

The azoles (imidazoles and triazoles) inhibit synthesis of ergosterol, a constituent of the fungal cell membrane. Side-effects vary but include gastrointestinal upset, hepatitis and rash. Azoles are inhibitors of cytochrome p450 enzymes, so tend to increase exposure to cytochrome p450-metabolised drugs (p. 28).

**Imidazoles**

Miconazole, econazole, clotrimazole and ketoconazole are relatively toxic and therefore mainly administered topically. Clotrimazole is used extensively to treat superficial fungal infections. Ketoconazole may be given orally, but causes severe hepatitis in 1:15000 cases and inhibits enzymes involved in steroid hormone biosynthesis. Triazoles are preferred for systemic administration because of their reduced toxicity.

**Triazoles**

Fluconazole is effective against yeasts (*Candida* and *Cryptococcus* spp.). It is well absorbed after oral administration, and has a long half-life (approximately 30 hours) and an excellent safety profile. The drug is highly water-soluble and distributes widely to all body sites and tissues, including CSF.

Itraconazole is lipophilic and distributes extensively, including to toenails and fingernails. CSF penetration
Principles of infectious disease

6.25 Antifungal agents

<table>
<thead>
<tr>
<th>Agent</th>
<th>Usual route(s) of administration</th>
<th>Clinically relevant antifungal spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imidazoles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miconazole</td>
<td>Topical</td>
<td>Candida spp., dermatophytes</td>
</tr>
<tr>
<td>Econazole/Clotrimazole</td>
<td>Topical, oral</td>
<td>Malassezia spp., dermatophytes, agents of eumycetoma</td>
</tr>
<tr>
<td>Ketoconazole</td>
<td>Topical, oral</td>
<td></td>
</tr>
<tr>
<td>Triazoles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluconazole</td>
<td>Oral, IV</td>
<td>Yeasts (Candida and Cryptococcus spp.)</td>
</tr>
<tr>
<td>Itraconazole</td>
<td>Oral, IV</td>
<td>Yeasts, dermatophytes, dimorphic fungi (p. 376), Aspergillus spp.</td>
</tr>
<tr>
<td>Voriconazole</td>
<td>Oral, IV</td>
<td>Yeasts and most filamentous fungi (excluding mucoraceous moulds)</td>
</tr>
<tr>
<td>Posaconazole</td>
<td>Oral</td>
<td>Yeasts and many filamentous fungi (including most mucoraceous moulds)</td>
</tr>
<tr>
<td>Echinocandins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anidulafungin</td>
<td>IV only</td>
<td>Candida spp., Aspergillus spp. (no activity against Cryptococcus spp. or mucoraceous moulds)</td>
</tr>
<tr>
<td>Caspofungin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micafungin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polyenes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphotericin B</td>
<td>IV</td>
<td>Yeasts and most dimorphic and filamentous fungi (including mucoraceous moulds)</td>
</tr>
<tr>
<td>Nystatin</td>
<td>Topical</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flucytosine</td>
<td>Oral, IV</td>
<td>Yeasts</td>
</tr>
<tr>
<td>Griseofulvin</td>
<td>Oral</td>
<td>Dermatophytes</td>
</tr>
<tr>
<td>Terbinafine</td>
<td>Topical, oral</td>
<td>Dermatophytes</td>
</tr>
</tbody>
</table>

is poor. Oral absorption is erratic and formulation-dependent, necessitating therapeutic drug monitoring.

Voriconazole is well absorbed (96% oral bioavailability) and used mainly in aspergillosis (p. 697).

Posaconazole is the broadest-spectrum antifungal azole, and the only one with consistent activity against mucoraceous moulds. It is currently available as an oral agent only.

Echinocandins

The echinocandins inhibit β-1,3-glucan synthesis in the fungal cell wall. They have few significant adverse effects. Caspofungin, anidulafungin and micafungin are used to treat systemic candidosis, and caspofungin is also used in aspergillosis.

Polyenes

Amphotericin B (AmB) deoxycholate causes cell death by binding to ergosterol and damaging the fungal cytoplasmic membrane. Its use in resource-rich countries has been largely supplanted by less toxic agents. It is lipophilic, insoluble in water and not absorbed orally. Its long half-life enables once-daily administration. CSF penetration is poor.

Adverse effects include immediate anaphylaxis, other infusion-related reactions and nephrotoxicity. Nephrotoxicity may be sufficient to require dialysis, and occurs in most patients who are adequately dosed. It may be ameliorated by concomitant infusion of normal saline. Irreversible nephrotoxicity occurs with large cumulative doses of AmB.

Nystatin has a similar spectrum of antifungal activity to AmB. Its toxicity limits it to topical use, e.g. in oral and vaginal candidiasis.

Lipid formulations of amphotericin B

Lipid formulations of AmB have been developed to reduce AmB toxicity. They consist of AmB encapsulated in liposomes (liposomal AmB, L-AmB) or complexed with phospholipids (AmB lipid complex, ABLC). The drug becomes active on dissociating from its lipid component. Adverse effects are similar to, but considerably less frequent than, those with AmB deoxycholate, and efficacy is similar. Lipid formulations of AmB are used in invasive fungal disease, as empirical therapy in patients with neutropenic fever (p. 1004), and also in visceral leishmaniasis (p. 362).

Other antifungal agents

Flucytosine

This drug has particular activity against yeasts. When used as monotherapy, resistance develops rapidly, so it should be administered in combination with another antifungal agent. Oral dosing is effective. Adverse effects include myelosuppression, gastrointestinal upset and hepatitis.

Griseofulvin

Griseofulvin has been largely superseded by terbinafine and itraconazole for treatment of dermatophyte infections, except in children, for whom these agents remain largely unlicensed. It demonstrates excellent oral bioavailability and is deposited in keratin precursor cells, which become resistant to fungal invasion. The duration of treatment is 2–4 weeks for tinea corporis/capitis, 4–8 weeks for tinea pedis, and 4–6 months for onychomycosis (fungal nail infections).

Terbinafine

Terbinafine is well absorbed orally, can be given once daily and distributes with high concentration to sebum and skin, with a half-life of more than 1 week. It is used topically for dermatophyte skin infections and orally for onychomycosis. The major adverse reaction is hepatic toxicity (approximately 1:50000 cases). Terbinafine is not recommended for breastfeeding mothers.

Antiviral agents

Most viral infections in immunocompetent individuals resolve without intervention. Antiviral therapy is
Anti-herpesvirus agents

Aciclovir, valaciclovir, penciclovir and famciclovir are acyclic analogues of guanosine, which inhibit viral DNA polymerase after being phosphorylated by available for a limited number of infections only (Box 6.26 and p. 407).

Antiretroviral agents

These agents, used predominantly against HIV, are discussed in Chapter 14.
Principles of Infectious Disease

Anti-influenza agents

**Zanamivir and oseltamivir**

These agents inhibit influenza A and B neuraminidase, which is required for release of virus from infected cells (see Fig. 6.2, p. 135). They are used in treatment and prophylaxis of influenza. Administration within 48 hours of disease onset reduces the duration of symptoms by approximately 1–1½ days. In the UK, their use is limited mainly to adults with chronic respiratory or renal disease, significant cardiovascular disease, immunosuppression or diabetes mellitus, during known outbreaks. Peramivir has a distinct chemical structure, which means that it retains activity against some oseltamivir and zanamivir-resistant strains. It has poor oral bioavailability and is being developed as an intravenous or intramuscular formulation. An intravenous formulation of zanamivir is in development for critically ill patients.

**Amantadine and rimantadine**

These drugs reduce replication of influenza A by inhibition of viral M2 protein ion channel function, which is required for uncoating (see Fig. 6.2, p. 135). Resistance develops rapidly and is widespread, and amantadine and rimantadine should be used only if the prevalence of resistance locally is known to be low. They are no longer recommended for treatment or prophylaxis in the UK or USA, having been superseded by zanamivir and oseltamivir. However, they may still be indicated to treat oseltamivir-resistant influenza A in patients unable to take zanamivir (e.g. ventilated patients).

Agents used against hepatitis viruses

**Ribavirin**

Ribavirin is a guanosine analogue that inhibits nucleic acid synthesis in a variety of viruses and is used in particular in the treatment of hepatitis C virus.

**Lamivudine, adefovir dipivoxil, tenofovir, entecavir and telbivudine**

These agents have excellent activity against hepatitis B virus DNA polymerase-reverse transcriptase. They are well tolerated after oral administration but resistance develops with monotherapy. Resistance seems to emerge most rapidly for lamivudine (via the tyrosine–methionine–aspartate–aspartate, or YMDD, mutation) and most slowly for entecavir (multiple mutations required). Organisms resistant to lamivudine are usually also resistant to telbivudine, but not to adefovir/tenofovir. The role of monotherapy for hepatitis B virus is currently a matter for debate, and combination therapy, as used in HIV treatment, is likely to be increasingly employed. Lamivudine and tenofovir are also used against HIV (p. 407).

**Telaprevir and boceprevir**

A number of antiviral inhibitors of the hepatitis C virus NS3 serine protease or NS5B polymerase are in development. Telaprevir and boceprevir have been licensed for use in chronic hepatitis C virus genotype 1 disease. Addition of these agents to standard interferon–ribavirin combination therapy improves sustained virological response rates. They are prone to drug–drug interactions, including those involving antiretrovirals. Resistance develops to these agents, so they are administered as part of combination treatment.

Interferon-α

The interferons are naturally occurring cytokines that are produced as an early response to viral infection (p. 74). The addition of a polyethylene glycol (PEG) moiety to the molecule significantly enhances pharmacokinetics and efficacy.

Antiparasitic agents

**Drugs used against helminths**

**Benzimidazoles (albendazole, mebendazole)**

These agents act by inhibiting both helminth glucose uptake, causing depletion of glycogen stores, and fumarate reductase. Albendazole is used for hookworm, ascariasis, threadworm, Strongyloides infection, trichinellosis, Taenia solium (cysticercosis) and hydatid disease. Mebendazole is used for hookworm, ascariasis, threadworm and whipworm. The drugs are administered orally. Absorption is relatively poor, but increased by a fatty meal. Significant adverse effects are uncommon.

**Bithionol**

Bithionol is used to treat fluke infections with Fasciola hepatica. It is well absorbed orally. Adverse effects are mild (e.g. nausea, vomiting, diarrhoea, rashes) but relatively common (approximately 30%).
**Diethylcarbamazine**
Diethylcarbamazine (DEC) is an oral agent used to treat filariasis and loiasis. Treatment of filariasis is often followed by fever, headache, nausea, vomiting, arthralgia and prostration. This is caused by the host response to dying microfilariae, rather than the drug, and may be reduced by pre-treatment with corticosteroids.

**Ivermectin**
Ivermectin binds to helminth nerve and muscle cell ion channels, causing increased membrane permeability. It is an oral agent, used in Strongyloides infection, filariasis and onchocerciasis. Significant side-effects are uncommon.

**Niclosamide**
Niclosamide inhibits oxidative phosphorylation, causing paralysis of helminths. It is an oral agent, used in Taenia saginata and intestinal T. solium infection. Systemic absorption is minimal and it has few significant side-effects.

**Piperazine**
Piperazine inhibits neurotransmitter function, causing helminth muscle paralysis. It is an oral agent, used in ascariasis and threadworm (Enterobius vermicularis) infection. Significant adverse effects are uncommon, but include neuropsychological reactions such as vertigo, confusion and convulsions.

**Praziquantel**
Praziquantel increases membrane permeability to Ca++, causing violent contraction of worm muscle. It is the drug of choice for schistosomiasis, and is also used in T. saginata and intestinal T. solium infection. Systemic absorption is minimal and it has few significant side-effects.

**Thiabendazole**
Thiabendazole inhibits fumarate reductase, which is required for energy production in helminths. It is used orally in ascariasis and threadworm infection. Systemic absorption is poor and adverse effects are uncommon.

**Antimalarial agents**

**Artemisinin (qinghaosu) derivatives**
Artemisinin originates from a herb (sweet wormwood, Artemisia annua), which was used in Chinese medicine to treat fever. Its derivatives, artemether and artesunate, were developed for use in malaria in the 1970s. Their mechanism of action is unknown. They are used in the treatment, but not prophylaxis, of malaria, usually in combination with other antimalarials, and are effective against strains of Plasmodium spp. that are resistant to other antimalarials. Artemether is lipid-soluble and may be administered via intramuscular and oral routes. Artesunate is water-soluble and is administered intravenously or orally. Serious adverse effects are uncommon. Current advice for malaria in pregnancy is that the artemisinin derivatives should be used to treat uncomplicated falciparum malaria in the second and third trimesters, but should not be prescribed in the first trimester until more information becomes available.

**Atovaquone**
Atovaquone inhibits mitochondrial function. It is an oral agent, used for treatment and prophylaxis of malaria, in combination with proguanil (see below), without which it is ineffective. It is also employed in the treatment of mild cases of Pneumocystis jiroveci (carinii) pneumonia, where there is intolerance to co-trimoxazole. Significant adverse effects are uncommon.

**Folate synthesis inhibitors (proguanil, pyrimethamine–sulfadoxine)**
Proguanil inhibits dihydrofolate reductase and is used for malaria prophylaxis. Pyrimethamine–sulfadoxine is used in the treatment of malaria (p. 356).

**Quinoline-containing compounds**
Chloroquine and quinine are believed to act by intraparasitic inhibition of haem polymerisation, resulting in toxic build-up of intracellular haem. The mechanisms of action of other agents in this group (quinidine, amodiaquine, mefloquine, primaquine, etc.) may differ. They are employed in the treatment and prophylaxis of malaria. Primaquine is used for radical cure of malaria due to Plasmodium vivax and P. ovale (destruction of liver hypnozoites). Chloroquine is also given for extra-intestinal amoebiasis.

Chloroquine can cause a pruritus sufficient to compromise compliance with therapy. If used in long-term, high-dose regimens, it causes an irreversible retinopathy. Overdosage leads to life-threatening cardiotoxicity. The side-effect profile of mefloquine includes neuropsychiatric effects ranging from mood change, nightmares and agitation to hallucinations and psychosis. Quinidine may cause hypoglycaemia and cardiotoxicity, especially when administered parenterally. Primaquine causes haemolysis in people with glucose-6-phosphate dehydrogenase deficiency (p. 1029), which should be excluded before therapy. Chloroquine is considered safe in pregnancy, but mefloquine should be avoided in the first trimester.

**Lumefantrine**
Lumefantrine is used in combination with artemether to treat uncomplicated falciparum malaria, including chloroquine-resistant strains. Its mechanism of action is unknown. Significant adverse effects are uncommon.

**Drugs used in trypanosomiasis**

**Benznidazole**
Benznidazole is an oral agent used to treat South American trypanosomiasis (Chagas’ disease, p. 360). Significant and common adverse effects include dose-related peripheral neuropathy, purpuric rash and granulocytopenia.
**Principles of infectious disease**

**Eflornithine**
Eflornithine inhibits biosynthesis of polyamines by ornithine decarboxylase inhibition, and is used in West African trypanosomiasis (T. brucei gambiense infection) of the CNS. It is administered as an intravenous infusion 4 times daily, which may be logistically difficult in the geographic areas affected by this disease. Significant adverse effects are common, and include convulsions, gastrointestinal upset and bone marrow depression. Eflornithine is also used (topically) to treat hirsutism (p. 763).

**Melarsoprol**
This is an arsenical agent, used to treat CNS infections in East and West African trypanosomiasis (T. brucei rhodesiense and gambiense). It is administered intravenously. Melarsoprol treatment is associated with peripheral neuropathy and reactive arsenical encephalopathy (RAE), which carries a significant mortality.

**Nifurtimox**
Nifurtimox is administered orally to treat South American trypanosomiasis (Chagas’ disease). Gastrointestinal and neurological adverse effects are common.

**Pentamidine isetionate**
Pentamidine is an inhibitor of DNA replication used in West African trypanosomiasis (T. brucei gambiense) and, to a lesser extent, in visceral and cutaneous leishmaniasis. It is also prescribed in Pneumocystis jirovecii (carinii) pneumonia. It is administered via intravenous or intramuscular routes. It is a relatively toxic drug, commonly causing rash, renal impairment, profound hypotension (especially on rapid infusion), electrolyte disturbances, blood dyscrasias and hypoglycaemia.

**Suramin**
Suramin is a naphthaline dye derivative, used to treat East African trypanosomiasis (T. brucei rhodesiense). It is administered intravenously. Adverse effects are common, and include rash, gastrointestinal disturbance, blood dyscrasias, peripheral neuropathies and renal impairment.

**Other antiprotozoal agents**

**Pentavalent antimonials**
Sodium stibogluconate and meglumine antimoniate inhibit protozoal glycolysis by phosphofructokinase inhibition. They are used parenterally (intravenous or intramuscular) to treat leishmaniasis. Adverse effects include arthralgia, myalgias, raised hepatic transaminases, pancreatitis and ECG changes. Severe cardiotoxicity leading to death is not uncommon.

**Diloxanide furoate**
This oral agent is used to eliminate luminal cysts following treatment of intestinal amoebiasis, or in asymptomatic cyst excreters. The drug is absorbed slowly (enabling luminal persistence) and has no effect in hepatic amoebiasis. It is a relatively non-toxic drug, the most significant adverse effect being flatulence.

**Iodoquinol (di-iodohydroxyquinoline)**
Iodoquinol is a quinoline derivative (p. 163) with activity against Entamoeba histolytica cysts and trophozoites. It is used orally to treat asymptomatic cyst excreters or, in association with another amoebicide (e.g. metronidazole), to treat extra-intestinal amoebiasis. Long-term use of this drug is not recommended, as neurological adverse effects include optic neuritis and peripheral neuropathy.

**Nitazoxanide**
Nitazoxanide is an inhibitor of pyruvate–ferredoxin oxidoreductase-dependent anaerobic energy metabolism in protozoa. It is a broad-spectrum agent, active against various nematodes, tapeworms, flukes and intestinal protozoa. Nitazoxanide also has activity against some anaerobic bacteria and viruses. It is administered orally in giardiasis and cryptosporidiosis. Adverse effects are usually mild and involve the gastrointestinal tract (e.g. nausea, diarrhoea and abdominal pain).

**Further information and acknowledgements**

**Websites**
www.cdc.gov Centers for Disease Control, Atlanta, USA. Provides information on all aspects of communicable disease, including prophylaxis against malaria.
www.dh.gov.uk UK Department of Health. The publications section provides current UK recommendations for immunisation.
www.hpa.org.uk Health Protection Agency. Provides information on infectious diseases relating mainly to the UK, including community infection control.
www.who.int World Health Organization. Provides up-to-date information on global aspects of infectious disease, including outbreak updates.

**Figure acknowledgements**
Fig. 6.10 Adapted from Samaranayake L. Essential microbiology for dentistry. 3rd edn. Edinburgh: Churchill Livingstone; 2006 (Fig. 1.1); copyright Elsevier.
Fig. 6.12 Based on the ‘How to Handwash’ URL: http://www.who.int/gpsc/5may/How_To_Handwash_Poster.pdf ©World Health Organization 2009. All rights reserved.
In the image, there is a diagram illustrating a comprehensive geriatric assessment. The assessment includes various components such as cognitive function, vision, hearing, pulse, hydration, nutrition, muscle, per rectum, skin, joints, gait and balance, full systems examination, with particular attention to the above.

Key points:
- **Cognitive function**: Mini-mental state examination (see Ch. 10).
- **Vision**: Visual acuity, glasses worn/present, cataract.
- **Hearing**: Wax, hearing aid used.
- **Erect and supine blood pressure**: Postural hypotension.
- **Pulse**: Atrial fibrillation.
- **Hydration**: Skin turgor, oedema.
- **Nutrition**: Body mass index (height calculated from arm demispan or knee height to compensate for loss of vertebral height), recent weight loss, e.g., loose skin folds, dentition/oral hygiene.
- **Muscle**: Wasting, strength.
- **Per rectum**: Faecal impaction, prostate size/consistency in men, anal tone.
- **Skin**: Wounds/ulcers, infection, swelling.
- **Joints**: Deformity, pain, swelling, range of movement.
- **Gait and balance**: Get up and go test (see opposite), walking aid used.

Insets (Wasted hand, kyphosis) From Atzal Mir 2003; (Senile purpura) Forbes and Jackson 2004; (Venous ulceration) Mosti 2012 – see p. 177.
**History**

- Slow down the pace.
- Ensure the patient can hear.
- Establish the speed of onset of the illness.
- If the presentation is vague, carry out a systematic enquiry.
- Obtain full details of:
  - all drugs, especially any recent prescription changes
  - past medical history, even from many years previously
  - usual function
  1. Can the patient walk normally?
  2. Has the patient noticed memory problems?
  3. Can the patient perform all household tasks?
- Obtain a collateral history: confirm information with a relative or carer and the general practitioner, particularly if the patient is confused or communication is limited by deafness or speech disturbance.

**Social assessment**

**Home circumstances**

- Living alone, with another or in a care home.

**Activities of daily living (ADL)**

- Tasks for which help is needed:
  - domestic ADL: shopping, cooking, housework
  - personal ADL: bathing, dressing, walking.
- Informal help: relatives, friends, neighbours.
- Formal social services: home help, meals on wheels.
- Carer stress.

**Examination**

- Thorough to identify all comorbidities.
- Tailored to the patient’s stamina and ability to cooperate.
- Include functional status: cognitive function, gait and balance, nutrition, hearing and vision.

**Get up and go test**

To assess gait and balance, ask the patient to stand up from a sitting position, walk 10 m, turn and go back to the chair. A normal performance takes less than 12 seconds.
Sweeping demographic change has meant that older people now represent the core practice of medicine in many countries. A good knowledge of the effects of ageing and the clinical problems associated with old age is thus essential in most medical specialties. The older population is extremely diverse; a substantial proportion of 90-year-olds enjoy an active healthy life, while some 70-year-olds are severely disabled by chronic disease. The terms ‘chronological’ and ‘biological’ ageing have been coined to describe this phenomenon. Biological rather than chronological age is taken into consideration when making clinical decisions about, for example, the extent of investigation and intervention that is appropriate.

Geriatric medicine is concerned particularly with frail older people, in whom physiological capacity is so reduced that they are incapacitated by even minor illness. They frequently have multiple comorbidities, and acute illness may present in non-specific ways, such as confusion, falls or loss of mobility and day-to-day functioning. These patients are prone to adverse drug reactions, partly because of polypharmacy and partly because of age-related changes in responses to drugs and their elimination (p. 36). Disability is common, but patients’ function can often be improved by the interventions of the multidisciplinary team (p. 167).

Older people have been neglected in research terms and, until recently, were rarely included in randomised controlled clinical trials. There is thus little evidence on which to base practice.

### Demography

The demography of developed countries has changed rapidly in recent decades. In the UK, for example, the total population grew by 11% over the last 30 years, but the number of people aged over 65 years rose by 24%. The steepest rise occurred in those aged over 85 – from 600 000 in 1981 to 1.5 million in 2011 – and this number is projected to increase to 2.4 million by 2026, whilst the working-age population (20–64 years) is expected to grow by only 4% between 2011 and 2026. This will have a significant impact on the old-age dependency ratio, i.e. the number of people of working age for each person over retirement age. Young people support older members of the population directly (e.g. through living arrangements) and financially (e.g. through taxation and pension contributions), so the consequences of a reduced ratio are far-reaching. However, many older people support the younger population, through care of children and other older people.

Life expectancy in the developed world is now prolonged, even in old age (Box 7.1); women aged 80 years can expect to live for a further 10 years. However, rates of disability and chronic illness rise sharply with ageing and have a major impact on health and social services. In the UK, the reported prevalence of a chronic illness or disability sufficient to restrict daily activities is around 25% in those aged 50–64, but is 66% in men and 75% in women aged over 85.

Although the proportion of the population aged over 65 years is greater in developed countries, two-thirds of the world population of people aged over 65 live in developing countries at present, and this is projected to rise to 75% in 2025. The rate of population ageing is much faster in developing countries (Fig. 7.1) and so they have less time to adjust to its impact.

### Functional Anatomy and Physiology

#### Biology of ageing

Ageing can be defined as a progressive accumulation through life of random molecular defects that build up within tissues and cells. Eventually, despite multiple repair and maintenance mechanisms, these result in age-related functional impairment of tissues and organs.

Many genes probably contribute to ageing, with those that determine durability and maintenance of somatic cell lines particularly important. However, genetic factors only account for around 25% of variance in human lifespan; nutritional and environmental factors determine the rest.

A major contribution to random molecular damage is made by reactive oxygen species produced during the metabolism of oxygen to produce cellular energy. These cause oxidative damage at a number of sites:

- **Nuclear chromosomal DNA**, causing mutations and deletions which ultimately lead to aberrant gene function and potential for malignancy.
- **Telomeres**, which are the protective end regions of chromosomes which shorten with each cell division...
because telomerase (which copies the end of the 3’ strand of linear DNA in germ cells) is absent in somatic cells. When telomeres are sufficiently eroded, cells stop dividing. It has been suggested that telomeres represent a ‘biological clock’ which prevents uncontrolled cell division and cancer. Telomeres are particularly shortened in patients with premature ageing due to Werner’s syndrome, in which DNA is damaged due to lack of a helicase.

- **Mitochondrial DNA and lipid peroxidation**, resulting in reduced cellular energy production and ultimately cell death.
- **Proteins** – e.g. those increasing formation of advanced glycosylation end-products from spontaneous reactions between proteins and sugars. These damage structure and function of the affected protein, which becomes resistant to breakdown.

The rate at which damage occurs is malleable and this is where the interplay with environment, particularly nutrition, takes place. There is evidence in some organisms that this interplay is mediated by insulin signalling pathways. Chronic inflammation also plays an important role, again in part by driving the production of reactive oxygen species.

### Physiological changes of ageing

The physiological features of normal ageing have been identified by examining disease-free populations of older people, to separate the effects of pathology from those due to time alone. However, the fraction of older people who age without disease ultimately declines to very low levels, so that use of the term ‘normal’ becomes debatable. There is a marked increase in inter-individual variation in function with ageing; many physiological processes deteriorate substantially when measured across populations, but some individuals show little or no change. This heterogeneity is a hallmark of ageing, meaning that each person must be assessed individually and that one cannot unthinkingly apply the same management to all people of a certain age.

Although some genetic influences contribute to heterogeneity, environmental factors, such as poverty, nutrition, exercise, cigarette smoking and alcohol misuse, play a large part, and a healthy lifestyle should be encouraged even when old age has been reached.

The effects of ageing are usually not enough to interfere with organ function under normal conditions, but reserve capacity is significantly reduced. Some changes

### Clinical consequences

| CNS | • Increased risk of delirium  
| • Presbyacusis/high-tone hearing loss  
| • Presbyopia/abnormal near vision  
| • Cataract  
| • Muscle weakness and wasting  
| • Reduced position and vibration sense  
| • Increased risk of falls |
| Respiratory system | • Reduced vital capacity and peak expiratory flow  
| • Increased residual volume  
| • Reduced inspiratory reserve volume  
| • Reduced arterial oxygen saturation  
| • Increased risk of infection |
| Cardiovascular system | • Reduced exercise tolerance  
| • Widened aortic arch on X-ray  
| • Widened pulse pressure  
| • Increased risk of postural hypotension  
| • Increased risk of atrial fibrillation |
| Endocrine system | • Increased risk of impaired glucose tolerance |
| Renal system | • Impaired fluid balance  
| • Increased risk of dehydration/overload  
| • Impaired drug metabolism and excretion |
| Gastrointestinal system | • Constipation |
| Bones | • Increased risk of osteoporosis |

---

Fig. 7.2 Features and consequences of normal ageing.
of ageing, such as depigmentation of the hair, are of no clinical significance. Figure 7.2 shows many factors that are clinically important.

Frailty

Frailty is defined as the loss of an individual’s ability to withstand minor stresses because the reserves in function of several organ systems are so severely reduced that even a trivial illness or adverse drug reaction may result in organ failure and death. The same stresses would cause little upset in a fit person of the same age.

It is important to understand the difference between ‘disability’, ‘comorbidity’ and ‘frailty’. Disability indicates established loss of function (e.g. mobility; see Box 7.13, p. 176), while frailty indicates increased vulnerability to loss of function. Disability may arise from a single pathological event (such as a stroke) in an otherwise healthy individual. After recovery, function is largely stable and the patient may otherwise be in good health. When frailty and disability coexist, function deteriorates markedly even with minor illness, to the extent that the patient can no longer manage independently. Similarly, comorbidity (the number of diagnoses present) is not equivalent to frailty; it is quite possible to have several diagnoses without major impact on homeostatic reserve.

Unfortunately, the term ‘frail’ is often used rather vaguely, sometimes to justify a lack of adequate investigation and intervention in older people. However, it can be specifically identified by assessing function in a number of domains. Two main approaches to evaluating frailty exist: measurement of physiological function across a number of domains (e.g. the Fried Frailty score, Box 7.2), or a score based on the number of deficits or problems – for example, the Rockwood score.

Frail older people particularly benefit from a clinical approach that addresses both the precipitating acute illness and their underlying loss of reserves. It may be possible to prevent further loss of function through early intervention; for example, a frail woman with myocardial infarction will benefit from specific cardiac investigation and drug treatment, but may benefit even further from an exercise programme to improve musculoskeletal function, balance and aerobic capacity, with nutritional support to restore lost weight. Establishing a patient’s level of frailty also helps inform decisions regarding further investigation and management, and the need for rehabilitation.

7.2 How to assess a Fried Frailty score

- Handgrip strength in bottom 20% of healthy elderly distribution
- Walking speed in bottom 20% of healthy elderly distribution
- Self-reported exhaustion
- Physically inactive
- At least 6 kg weight loss within 1 year

Patient is defined as frail if 3 or more factors are present.

*Varies between populations. Grip cutoff is 30 kg for men, 18 kg for women in US adults; 5 m walk time cutoff is 7 seconds in US adults for both sexes.

INVESTIGATIONS

Comprehensive geriatric assessment

Although not strictly an investigation, one of the most powerful tools in the management of older people is the Comprehensive Geriatric Assessment, which identifies all the relevant factors contributing to their presentation (p. 166). In frail patients with multiple pathology, it may be necessary to perform the assessment in stages to allow for their reduced stamina. The outcome should be a management plan that not only addresses the acute presenting problems, but also improves the patient’s overall health and function (Box 7.3).

Comprehensive Geriatric Assessment is performed by a multidisciplinary team (p. 167). Such an approach was pioneered by Dr Marjory Warren at the West Middlesex Hospital in London in the 1930s; her comprehensive assessment and rehabilitation of supposedly incurable, long-term bedridden older people revolutionised the approach of the medical profession to older, frail people and laid the foundations for the modern specialty of geriatric medicine.

EBM 7.3 Comprehensive geriatric assessment

‘Inpatient comprehensive geriatric assessment reduces short-term mortality and increases the chance of patients living at home in the long term.’


Decisions about investigation

Accurate diagnosis is important at all ages but frail older people may not be able to tolerate lengthy or invasive procedures, and diagnoses may be revealed for which patients could not withstand intensive or aggressive treatment. On the other hand, disability should never be dismissed as due to age alone. For example, it would be a mistake to supply a patient no longer able to climb stairs with a stair lift, when simple tests would have revealed osteoarthritis of a hip and vitamin D deficiency, for which appropriate treatment would have restored his or her strength. So how do doctors decide when and how far to investigate?

The patient’s general health

Does this patient have the physical and mental capacity to tolerate the proposed investigation? Does he have the aerobic capacity to undergo bronchoscopy? Will confusion prevent her from remaining still in the magnetic resonance imaging (MRI) scanner? The more comorbidities a patient has, the less likely he or she will be able to withstand an invasive intervention.

Will the investigation alter management?

Would the patient be fit for, or benefit from, the treatment that would be indicated if investigation proved positive? The presence of comorbidity is more important than age itself in determining this. When a patient with
severe heart failure and a previous disabling stroke presents with a suspicious mass lesion on chest X-ray, detailed investigation and staging may not be appropriate if he is not fit for surgery, radical radiotherapy or chemotherapy. On the other hand, if the same patient presented with dysphagia, investigation of the cause would be important, as he would be able to tolerate endoscopic treatment (for example, to palliate an obstructing oesophageal carcinoma).

The views of the patient and family
Older people may have strong views about the extent of investigation and the treatment they wish to receive, and these should be sought from the outset. If the patient wishes, the views of relatives can be taken into account. If the patient is not able to express a view or lacks the capacity to make decisions because of cognitive impairment or communication difficulties, then relatives’ input becomes particularly helpful. They may be able to give information on views previously expressed by the patient or on what the patient would have wanted under the current circumstances. However, families should never be made to feel responsible for difficult decisions.

Advance directives
Advance directives or ‘living wills’ are statements made by adults at a time when they have the capacity to decide about the interventions they would refuse or accept in the future, should they no longer be able to make decisions or communicate them. An advance directive cannot authorise a doctor to do anything that is illegal and doctors are not bound to provide a specific treatment requested if, in their professional opinion, it is not clinically appropriate. However, any advance refusal of treatment, made when the patient was able to make decisions based on adequate information about their implications, is legally binding in the UK. It must be respected when it clearly applies to the patient’s present circumstances and when there is no reason to believe that the patient has changed his or her mind.

PRESENTING PROBLEMS IN GERIATRIC MEDICINE

Characteristics of presenting problems in old age
Problem-based practice is central to geriatric medicine. Most problems are multifactorial and there is rarely a single unifying diagnosis. All contributing factors have to be taken into account and attention to detail is paramount. Two patients who share the same presenting problem may have completely disparate diagnoses. A wide knowledge of adult medicine is required, as disease in any, and often many, of the organ systems has to be managed at the same time. There are a number of features that are particular to older patients.

Late presentation
Many people (of all ages) accept ill health as a consequence of ageing and may tolerate symptoms for lengthy periods before seeking medical advice. Comorbidities may also contribute to late presentation; in a patient whose mobility is limited by stroke, angina may only present when coronary artery disease is advanced, as the patient has been unable to exercise sufficiently to cause symptoms at an earlier stage.

Atypical presentation
Infection may present with delirium and without clinical pointers to the organ system affected. Stroke may present with falls rather than symptoms of focal weakness. Myocardial infarction may present as weakness and fatigue, without the chest pain or dyspnoea. The reasons for these atypical presentations are not always easy to establish. Perception of pain is altered in old age, which may explain why myocardial infarction presents in other ways. The pyretic response is blunted in old age so that infection may not be obvious at first. Cognitive impairment may limit the patient’s ability to give a history of classical symptoms.

Acute illness and changes in function
Atypical presentations in frail elderly patients include ‘failure to cope’, ‘found on floor’, ‘confusion’ and ‘off feet’, but these are not diagnoses. The possibility that an acute illness has been the precipitant must always be considered. To establish whether the patient’s current status is a change from his or her usual level of function, it helps to ask a relative or carer (by phone if necessary). Investigations aimed at uncovering an acute illness will not be fruitful in a patient whose function has been deteriorating over several months, but are important if function has suddenly changed.

Multiple pathology
Presentations in older patients have a more diverse differential diagnosis because multiple pathology is so common. There are frequently a number of causes for any single problem, and adverse effects from medication often contribute. A patient may fall because of osteoarthritis of the knees, postural hypotension due to diuretic therapy for hypertension, and poor vision due to cataracts. All these factors have to be addressed to prevent further falls, and this principle holds true for most of the common presenting problems in old age.

Approach to presenting problems in old age
For the sake of clarity, the common presenting problems are described individually, but in real life, older patients often present with several at the same time, particularly confusion, incontinence and falls. These share some underlying causes and may precipitate each other.

The approach to most presenting problems in old age can be summarised as follows:
- Obtain a collateral history. Find out the patient’s usual status (e.g. mobility, cognitive state) from a relative or carer. Call these people by phone if they are not present.
- Check all medication. Have there been any recent changes?
- Search for and treat any acute illness. See Box 7.4.
- Identify and reverse predisposing risk factors. These depend on the presenting problem.
Falls

Around 30% of those over 65 years of age fall each year, this figure rising to more than 40% in those aged over 80. Although only 10–15% of falls result in serious injury, they are the cause of more than 90% of hip fractures in this age group, compounded by the rising prevalence of osteoporosis. Falls also lead to loss of confidence and fear, and are frequently the ‘final straw’ that makes an older person decide to move to institutional care. Management will vary according to the underlying cause.

Acute illness

Falls are one of the classical atypical presentations of acute illness in frail people. The reduced reserves in older people’s neurological function mean that they are less able to maintain their balance when challenged by an acute illness. Suspicion should be high when falls have suddenly occurred over a period of a few days. Common underlying illnesses include infection, stroke, metabolic disturbance and heart failure. Thorough examination and investigation are required (see Box 7.4). It is also important to establish whether any drug which precipitates falls, such as a psychotropic or hypotensive agent, has been started recently. Once the underlying acute illness has been treated, falls may stop.

Blackouts

A proportion of older people who ‘fall’ have, in fact, had a syncopal episode. A collateral history from a witness is of utmost importance in anyone falling over; people who lose consciousness do not always remember having done so. If loss of consciousness is suggested by the patient or witness, it is important to perform appropriate investigations (pp. 554 and 1157).

Mechanical and recurrent falls

Amongst patients who have tripped or are uncertain how they fell, those who have fallen more than once in the past year and those who are unsteady during a ‘get up and go’ test (p. 167) require further assessment. Patients with recurrent falls are commonly frail, with multiple medical problems and chronic disabilities. Obviously, such patients may present with a fall resulting from an acute illness or syncope, but they will remain at risk of further falls even when the acute illness has resolved. The risk factors for falls (Box 7.5) should be considered. If problems are identified with muscle strength, balance, vision or cognitive function, the causes of these must be identified by specific investigation, and treatment commenced if appropriate. Careful assessment of the patient’s gait may provide important clues to an underlying diagnosis (Box 7.6). Common pathologies identified include cerebrovascular disease (Ch. 27), Parkinson’s disease (p. 1195) and osteoarthritis of weight-bearing joints (p. 1081). Osteoporosis risk factors should also be sought and dual energy X-ray absorptiometry (DEXA) bone density scanning considered in all older patients who have recurrent falls, particularly if they have already sustained a fracture (p. 1065).

Prevention of falls and fractures

Falls can be prevented by multiple risk factor intervention (Box 7.7). The most effective intervention is balance and strength training by physiotherapists; an alternative with good evidence is tai chi training. An assessment of the patient’s home environment for hazards should be undertaken by an occupational therapist, who may also provide personal alarms so that patients can summon help, should they fall again. Rationalising psychotropic medication may help to reduce sedation, although many older patients are reluctant to stop hypnotics. If postural hypotension is present (defined as a drop in blood pressure of > 20 mmHg systolic or > 10 mmHg diastolic pressure on standing from supine), reducing or stopping hypotensive drugs may be helpful. Evidence supporting the efficacy of other interventions for postural hypotension is lacking, but drugs, including fludrocortisone and midodrine, are sometimes used to try to improve dizziness on standing. Simple interventions, such as new glasses to correct visual acuity, and podiatry, can also have a significant impact on function in those who fall.

If osteoporosis is diagnosed, specific drug therapy should be commenced (p. 1122). In patients in institutional care, calcium and vitamin D₃ administration has

### 7.4 Screening investigations for acute illness

- Full blood count
- Urea and electrolytes, liver function tests, calcium and glucose
- Chest X-ray
- Electrocardiogram
- C-reactive protein: useful marker for occult infection or inflammatory disease
- Blood cultures if pyrexial

### 7.5 Risk factors for falls

- Muscle weakness
- History of falls
- Gait or balance abnormality
- Use of a walking aid
- Visual impairment
- Arthritis
- Impaired activities of daily living
- Depression
- Cognitive impairment
- Age over 80 years
- Psychotropic medication

### 7.6 Abnormal gaits and probable causes

<table>
<thead>
<tr>
<th>Gait abnormality</th>
<th>Probable cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antalgic</td>
<td>Arthropathy</td>
</tr>
<tr>
<td>Waddling</td>
<td>Proximal myopathy</td>
</tr>
<tr>
<td>Stamping</td>
<td>Sensory neuropathy</td>
</tr>
<tr>
<td>Foot drop</td>
<td>Peripheral neuropathy or radiculopathy</td>
</tr>
<tr>
<td>Ataxic</td>
<td>Sensory neuropathy or cerebellar disease</td>
</tr>
<tr>
<td>Shuffling/festination</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>Marche à petits pas</td>
<td>Small-vessel cerebrovascular disease</td>
</tr>
<tr>
<td>Hemiplegic</td>
<td>Cerebral hemisphere lesion</td>
</tr>
<tr>
<td>Apraxic</td>
<td>Bilateral hemisphere lesions</td>
</tr>
</tbody>
</table>
been shown to reduce both falls and fracture rates, through effects on both bone mineral density and neuromuscular function. They are not effective in those with osteoporosis living in the community, in whom bisphosphonates are first-line therapy.

In the UK, government policy and National Institute for Health and Clinical Excellence guidelines (www.nice.org.uk) for falls prevention have led to the development of specific Falls and Fracture Prevention Services in many parts of the country.

Dizziness

Dizziness is very common, affecting at least 30% of those aged over 65 years in community surveys. Dizziness can be disabling in its own right and is also a risk factor for falls. Acute dizziness is relatively straightforward and common causes include:

- hypotension due to arrhythmia, myocardial infarction, gastrointestinal bleed or pulmonary embolism
- onset of posterior fossa stroke
- vestibular neuritis.

Although older people more commonly present with recurrent dizzy spells and often find it difficult to describe the sensation they experience, the most effective way of establishing the cause(s) of the problem is nevertheless to determine which of the following is predominant (even if more than one is present):

- lightheadedness, suggestive of reduced cerebral perfusion
- vertigo, suggestive of labyrinthine or brainstem disease (p. 1167)
- unsteadiness/poor balance, suggestive of joint or neurological disease.

In lightheaded patients, structural cardiac disease (such as aortic stenosis) and arrhythmia must be considered, but disorders of autonomic cardiovascular control, such as vasovagal syndrome and postural hypotension, are the most common causes in old age. Hypotensive medication may exacerbate these. Further investigation and treatment are described on page 1157.

Vertigo in older patients is most commonly due to benign positional vertigo (p. 1158), but if other brainstem symptoms or signs are present, MRI of the brain is required to exclude a cerebello-pontine angle lesion.

Delirium

Delirium is a syndrome of transient, reversible cognitive dysfunction. It is very common, affecting up to 30% of older hospital inpatients, either at admission or during their hospital stay. It is associated with high rates of mortality, complication and institutionalisation, and with longer lengths of stay. Risk factors are shown in Box 7.8. Its pathophysiology is unclear; it may in part be due to the effect of increased cortisol release in acute illness, or it may reflect a sensitivity of cholinergic neurotransmission to toxic insults. Older terms for delirium, e.g. acute confusion or toxic confusional state, lack diagnostic precision and should be avoided.

Clinical assessment

Assessment has two main goals: firstly, to establish the diagnosis of delirium; and secondly, to identify all of the reversible precipitating factors to allow optimal treatment.

Delirium may be missed unless routine cognitive testing with an Abbreviated Mental Test, CLOX test or mini-mental state examination (MMSE; p. 234) is performed. Delirium often occurs in patients with dementia, and a history from a relative or carer about the onset and course of confusion is needed to distinguish acute from chronic features. The Confusion Assessment Method (Box 7.9) is a useful tool to diagnose delirium accurately and to differentiate the condition from dementia.

More than one of the precipitating causes of delirium (Fig. 7.3) is often present. Symptoms suggestive of a physical illness, such as an infection or stroke, should be elicited. An accurate drug and alcohol history is required, especially to ascertain whether any drugs have been recently stopped or started.

A full physical examination should be performed, noting in particular:

<table>
<thead>
<tr>
<th>7.8 Risk factors for delirium</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predisposing factors</strong></td>
</tr>
<tr>
<td>Old age</td>
</tr>
<tr>
<td>Dementia</td>
</tr>
<tr>
<td>Frailty</td>
</tr>
<tr>
<td>Intercurrent illness</td>
</tr>
<tr>
<td>Surgery</td>
</tr>
<tr>
<td>Change of environment or ward</td>
</tr>
<tr>
<td>Sensory deprivation (e.g.</td>
</tr>
<tr>
<td>darkness) or overload (e.g.</td>
</tr>
<tr>
<td>noise)</td>
</tr>
<tr>
<td>Medications (e.g. opioids,</td>
</tr>
<tr>
<td>psychotropics)</td>
</tr>
</tbody>
</table>
**7.9 How to make a diagnosis of delirium: the Confusion Assessment Method (CAM)**

**Talk to the patient and assess:**
- **Cognition** (e.g. MMSE, p. 234). A normal score makes delirium unlikely.
- **Inattention.** Can the patient converse with you? If in doubt, give 6–7 digits (between 1 and 9) to remember and repeat back to you; failure suggests inattention.
- **Conscious level.** Alert, hyper-alert or drowsy?
- **Thinking.** Is speech rambling? Does it make sense? Is the patient hallucinating?

**Obtain a collateral history (e.g. from carer, nurse or general practitioner):**
- What is the patient normally like?
- Has there been a sudden deterioration, e.g. over a few days?
- Does confusion fluctuate through the day?

**Consider the diagnosis. Delirium is present if there is:**

- Acute deterioration in cognition, which fluctuates over time
- **AND** Evidence of inattention
- WITH EITHER Evidence of disorganised thinking
- OR Altered level of consciousness (either drowsy/stupor/coma or hyper-alert/agitated/irritable)

---

**Common causes and investigation of delirium.** All investigations are performed routinely, except those in italics. *Tend to present over weeks to months rather than hours to days. The chest X-ray shows consolidation in pneumonia. The CT scan shows a cerebral haemorrhage. (COPD = chronic obstructive pulmonary disease; CRP = C-reactive protein; MI = myocardial infarction; SSRI = selective serotonin re-uptake inhibitor; UTI = urinary tract infection)*

- **Infection**
  - Full blood count, CRP
  - Chest X-ray
  - Urinalysis and culture
  - Others as appropriate: sputum, blood cultures, wound swabs

- **Metabolic disturbance**
  - Urea and electrolytes
  - Plasma calcium
  - Capillary blood and plasma glucose
  - Liver function tests
  - Thyroid function tests
  - B₁₂ and folate

- **Toxic insult**
  - Digoxin level if prescribed

- **Acute neurological conditions**
  - CT brain: only when intracranial lesion is suspected (focal neurological signs, recent fall or head injury) or no other physical cause of delirium is identified
  - Lumbar puncture: only if meningitis or encephalitis is suspected

- **Hypoxia**
  - Pulse oximetry (arterial blood gases if low)
  - Chest X-ray
  - ECG

---

**Fig. 7.3 Common causes and investigation of delirium.** All investigations are performed routinely, except those in italics. *Tend to present over weeks to months rather than hours to days. The chest X-ray shows consolidation in pneumonia. The CT scan shows a cerebral haemorrhage. (COPD = chronic obstructive pulmonary disease; CRP = C-reactive protein; MI = myocardial infarction; SSRI = selective serotonin re-uptake inhibitor; UTI = urinary tract infection)*
Urinary incontinence

Urinary incontinence is defined as the involuntary loss of urine and comes to medical attention when sufficiently severe to cause a social or hygiene problem. It occurs in all age groups but becomes more prevalent in old age, affecting about 15% of women and 10% of men aged over 65. It may lead to skin damage if severe and can be socially restricting. While age-dependent changes in the lower urinary tract predispose older people to incontinence, it is not an inevitable consequence of ageing and requires investigation and appropriate treatment. Urinary incontinence is frequently precipitated by acute illness in old age and is commonly multifactorial (Fig. 7.4).

Initial management is to identify and address contributory factors. If incontinence fails to resolve, further diagnosis and management should be pursued, as described on page 472.

- Urge incontinence is usually due to detrusor over-activity and results in urgency and frequency.
- Stress incontinence is almost exclusive to women and is due to weakness of the pelvic floor muscles, which allows leakage of urine when intra-abdominal pressure rises, e.g., on coughing. It may be compounded by atrophic vaginitis, associated with oestrogen deficiency in old age, which can be treated with oestrogen pessaries.
- Overflow incontinence is most commonly seen in elderly men with prostatic enlargement, which obstructs bladder outflow.

In patients with severe stroke disease or dementia, treatment may be ineffective, as frontal cortical inhibitory signals to bladder emptying are lost. A timed/prompted toileting programme may help. Other than in overflow incontinence, urinary catheterisation should never be viewed as first-line management, but may be required as a final resort if the perineal skin is at risk of breakdown or quality of life is affected.

Adverse drug reactions

Adverse drug reactions (ADRs) and the effects of drug interactions are discussed on pages 24–28. They may result in symptoms, abnormal physical signs and altered laboratory test results (Box 7.10). ADRs are the cause of around 5% of all hospital admissions but account for up to 20% of admissions in those aged over 65. This is partly because older people receive many more prescribed drugs than younger people. Polypharmacy has been defined as the use of four or more drugs; this should be avoided if possible, but is not always inappropriate because many conditions, such as hypertension and heart failure, necessitate the use of several drugs, and older people may have several coexisting medical problems (Box 7.11). However, the more drugs that are taken, the greater the risk of an ADR. This risk is compounded by age-related changes in pharmacodynamic and pharmacokinetic factors (pp. 18–21), and by impaired homeostatic mechanisms, such as baroreceptor responses, plasma volume and electrolyte control. Older people are thus especially sensitive to drugs that can cause postural hypotension or volume depletion (see Box 7.10). Non-adherence to drug therapy also rises with the number of drugs prescribed.

The clinical presentations of ADRs are diverse, so for any presenting problem in old age the possibility that the patient’s medication is a contributory factor should always be considered. Failure to recognise this may lead to the use of a further drug to treat the problem, making matters worse, when the better course would be to stop or reduce the dose of the offending drug or to find an alternative.
### 7.10 Common adverse drug reactions in old age

<table>
<thead>
<tr>
<th>Drug class</th>
<th>Adverse reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSAIDs</td>
<td>Gastrointestinal bleeding and peptic ulceration</td>
</tr>
<tr>
<td></td>
<td>Renal impairment</td>
</tr>
<tr>
<td>Diuretics</td>
<td>Renal impairment, electrolyte disturbance</td>
</tr>
<tr>
<td></td>
<td>Gout</td>
</tr>
<tr>
<td></td>
<td>Hypotension, postural hypotension</td>
</tr>
<tr>
<td>Warfarin</td>
<td>Bleeding</td>
</tr>
<tr>
<td>ACE inhibitors</td>
<td>Renal impairment, electrolyte disturbance</td>
</tr>
<tr>
<td></td>
<td>Hypotension, postural hypotension</td>
</tr>
<tr>
<td>β-blockers</td>
<td>Bradycardia, heart block</td>
</tr>
<tr>
<td></td>
<td>Hypotension, postural hypotension</td>
</tr>
<tr>
<td>Opiates</td>
<td>Constipation, vomiting</td>
</tr>
<tr>
<td></td>
<td>Delirium</td>
</tr>
<tr>
<td></td>
<td>Urinary retention</td>
</tr>
<tr>
<td>Antidepressants</td>
<td>Delirium</td>
</tr>
<tr>
<td></td>
<td>Hypotension, postural hypotension</td>
</tr>
<tr>
<td></td>
<td>Falls</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>Delirium</td>
</tr>
<tr>
<td></td>
<td>Falls</td>
</tr>
<tr>
<td>Anticholinergics</td>
<td>Delirium</td>
</tr>
<tr>
<td></td>
<td>Urinary retention</td>
</tr>
<tr>
<td></td>
<td>Constipation</td>
</tr>
</tbody>
</table>

(ACE = angiotensin-converting enzyme; NSAID = non-steroidal anti-inflammatory drug; SSRI = selective serotonin re-uptake inhibitor)

### 7.12 Other presenting problems in old age

- Hypothermia  p. 104
- Under-nutrition  p. 120
- Dementia  p. 250
- Infection  pp. 296 and 306
- Fluid balance problems  p. 439
- Heart failure  p. 546
- Hypertension  p. 606
- Dizziness and blackouts  pp. 554 and 1157
- Atrial fibrillation  p. 564
- Diabetes mellitus  p. 806
- Peptic ulceration  p. 872
- Anaemia  p. 1001
- Painful joints  p. 1069
- Bone disease and fracture  pp. 1120 and 1071
- Stroke  p. 1231

Other problems in old age

There is a vast range of other presenting problems in older people and they present to many medical specialties. End-of-life care is an important facet of clinical practice in old age and is discussed on page 290. Relevant sections in other chapters are referenced in Box 7.12.

Within each chapter, ‘In Old Age’ boxes highlight the areas in which presentation or management differs from that in younger individuals. These are listed on page 178.

### REHABILITATION

Rehabilitation aims to improve the ability of people of all ages to perform day-to-day activities, and to restore their physical, mental and social capabilities as far as possible. Acute illness in older people is often associated with loss of their usual ability to walk or care for themselves, and common disabling conditions such as stroke, fractured neck of femur, arthritis and cardio-respiratory disease become increasingly prevalent with advancing age.

Disability is an interaction between factors intrinsic to the individual and the context in which they live, and both medical and social interventions are needed to address this (Box 7.13). Doctors tend to focus on health conditions and impairments, but patients are more concerned with the effect on their activities and ability to participate in everyday life.

### 7.11 Factors leading to polypharmacy in old age

- Multiple pathology
- Poor patient education (see Box 2.20, p. 35)
- Lack of routine review of all medications
- Patient expectations of prescribing
- Over-use of drug interventions by doctors
- Attendance at multiple specialist clinics
- Poor communication between specialists

Regular review of medications is important in preventing ADRs. The patient or carer should be asked to bring all medication for review rather than the doctor relying on previous records. Those drugs that are no longer needed or are contraindicated can be discontinued.

### 7.13 International classification of functioning and disability

<table>
<thead>
<tr>
<th>Factor</th>
<th>Intervention required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health condition</td>
<td></td>
</tr>
<tr>
<td>Underlying disease, e.g. stroke, osteoarthritis</td>
<td>Medical or surgical treatment</td>
</tr>
<tr>
<td>Impairment</td>
<td></td>
</tr>
<tr>
<td>Symptoms or signs of the condition, e.g. hemiparesis, visual loss</td>
<td>Medical or surgical treatment</td>
</tr>
<tr>
<td>Activity limitation</td>
<td></td>
</tr>
<tr>
<td>Resultant loss of function, e.g. walking, dressing</td>
<td>Rehabilitation, assistance, aids</td>
</tr>
<tr>
<td>Participation restriction</td>
<td></td>
</tr>
<tr>
<td>Resultant loss of social function, e.g. cooking, shopping</td>
<td>Adapted accommodation Social services</td>
</tr>
</tbody>
</table>
The rehabilitation process

Rehabilitation is a problem-solving process focused on improving the patient’s physical, psychological and social function. It entails:

• **Assessment.** The nature and extent of the patient’s problems can be identified using the framework in Box 7.13. Specific assessment scales, such as the Elderly Mobility Scale or Barthel Index of Activities of Daily Living (Box 7.14), are useful to quantify components of disability, but additional assessment is needed to determine the underlying causes or the interventions required in individual patients.

• **Goal-setting.** Goals should be specific to the patient’s problems, realistic, and agreed between the patient and the rehabilitation team.

• **Intervention.** This includes the active treatments needed to achieve the established goals and to maintain the patient’s health and quality of life. Interventions include hands-on treatment by therapists using a functional, task-orientated approach to improve day-to-day activities, and also psychological support and education. The emphasis on the type of intervention will be individualised, according to the patient’s disabilities, psychological status and progress. The patient and carer(s) must be active participants.

**Multidisciplinary team working**

The core rehabilitation team includes all members of the multidisciplinary team (p. 167). Others may be involved, e.g. audiometry to correct hearing impairment, podiatry for foot problems, and orthotics where a prosthesis or splinting is required. Good communication and mutual respect are essential. Regular team meetings allow sharing of assessments, agreement of rehabilitation goals and interventions, evaluation of progress and planning for the patient’s discharge home. Rehabilitation is not when the doctor orders ‘physio’ or ‘a home visit’, and takes no further role.

**Rehabilitation outcomes**

There is evidence that rehabilitation improves functional outcomes in older people following acute illness, stroke and hip fracture. It also reduces mortality after stroke and hip fracture. These benefits accrue from complex multi-component interventions, but occupational therapy to improve personal ADLs and individualised exercise interventions have now been shown to be effective in improving functional outcome in their own right.

**Further information and acknowledgements**

Websites

http://profane.co Prevention of Falls Network Earth; focuses on the prevention of falls and improvement of postural stability in older people.


www.bgs.org.uk British Geriatrics Society; useful publications on management of common problems in older people and links to other relevant websites.

www.eugms.org European Union Geriatric Medicine Society. Research, position papers and educational resources.

www.iagg.info International Association of Gerontology and Geriatrics. Promoting care of older people and the science of gerontology globally; research, policy and educational resources.


**Figure acknowledgements**

Page 166 insets (Wasted hand, kyphosis) Afzal Mir M. Atlas of clinical diagnosis. 2nd edn. Edinburgh: Saunders; 2003; copyright Elsevier; (Senile purpura) Forbes CD, Jackson WF. Clinical medicine. 3rd edn. Edinburgh: Mosby; 2004 (p. 438, Fig. 10.101); copyright Elsevier; (Venous ulceration) Giovanni Mosti, Clinics in Plastic Surgery, 2012, 39(3): 269–280, Figure 1; copyright © 2012 Elsevier Inc.; all reproduced with permission.
### 7.15 Index of ‘In Old Age’ boxes

<table>
<thead>
<tr>
<th>Environmental and nutritional factors in disease</th>
<th>Diabetes mellitus</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Environmental factors in disease p. 102</td>
<td>• Diagnosis of diabetes mellitus p. 806</td>
</tr>
<tr>
<td>• Thermoregulation p. 104</td>
<td>• Diabetes management p. 828</td>
</tr>
<tr>
<td>• Energy balance p. 122</td>
<td><strong>Alimentary tract and pancreatic disease</strong></td>
</tr>
<tr>
<td>• Vitamin deficiency p. 127</td>
<td>• Endoscopy p. 850</td>
</tr>
<tr>
<td><strong>Infectious diseases</strong></td>
<td>• Acute abdominal pain p. 863</td>
</tr>
<tr>
<td>• Problems with antimicrobial therapy p. 154</td>
<td>• Oral health p. 864</td>
</tr>
<tr>
<td>• Fever p. 236</td>
<td>• Gastro-oesophageal reflux disease p. 868</td>
</tr>
<tr>
<td>• Infectious diarrhoea p. 306</td>
<td>• Peptic ulcer disease p. 875</td>
</tr>
<tr>
<td><strong>Clinical biochemistry and metabolism</strong></td>
<td>• Malabsorption p. 883</td>
</tr>
<tr>
<td>• Hyponatraemia and hypernatraemia p. 439</td>
<td>• Constipation p. 918</td>
</tr>
<tr>
<td>• Management of hyperlipidaemia p. 458</td>
<td><strong>Liver and biliary tract disease</strong></td>
</tr>
<tr>
<td><strong>Kidney and urinary tract disease</strong></td>
<td>• Liver disease p. 980</td>
</tr>
<tr>
<td>• Incontinence p. 473</td>
<td>• Gallbladder disease p. 988</td>
</tr>
<tr>
<td>• Acute kidney injury p. 483</td>
<td><strong>Blood disease</strong></td>
</tr>
<tr>
<td>• Renal replacement therapy p. 491</td>
<td>• Haematological investigations p. 1001</td>
</tr>
<tr>
<td>• Urinary infection p. 512</td>
<td>• Anaemia p. 1034</td>
</tr>
<tr>
<td><strong>Cardiovascular disease</strong></td>
<td>• Haematological malignancy p. 1047</td>
</tr>
<tr>
<td>• Congestive cardiac failure p. 551</td>
<td>• Haemostasis and thrombosis p. 1056</td>
</tr>
<tr>
<td>• Atrial fibrillation p. 566</td>
<td><strong>Rheumatology and bone disease</strong></td>
</tr>
<tr>
<td>• Angina p. 595</td>
<td>• Use of oral NSAID p. 1079</td>
</tr>
<tr>
<td>• Myocardial infarction p. 599</td>
<td>• Osteoarthritis p. 1085</td>
</tr>
<tr>
<td>• Atherosclerotic vascular disease p. 602</td>
<td>• Gout p. 1090</td>
</tr>
<tr>
<td>• Hypertension p. 610</td>
<td>• Joint and bone infection p. 1094</td>
</tr>
<tr>
<td>• Aortic stenosis p. 622</td>
<td>• Osteoporosis p. 1124</td>
</tr>
<tr>
<td>• Endocarditis p. 626</td>
<td><strong>Neurological disease</strong></td>
</tr>
<tr>
<td><strong>Respiratory disease</strong></td>
<td>• Neurological examination p. 1139</td>
</tr>
<tr>
<td>• Respiratory function p. 648</td>
<td>• Dizziness p. 1158</td>
</tr>
<tr>
<td>• Obstructive pulmonary disease p. 677</td>
<td>• Epilepsy p. 1185</td>
</tr>
<tr>
<td>• Respiratory infection p. 686</td>
<td><strong>Other</strong></td>
</tr>
<tr>
<td>• Interstitial lung disease p. 714</td>
<td>• Prescribing p. 36</td>
</tr>
<tr>
<td>• Thromboembolic disease p. 724</td>
<td>• Genetic disease and counselling p. 64</td>
</tr>
<tr>
<td>• Pleural disease p. 731</td>
<td>• Immune senescence p. 82</td>
</tr>
<tr>
<td><strong>Endocrine disease</strong></td>
<td>• The critically ill older patient p. 204</td>
</tr>
<tr>
<td>• The thyroid gland p. 755</td>
<td>• Poisoning p. 208</td>
</tr>
<tr>
<td>• Gonadal function p. 766</td>
<td>• Medical psychiatry p. 238</td>
</tr>
<tr>
<td>• The parathyroid glands p. 770</td>
<td>• Cancer p. 271</td>
</tr>
<tr>
<td>• Glucocorticoids p. 779</td>
<td>• HIV infection p. 410</td>
</tr>
<tr>
<td>• Spontaneous hypoglycaemia p. 784</td>
<td>• Skin changes p. 1254</td>
</tr>
<tr>
<td>• The pituitary and hypothalamus p. 795</td>
<td></td>
</tr>
</tbody>
</table>
Clinical examination of the critically ill patient 180

Physiology of critical illness 182
- Oxygen transport 182
- Cardiovascular component of oxygen delivery: flow 182
- Oxygenation component of oxygen delivery: content 183
- Oxygen consumption 184
- Relationship between oxygen consumption and delivery 184
- Pathophysiology of the inflammatory response 184

Monitoring 185
- Monitoring the circulation 185
- Monitoring respiratory function 187

Recognition of critical illness 188
- Assessment and initial resuscitation of the critically ill patient 188
- Clinical decision-making and referral to critical care 189

Presenting problems/Management of major organ failure 190
- Circulatory failure: ‘shock’ 190
- Circulatory support 190
- Respiratory failure and acute respiratory distress syndrome 191
- Respiratory support 193
- Acute kidney injury 197
- Renal support 197
- Gastrointestinal and hepatic disturbance 198
- Gastrointestinal and hepatic support 198
- Neurological failure (coma) 198
- Neurological support 199
- Neurological complications in intensive care 199
- Sepsis 200
- Management 201
- Disseminated intravascular coagulation 201

General principles of critical care management 201
- Daily clinical management in the ICU 202
- Sedation and analgesia 202

Discharge from intensive care 203
- Withdrawal of intensive support 203
- Brainstem death 203

Outcome of intensive care 204
- Scoring systems 204
**Referral**

**Initial assessment and resuscitation**

A. **Airway clear**
   - No: head tilt, chin lift
   - Yes: take presenting history

B. **Breathing**
   - Distressed: Yes: oxygen
   - Respiratory rate: oxygen if tachypnoeic
   - Auscultation: nebulised salbutamol if wheeze
   - Monitor SpO\textsubscript{2}
   - Obtain arterial blood gases and chest X-ray

C. **Circulation**
   - Pulse rate, rhythm and volume
   - Blood pressure
   - ECG monitor
   - Check peripheral perfusion
   - Establish IV access and 'give fluid'
   - 12-lead ECG

D. **Disability**
   - Establish Glasgow Coma Score (GCS):
   - if reduced, exclude hypoglycaemia
   - Check pupils and limbs for focal signs

E. **Evidence**
   - Look for information to assess severity and establish diagnosis
   - Examination: target initially to systems likely to give most information, e.g. chest and heart if presentation is with breathlessness

*ABG, potassium, glucose, haemoglobin, lactate

**Re-assessment and further management**

A. **Airway compromised**
   - Consider intubation and ventilation
   - Give more oxygen

B. **Breathing**
   - Oxygenation failure: give more oxygen; continuous positive airways pressure (CPAP)
   - Respiratory failure: non-invasive ventilation (NIV); intubation and ventilation

C. **Shock**
   - Large-bore IV access
   - Volume resuscitation
   - Arterial line insertion
   - Central line insertion
   - Monitoring
   - Vasoactive drugs

D. **Reduced GCS**
   - A + B + C and Is CT scan required?

E. **Extricate**
   - To ICU or HDU as appropriate
   - Plan safe transport

F. **Further investigations and full examination**
   - FBC, acute biochemistry, coagulation, cultures, targeted specialist investigations

**Respiratory signs**
- Respiratory arrest
- Threatened or obstructed airway
- Stridor, intercostal recession, paradoxical breathing (seesaw pattern)
- Respiratory rate < 8 or > 35/min
- Respiratory distress: use of accessory muscles; unable to speak in complete sentences
- SpO\textsubscript{2} < 90% on high-concentration oxygen
- Rising PaCO\textsubscript{2} > 7 kPa (52.5 mmHg) or > 2 kPa (> 15 mmHg) above ‘normal’ with acidosis

**Cardiovascular signs**
- Cardiac arrest
- Pulse rate < 40 or > 140 bpm
- Systolic blood pressure < 100 mmHg
- Poor peripheral perfusion
- Evidence of inadequate oxygen delivery
- Metabolic acidosis
- Hyperlactataemia
- Poor response to volume resuscitation
- Oliguria < 0.5 mL/kg/hr (check urea, creatinine, K+)

**Neurological signs**
- Threatened or obstructed airway
- Absent gag or cough reflex
- Failure to maintain normal PaO\textsubscript{2} and PaCO\textsubscript{2}
- Failure to obey commands
- GCS < 10
- Sudden fall in level of consciousness (GCS by > 2 points)
- Repeated or prolonged seizures
Recognition of critical illness: early warning scores

- Record standard observations:
  - Respiratory rate
  - \( \text{SpO}_2 \)
  - Temperature
  - Blood pressure (BP)
  - Heart rate
  - Neurological response

- Note whether the observation falls in a shaded ‘at-risk zone’ (see SEWS key)
- Add the points scored and record total SEWS score on chart
- Do not add ‘Pain’ score to SEWS score

If SEWS score ≥ 4, a doctor should assess the patient within 20 mins.
If SEWS score ≥ 6, a senior doctor should assess the patient within 10 mins.

Clinical features of shock

Low-flow shock, e.g. hypovolaemia, cardiogenic shock

- Rapid, shallow respiration
- Cold, clammy skin
- Tachycardia (> 100/min)
- Hypotension (systolic BP < 100 mmHg)
- Drowsiness, confusion, irritability (usually occurs late)
- Oliguria
- Multi-organ failure

Vasodilated shock, e.g. sepsis, anaphylaxis

- Rapid, shallow respiration (very early)
- Warm peripheries
- Tachycardia (> 100/min)
- Hypotension (systolic BP < 100 mmHg and disproportionately low diastolic BP – early)
- Drowsiness, confusion, irritability (can occur early)
- Oliguria
- Multi-organ failure

*Peripheries may be cool in sepsis with hypovolaemia, or if myocardial depression is present.
A critically ill patient is at imminent risk of death. Recognition, assessment and management of critical illness are thus fundamental to clinical care in any area of medicine. The principle underpinning intensive care is the simultaneous assessment of illness severity and stabilisation of life-threatening physiological abnormalities. The goal is to prevent deterioration and effect improvements as the diagnosis is established, and treatment of the underlying definitive disease process(es) is initiated. Blinkered attention to either resuscitation or diagnosis in isolation results in worse outcomes and increased mortality; the two processes are inextricably interlinked. Appropriate physiological monitoring is required to allow continuing assessment and reassessment of response to therapy, wherever the clinical environment.

### Physiology of Critical Illness

#### Oxygen Transport

The principal function of the heart, lungs and circulation is the provision of oxygen (and other nutrients) to the various organs and tissues of the body. During this process, carbon dioxide and other metabolic waste products are removed. The rate of supply and removal should match the specific metabolic requirements of the individual tissues. This requires adequate oxygen uptake in the lungs, global matching of delivery and consumption, and regional control of the circulation. Failure to supply sufficient oxygen to meet the metabolic requirements of the tissues is the cardinal feature of circulatory failure or ‘shock’, and optimisation of tissue oxygen delivery and consumption is the goal of resuscitation.

Atmospheric oxygen moves down a partial pressure gradient from air, through the respiratory tract, from alveoli to arterial blood and then to the capillary beds and cells, diffusing into the mitochondria, where it is utilised at cytochrome a3 (Fig. 8.1). The movement of oxygen from the left ventricle to the systemic tissue capillaries is known as oxygen delivery (DO₂), and is the product of cardiac output (flow) × arterial oxygen content (CaO₂). The latter is the product of haemoglobin (Hb) × arterial oxygen saturation of haemoglobin (SaO₂) × 1.34. By increasing cardiac output, arterial oxygen saturation or haemoglobin concentration, DO₂ will be increased.

The regional distribution of oxygen delivery is important. If skin and muscle receive high blood flows but the splanchnic bed does not, the gut will become hypoxic even if overall DO₂ is high.

The movement of oxygen from tissue capillary to cell occurs by diffusion and depends on the gradient of oxygen partial pressures, diffusion distance and the ability of the target cell to take up and use oxygen. Microcircular, tissue diffusion and cellular factors thus also influence the oxygen status of the cell.

#### Cardiovascular Component of Oxygen Delivery: Flow

A key determinant of DO₂ is cardiac output, which is determined by the ventricular ‘preload’ and ‘afterload’, myocardial contractility and heart rate.

### Preload

The atrial filling pressures, or preload, determine the end-diastolic ventricular volume, which, according to Starling’s Law and depending on myocardial contractility, defines the force of cardiac contraction and the stroke volume (see Fig. 18.22, p. 547). The principal determinant of preload is venous return, determined by the intravascular volume, venous ‘tone’ and intrathoracic pressure. This can be measured as the central venous pressure (CVP), as described on page 185 (Box 8.1).

When volume is lost (e.g. in major haemorrhage), venous ‘tone’ increases and this helps to offset the consequent fall in atrial filling pressure and stroke volume. If the equivalent volume is restored gradually by intravenous fluid administration, the right atrial pressure will return to normal as the intravascular volume is normalised and the reflex increase in venous tone abates. However, if fluid is infused too rapidly, there is insufficient time for the venous and arteriolar tone to fall and pulmonary oedema may occur, even though the intravascular volume has only been restored to the pre-morbid level.

If the preload is low, volume loading with intravenous fluids is the priority and is the most appropriate means of improving cardiac output and tissue perfusion. The choice of fluid for volume loading is controversial, but as there is no clear advantage of colloid over crystalloid, sodium chloride is used. Fluid challenges of 200–250 mL should be administered rapidly over a couple of minutes, and titrated against heart rate, blood pressure (BP), peripheral circulation, and measurements of CVP (Fig. 8.2). Red cells have traditionally been transfused to...
achieve and maintain a haemoglobin concentration of 100 g/L, but in the absence of significant heart disease, the target is 70–90 g/L (p. 184).

When the preload is high due to excessive intravascular volume or impaired myocardial contractility, removing volume from the circulation by using diuretics or haemofiltration, or increasing the capacity of the vascular bed by using venodilator therapy (glyceryl trinitrate, morphine) often improves stroke volume.

**Afterload**

Afterload is the tension in the ventricular myocardium during systole, and is determined by the resistance to ventricular outflow, which is a function of the peripheral arteriolar resistance.

Understanding the reciprocal relationship between pressure, flow and resistance is crucial for appropriate circulatory management. High resistances produce lower flows at higher pressures for a given amount of ventricular work. Therefore, a systemic vasodilator (see below) will allow the same cardiac output to be maintained for less ventricular work but with a reduced arterial BP. In hyperdynamic sepsis, the peripheral arteriolar tone and BP are low but the cardiac output is often high; therefore the vasoconstrictor noradrenaline (norepinephrine) is appropriate to restore BP, usually at the price of some reduction in cardiac output.

**Myocardial contractility**

This determines the stroke volume that the ventricle can generate against a given afterload for a particular preload. The ventricular stroke work is the external work performed by the ventricle with each beat. The relationship between stroke work and filling pressure is shown in Figure 18.22 (p. 547). Myocardial contractility is frequently reduced in critically ill patients due to pre-existing cardiac disease (usually ischaemic), drugs (e.g. β-blockers, verapamil) or to the disease process itself (particularly sepsis, as the associated low diastolic BP may compromise coronary arterial perfusion). It is thus important to maintain satisfactory perfusion and oxygen delivery to all organs at maximum cardiac efficiency, to minimise myocardial ischaemia.

---

**Oxygenation component of oxygen delivery: content**

The major determinants of the oxygen content of arterial blood (CaO₂) are the arterial oxygen saturation of haemoglobin (SaO₂) and the haemoglobin concentration. Over 95% of oxygen carried in the blood is bound to haemoglobin.

The oxyhaemoglobin dissociation curve (Fig. 8.3) describes the relationship between the saturation of haemoglobin (SO₂) and the partial pressure (PO₂) of oxygen in the blood. A shift in the curve will influence the uptake and release of oxygen by the haemoglobin molecule. If the curve moves to the right, the haemoglobin saturation will be lower for any given oxygen tension: less oxygen will be taken up in the lungs but more will be released to the tissues. As capillary PO₂ rises, the curve moves to the right, increasing the unloading of oxygen in the tissues – a phenomenon known as the Bohr effect. Thus a shift to the right increases capillary PO₂ and hence cellular oxygen supply.

Due to the shape of the curve, a small drop in arterial PO₂ (PaO₂) below 8 kPa (60 mmHg) will cause a marked fall in SaO₂. Its position and the effect of various physico-chemical factors are defined by the PO₂ at which 50% of the haemoglobin is saturated (P₅₀), which is normally 3.5 kPa (26 mmHg). The shape of the curve also means that increases in PaO₂ beyond the level that ensures SaO₂ is greater than 90% produce relatively small additional increases in CaO₂ (Fig. 8.3). Thus, in a patient who is both anaemic (Hb 60 g/L or 6 g/dL) and hypoxaemic (SaO₂ 75%) when breathing air (fractional inspired oxygen concentration (FiO₂) 20%), supplementary oxygen at FiO₂ 40% will increase SaO₂ to 93% and CaO₂ by 24%. However, further increases in FiO₂ while raising PaO₂ cannot produce any further useful increases in SaO₂ or CaO₂. However, increasing haemoglobin to 90 g/L
CRITICAL ILLNESS

(9 g/dL) by blood transfusion will result in a further 50% increase in \( \text{CaO}_2 \).

Traditionally, the optimum haemoglobin concentration for critically ill patients was considered to be approximately 100 g/L (10 g/dL), representing a balance between maximising the oxygen content of the blood and avoiding regional microcirculatory problems due to increased viscosity. However, improved outcomes have been demonstrated when the haemoglobin is maintained between 70 and 90 g/L (7–9 g/dL). A target haemoglobin of 100 g/L remains appropriate in the elderly and in patients with coronary artery disease, cardiogenic shock, significant aortic stenosis or acute brain trauma.

Oxygen consumption

The sum of the oxygen consumed by the various organs represents the global oxygen consumption (\( \text{VO}_2 \)), and is approximately 250 mL/min for an adult of 70 kg undertaking normal daily activities.

The oxygen saturation in the pulmonary artery, or ‘mixed venous oxygen saturation’ (\( \text{SvO}_2 \)), is a measure of the oxygen not consumed by the tissues (\( \text{DO}_2 - \text{VO}_2 \)). The saturation of venous blood from different organs varies considerably; the hepatic venous saturation usually does not exceed 60% but the renal venous saturation may reach 90%, reflecting the difference in the metabolic requirements of these organs, and the oxygen content of the blood delivered to them. The \( \text{SvO}_2 \) is a flow-weighted average measured in the mixed effluent blood from all perfused tissues, and is influenced by changes in both oxygen delivery (\( \text{DO}_2 \)) and consumption (\( \text{VO}_2 \)). Provided the microcirculation and the mechanisms for cellular oxygen uptake are intact, it can be used to monitor whether global oxygen delivery is adequate to meet overall demand, so its measurement is particularly useful in low-flow situations such as cardiogenic shock. Central venous oxygen saturation (\( \text{ScvO}_2 \)) is used in the same way, but as it does not reflect hepato-splanchnic oxygen consumption, it may be less helpful than \( \text{SvO}_2 \).

The re-oxygenation of the blood that returns to the lungs and the resulting arterial saturation (\( \text{SaO}_2 \)) will depend on how closely pulmonary ventilation and perfusion are matched. If part of the pulmonary blood flow perfuses non-ventilated parts of the lung (‘shunting’), the blood entering the left atrium will be desaturated in proportion to the size of the shunt and the level of \( \text{ScvO}_2 \).

Relationship between oxygen consumption and delivery

The tissue oxygen extraction ratio (OER) is 20–25% in a normal individual at rest, but rises as consumption increases or supply diminishes. The maximum OER is approximately 60% for most tissues; at this point, no further increase in extraction can occur and any further increase in oxygen consumption or decline in oxygen delivery will cause tissue hypoxia, anaerobic metabolism and increased lactic acid production. This ultimately results in multiple organ failure and an increased risk of death.

In practice, if there is a metabolic acidosis, hyperlactataemia and/or oliguria that could be due to inadequate oxygen delivery, a therapeutic trial of increased oxygen delivery (while maintaining an adequate BP) may be helpful clinically. If oxygen consumption rises, it can indicate an oxygen debt that is being repaid.

Pathophysiology of the inflammatory response

The mediators and clinical manifestations of the inflammatory response are described on page 82. In critically ill patients, these have important consequences (Box 8.2).

Fever, tachycardia with warm peripheries, tachypnoea and a raised white cell count prompt a diagnosis of sepsis, with the presentation caused by invading microorganisms and their breakdown products. Other conditions, such as pancreatitis, trauma, malignancy, tissue necrosis (e.g. burns), aspiration syndromes, liver failure, blood transfusion and drug reactions, can also present in this way in the absence of infection.

Local inflammation

The body’s initial response to a noxious local insult is to produce a local inflammatory response, with sequestration and activation of white blood cells and the release of a

8.2 Terminology in the inflammatory state

Infection

- Invasion of normally sterile tissue by microorganisms

Bacteraemia

- Viable bacteria in the blood

Systemic inflammatory response syndrome (SIRS)

- Defined by the presence of two or more of
  - Respiratory rate > 20/min
  - Heart rate > 90/min
  - White blood count > 12 × 10⁹/L or < 4 × 10⁹/L
  - Temperature > 38.0°C or < 36.0°C
  - \( \text{PaCO}_2 \) < 4.3 kPa (< 32 mmHg) or ventilated
  - A wide pulse pressure, e.g. 115/42 mmHg, may be an early pointer to systemic sepsis
  - Cause may be infection or a non-infective condition, e.g. pancreatitis, trauma, cardiopulmonary bypass, vasculitis etc.
  - Hypothermia and septic neutropenia indicate more severe infection

Sepsis

- Systemic inflammatory response caused by documented infection

Severe sepsis/SIRS

- Sepsis/SIRS with evidence of early organ dysfunction or hypotension

Septic/SIRS shock

- Sepsis associated with organ failure and hypotension (systolic BP < 90 mmHg or > 40 mmHg fall from baseline) unresponsive to fluid resuscitation

Multiple organ dysfunction syndrome (MODS)

- Development of impaired organ function in a patient with SIRS
  - Multiple organ failure (MOF) ensues unless there is prompt treatment of the underlying cause and appropriate organ support
variety of mediators to overcome the primary ‘insult’ and prevent further damage locally or in distant organs.

Normally, a delicate balance is achieved between pro- and anti-inflammatory mediators. However, if the response is excessive, a large array of pro-inflammatory mediators may be released into the circulation (p. 74). The inflammatory and coagulation cascades are intimately linked, as the latter cause not only platelet activation and fibrin deposition, but also activation of leukocytes and endothelial cells. Conversely, leukocyte activation induces tissue factor expression and initiates coagulation pathways. The natural anticoagulants, antithrombin (AT III), activated protein C (APC) and tissue factor pathway inhibitor (TFPI), inhibit pro-inflammatory cytokines. Deficiency of AT III and APC (features of disseminated intravascular coagulation (DIC), p. 1056) facilitates thrombin generation and promotes further endothelial cell dysfunction.

**Systemic inflammation**

In a severe inflammatory response, systemic release of cytokines and other mediators triggers widespread interaction between the coagulation pathways, platelets, endothelial cells and monocytes, tissue macrophages, and neutrophils. Activated neutrophils express adhesion factors, which make them adhere to and initially roll along the endothelium, before adhering firmly and migrating through the damaged and disrupted endothelium into the extravascular interstitial space (together with fluid and proteins), resulting in tissue oedema and inflammation. A vicious circle of endothelial injury, intravascular coagulation, microvascular occlusion, tissue damage and further release of inflammatory mediators ensues. This can occur in all organs, manifesting in the lungs as acute lung injury and in the kidneys as acute tubular necrosis (ATN). Similar processes probably account for damage to other organs, including the heart.

The endothelium itself produces mediators that control local blood vessel tone. The profound vasodilatation that characterises septic shock and some other acute systemic inflammatory states, such as pancreatitis, results from excessive production of nitric oxide (NO, p. 82), due to activation of inducible NO synthase enzymes.

Systemic inflammatory processes also have important effects on mitochondrial function, resulting in impaired oxidative phosphorylation and aerobic energy generation. This block to oxygen utilisation by cells is sometimes called cytopathic hypoxia. Patients typically have a reduced arteriovenous oxygen difference, a low oxygen extraction ratio, a raised plasma lactate and a paradoxically high mixed venous oxygen saturation (SvO₂), despite normal or supranormal oxygen delivery. This is associated with the development of multiple organ failure (MOF) and reduced survival.

**MONITORING**

Monitoring in intensive care includes a combination of clinical and automated recordings. Electrocardiogram (ECG), SpO₂ (oxygen saturation), BP and usually CVP recordings are taken at least hourly, using either a 24-hour chart or a computerised system. Urine output measurement requires early catheterisation. All invasive haemodynamic monitoring should be referenced to the mid-axillary line as ‘zero’. Clinical monitoring of physical signs, such as respiratory rate, the appearance of the patient, restlessess, conscious level and indices of peripheral perfusion (pale, cold skin; delayed capillary refill in the nailbed), is just as important as a set of blood gases or monitor readings.

**Monitoring the circulation**

**Electrocardiogram**

Standard monitors display a single-lead ECG, record heart rate and identify rhythm changes. More sophisticated machines can print out rhythm strips and monitor ST segment shift, which is useful in patients with ischaemic heart disease.

**Blood pressure**

In critically ill patients, continuous intra-arterial monitoring is necessary using a line placed in the radial artery (or the femoral in vasoconstricted patients or where access is difficult). The brachial artery should be avoided, as it is an end artery of relatively small calibre and occlusion leads to ischaemia of the hand. When there is systemic vasoconstriction, the mean arterial pressure (MAP) may be normal or even high, although the cardiac output is low. Conversely, if there is peripheral vasodilatation, as in sepsis, the MAP may be low, although the cardiac output is high.

**Central venous pressure**

CVP or right atrial pressure (RAP) is monitored using a catheter inserted via either the internal jugular or the subclavian vein, with the distal end sited in the upper right atrium. The CVP may help in assessing the need for intravascular fluid replacement and the rate at which this should be given (see Box 8.1, p. 182). If the CVP is low in the presence of a low MAP or cardiac output, fluid resuscitation is necessary. However, a raised level does not necessarily mean that the patient is adequately volume-resuscitated. Right heart function, pulmonary artery pressure, intrathoracic pressure and venous ‘tone’ also influence CVP, and may lead to a raised CVP even when the patient is hypovolaemic (Box 8.3). In addition, positive pressure ventilation raises intrathoracic pressure and causes marked swings in atrial pressures and systemic BP in time with respiration. Pressure measurements should be recorded at end-expiration.

In severe hypovolaemia, the RAP may be sustained by peripheral vasoconstriction, and transfusion may initially produce little or no change in the CVP (see Fig. 8.3, p. 183).

**Pulmonary artery catheterisation and pulmonary artery ‘wedge’ pressure**

The CVP is usually an adequate guide to the filling pressures of both sides of the heart. However, certain conditions, such as pulmonary hypertension or right ventricular dysfunction, may lead to raised CVP levels even in the presence of hypovolaemia. In these circumstances, it may be appropriate to insert a pulmonary artery flotation catheter (Fig. 8.4) so that pulmonary artery pressure and pulmonary artery ‘wedge’ pressure
Critically ill illness (PAWP), which approximates to left atrial pressure, can be measured. The mean PAWP normally lies between 6 and 12 mmHg (measured from the mid-axillary line) but in left heart failure it may be grossly elevated, exceeding 30 mmHg. Provided the pulmonary capillary membranes are intact, the optimum PAWP when managing acute circulatory failure in the critically ill patient is generally 12–15 mmHg, because this will ensure good left ventricular filling without risking hydrostatic pulmonary oedema.

Pulmonary artery catheters also allow measurement of cardiac output and sampling of blood from the pulmonary artery (‘mixed venous’ samples), permitting continuous monitoring of the mixed venous oxygen saturation ($SvO_2$) by oximetry. Measurement of $SvO_2$ gives an indication of the adequacy of cardiac output (and hence $DO_2$) in relation to the body’s metabolic requirements. It is especially useful in low cardiac output states.

### Cardiac output
Measurement of cardiac output is important, particularly when large doses of a vasopressor are being administered, when there is underlying cardiac disease (acute or chronic), and when volume resuscitation and vaso-active drug therapy are not achieving resolution of lactic acidosis or oliguria. It is most accurately measured by indicator dilution methods. Most PA catheters incorporate a heating element, which raises blood temperature at frequent intervals, and the resultant temperature change is detected by a thermistor at the tip of the catheter.

---

**Fig. 8.4 A pulmonary artery catheter.**

There is a small balloon at the tip of the catheter and pressure can be measured through the central lumen. The catheter is inserted via an internal jugular, subclavian or femoral vein and advanced through the right heart until the tip lies in the pulmonary artery. When the balloon is deflated, the pulmonary artery pressure can be recorded. Advancing the catheter with the balloon inflated will ‘wedge’ the catheter in the pulmonary artery. Blood cannot then flow past the balloon, so the tip of the catheter will now record the pressure transmitted from the pulmonary veins and left atrium (known as the pulmonary artery wedge pressure), which provides an indirect measure of the left atrial pressure. ($LA = left\ atrium; LV = left\ ventricle; RA = right\ atrium; RV = right\ ventricle$).
Monitoring

Peripheral skin temperature

In general, resuscitation is not complete until the patient’s feet are warm and well perfused.

Blood lactate, hydrogen ion and base excess/deficit

Acid–base balance is discussed on page 443. Base excess or deficit is calculated as the difference between the patient’s bicarbonate and the normal bicarbonate after the \( P_{CO_2} \) has been maintained in a blood gas machine at 5.33 kPa (40 mmHg). This is particularly useful, as it describes patients’ underlying metabolic status independently of their current respiratory status. A metabolic acidosis with base deficit of more than 5 mmol/L requires investigation (p. 445). It often indicates increased lactic acid production in poorly perfused, hypoxic tissues, and impaired lactate metabolism and clearance due to poor hepatic perfusion. Serial lactate measurements may therefore be helpful in monitoring tissue perfusion and response to treatment. Other conditions, such as acute renal failure, ketoacidosis and poisoning, may be the cause, and infusions of large volumes of fluids containing sodium chloride may lead to a hyperchloremic acidosis.

Monitoring respiratory function

Oxygen saturation

Oxygen saturation (\( SpO_2 \)) is measured by a probe attached to a finger or earlobe. Spectrophotometric analysis determines the relative proportions of saturated and desaturated haemoglobin. It is unreliable if peripheral perfusion is poor, in the presence of nail polish, excessive movement or high ambient light. It is not useful in carbon monoxide poisoning, as it does not detect carboxy-haemoglobin. If this is suspected, \( PO_2 \) must be measured in an arterial blood gas sample. In general, arterial oxygenation is safe if \( SpO_2 \) is above 90%. Box 8.4 lists the causes of sudden falls in \( SpO_2 \).

Arterial blood gases

Arterial blood gases (ABGs) are measured several times a day in a ventilated patient so that inspired oxygen (\( FiO_2 \)) and minute volume can be adjusted to achieve the

---

**Box 8.4 Causes of sudden changes in oxygen saturation**

**Patient factors**

- Bronchospasm
- Lung collapse due to thick secretions blocking the proximal bronchial tree
- Pneumothorax
- Impaired peripheral perfusion

**Equipment factors**

- Displacement of the endotracheal tube (extubation, endobronchial intubation)
- Blockage of the endotracheal tube
- Disconnection from the ventilator
- Oxygen supply failure
- Detached probe

---

**Fig. 8.5 Oesophageal Doppler ultrasonography.**

Oesophageal Doppler ultrasonography provides a rapid and useful assessment of volume status and cardiac performance to guide early fluid and vasoactive therapy. A 6 mm probe is inserted into the distal oesophagus, allowing continuous monitoring of the aortic flow signal from the descending aorta (Fig. 8.5). Using the stroke distance (area under the velocity/time waveform) and a correction factor that incorporates the patient’s age, height and weight, an estimate of left ventricular stroke volume and hence cardiac output can be made. Peak velocity is an indicator of left ventricular performance, while flow time is an indicator of left ventricular filling and peripheral resistance.

Analysis of arterial pressure waveform is another means of continuously estimating cardiac output, and can be calibrated either by transpulmonary thermodilution (PiCCO) or lithium dilution methods (LiDCO). The Vigileo/Flotrac system derives cardiac output from arterial pressure waveform analysis with no external calibration.

Echocardiography

In many centres, echocardiography is increasingly being used for rapid assessment of myocardial function and volume status at the bedside. Continuous transoesophageal echocardiography allows direct assessment of cardiac filling status and ventricular function in real time.

Urine output

This is a sensitive measure of renal perfusion, provided that the kidneys are not damaged (e.g. in ATN) or affected by drugs such as diuretics or dopamine. Output is measured hourly and a lower limit of normal of 0.5 mL/hr/kg body weight is widely used. It reflects renal perfusion over the hours preceding measurement rather than in real time.
CRITICAL ILLNESS

desired $PaO_2$ and $PaCO_2$ respectively. ABG results are also used to monitor disturbances of acid–base balance.

**Lung function**

In ventilated patients, lung function is monitored by:
- arterial $PO_2$ taken in relation to the fractional inspired oxygen concentration ($PO_2$/FiO$_2$ ratio) and level of end-expiratory pressure
- arterial and end-tidal CO$_2$, reflecting alveolar ventilation
- airway pressures and tidal volumes, reflecting lung compliance and airways resistance.

**Capnography**

The CO$_2$ concentration in inspired gas is zero, but during expiration, after clearing the physiological dead space, it rises progressively to reach a plateau that represents the alveolar or end-tidal CO$_2$ concentration. This cyclical change in CO$_2$ concentration, or capnogram, is measured using an infrared sensor inserted between the ventilator tubing and the endotracheal tube (Fig. 8.6). In normal lungs, the end-tidal CO$_2$ closely mirrors $PaCO_2$, and can be used to assess the adequacy of alveolar ventilation. However, its use is limited as there may be marked discrepancies in the presence of lung disease or impaired pulmonary perfusion (e.g. due to hypovolaemia). In combination with the gas flow and respiratory cycle data from the ventilator, CO$_2$ production and hence metabolic rate may be calculated. In clinical practice, end-tidal CO$_2$ is used to confirm correct placement of an endotracheal tube, in the management of head injury, and during the transport of ventilated patients. Continuous measurement of end-tidal CO$_2$ is important in the minute-to-minute monitoring of any patient ventilated through an endotracheal tube or tracheostomy in the acute setting.

**Transcutaneous PCO$_2$**

Monitors that measure transcutaneous PCO$_2$ are now available with an earlobe probe that incorporates a pulse oximeter and CO$_2$ electrode. The transcutaneous PCO$_2$ closely approximates to $PaCO_2$ and gives continuous monitoring. This is useful in patients with no arterial cannula but who require close monitoring: for example, during ventilatory weaning (see below).

---

**RECOGNITION OF CRITICAL ILLNESS**

The immediate appearance of the patient yields a wealth of information. Introducing yourself, shaking hands and asking ‘How are you?’ allow assessment of:
- the airway (for patency and noises, e.g. stridor, snoring, gurgling, none)
- breathing (rate, symmetry, work of breathing, including accessory muscle use, paradoxical chest/abdominal movement or see-saw pattern)
- peripheral circulation (temperature of the extremities)
- conscious level (the response of the patient).

Tachypnoea is often the earliest abnormality to appear and the most sensitive sign of a worsening clinical state, but it is the least well documented. In the UK, the use of early warning scores, such as the Standard Early Warning System chart (SEWS, p. 181), has been adopted to improve the recognition of critical illness. These alert staff to severely ill patients, complement clinical judgement and facilitate the prioritisation of clinical care. A patient with a SEWS score of 4 or more requires urgent review and appropriate interventions. An elevated score correlates with increased mortality.

**Assessment and initial resuscitation of the critically ill patient**

**Airway and breathing**

If the patient is talking, the airway is clear and breathing is adequate. A rapid history should be obtained whilst initial assessment is undertaken.

Assess breathing as described above. Supplemental oxygen should be administered to patients who are breathless, tachypnoeic or bleeding, or who have chest pain or reduced conscious level. The clinical status of the patient determines how much oxygen to give, but the critically ill should receive at least 60% oxygen initially. High-concentration oxygen is best given using a mask with a reservoir bag, which, at 15 L/min, can provide nearly 90% oxygen. ABGs should be checked early to assess oxygenation, ventilation ($PaCO_2$) and metabolic state (pH or $H^+$, HCO$_3$ and base deficit). Oxygen therapy should be adjusted in light of the ABGs, remembering...
that oxygen requirements may subsequently increase or decrease. Early application of pulse oximeter monitoring is ideal, although this may not be reliable if the patient is peripherally shut down. Intubation, while often essential, may be hazardous in a patient with cardiorespiratory failure, and full monitoring and resuscitation facilities must be available.

**Circulation**

The carotid pulse should be sought in the collapsed or unconscious patient, but peripheral pulses checked in the conscious. The radial, brachial, foot and femoral pulses may disappear as shock progresses, and this indicates the severity of circulatory compromise.

Venous access for the administration of drugs and/or fluids is vital but often difficult in sick patients. The gauge of cannula needed is dictated by its purpose. Wide-bore cannulae are required for rapid fluid administration. Ideally, two 16G or larger cannulae should be inserted, one in each arm, in the severely hypovolaemic patient. If the two cannulae are of different sizes, the pulse oximeter should be placed on the same side as the larger one, and the BP cuff on the same side as the smaller one. This facilitates unimpeded volume resuscitation and uninterrupted oxygen saturation monitoring. Pressure infusors and blood warmers should be utilised for rapid, high-volume fluid resuscitation, particularly of blood products. An 18G cannula is adequate for drug administration.

Machine-derived cuff BP measurement is inaccurate at extremes of BP and in tachycardia, especially atrial fibrillation. Manual sphygmomanometer BP readings tend to be more accurate in hypotension. If severe hypotension is not readily corrected with fluid, early consideration should be given to arterial line insertion and vasoactive drug therapy.

**Disability**

Conscious level should be assessed using the Glasgow Coma Scale (GCS; see Box 26.15, p. 1160). Best eye, verbal and motor responses should be assessed and documented. Appropriate painful stimuli include supraorbital pressure and trapezius pinch. A score of 8 or less denotes coma with associated airway compromise and loss of airway protection, which necessitates intervention. Focal neurological signs may indicate unilateral cerebral pathology. Abnormal pupil size, symmetry or reaction to light may indicate primary cerebral disease or global cerebral insults induced by drugs (e.g. opioids), hypoxia or hypoglycaemia.

**Exposure, evidence and examination**

‘Exposure’ indicates the need for targeted clinical examination, and ‘evidence’ may be gathered from any recent investigations, prescription or monitoring charts.

**Clinical decision-making and referral to critical care**

During the initial assessment and resuscitation, several decisions must be made (Box 8.5), but particularly whether referral to the critical care service is necessary.

**8.5 Clinical decisions in the critically ill**

- How ill is the patient?
- How much help is needed and how quickly?
- Where would the patient be best managed?
- When should the patient be moved?
  - All critically ill patients require an appropriately trained escort during transfer
- What is required before transporting the patient?
  - ABCDE resuscitation ± endotracheal intubation and ventilation
  - Monitoring, including invasive arterial ± CVP
  - Volume resuscitation and vasoactive support
  - Imaging/diagnostic processes
- Is specialist involvement required?
  - E.g. transfer to specialist liver, burns, neurosurgical or cardiac surgical units or for specialised investigations
  - The urgency of specialist treatment needs balanced against the patient’s condition; it may be necessary to stabilise the patient in the ICU first

**8.6 Admission criteria for intensive care (ICU) and high-dependency units (HDU)**

**ICU**

- Patients requiring/likely to require endotracheal intubation and invasive mechanical ventilatory support
- Patients requiring support of two or more organ systems (e.g. inotropes and haemofiltration)
- Patients with chronic impairment of one or more organ systems (e.g. COPD or severe ischaemic heart disease) who require support for acute reversible failure of another organ

**HDU**

- Patients requiring detailed observation or monitoring that cannot be provided at ward level
  - Direct arterial BP monitoring
  - CVP monitoring
  - Fluid balance
  - Neurological observations, regular GCS recording
- Patients requiring support for a single failing organ system, excluding invasive ventilatory support:
  - CPAP or NIV – see p. 193
  - Moderate inotropic or vasopressor support
  - Renal replacement therapy in an otherwise stable patient
- Step down from intensive care

This requires local knowledge about the clinical areas providing enhanced care, whether intermediate high-dependency or advanced intensive care, and the mechanism of referral.

- **Intensive care units** allow management of the sickest patients who require invasive ventilation, multimodal monitoring and multiple organ system support (Box 8.6).
- **High-dependency care** allows a greater degree of monitoring, physiological support and nursing/medical input than the standard ward, for patients following major surgery, or for the septic patient requiring invasive haemodynamic monitoring and...
circulatory support alone, or for the patient with respiratory failure manageable with non-invasive ventilation (NIV) or continuous positive airway pressure (CPAP).

The mechanism of referral to critical care varies between hospitals, and all clinical staff must be aware of the local system and how to access it. Many hospitals have medical emergency or outreach teams that facilitate this. A clear understanding of what is available in critical care and what is achievable allows early referral of appropriate patients, which will improve their survival and reduce length of stay. It prevents referral of patients who have no realistic prospect of meaningful survival, due to either the overwhelming nature of their acute condition or the lack of definitive therapy for the underlying disease process.

**PRESENTING PROBLEMS/MANAGEMENT OF MAJOR ORGAN FAILURE**

As the diagnosis and management of presenting problems take place simultaneously in critical care, these are described together.

**Circulatory failure: ‘shock’**

The defining feature of ‘shock’ is a level of oxygen delivery (DO2) that fails to meet the metabolic requirements of the tissues. ‘Shock’ is not synonymous with hypotension, which is often a late manifestation. The cardiac output and oxygen delivery may be critically low, even though the BP remains normal, and the underlying problem should be identified and treated before the BP falls. Objective markers of inadequate tissue oxygen delivery, such as increasing base deficit, elevated blood lactate and reduced urine output, can aid earlier identification of shock.

The causes of circulatory failure or ‘shock’ may be categorised as either low flow or stroke volume, or low peripheral arteriolar resistance (vasodilatation).

**Low stroke volume**

- **Hypovolaemic**: any condition provoking a major reduction in blood volume, e.g. internal or external haemorrhage, severe burns, salt and water depletion.
- **Cardiogenic**: severe cardiac impairment, e.g. myocardial infarction, acute mitral regurgitation. Subarachnoid haemorrhage may cause catecholamine-mediated myocardial stunning that can result in pulmonary oedema or cardiogenic shock.
- **Obstructive**: obstruction to blood flow around the circulation, e.g. major pulmonary embolism, cardiac tamponade, tension pneumothorax.

**Vasodilatation**

- **Sepsis/SIRS**: infection or other causes of a systemic inflammatory response that produce widespread endothelial damage with vasodilatation, arteriovenous shunting, microvascular occlusion, capillary leak and tissue oedema.

- **Anaphylactic**: inappropriate vasodilatation triggered by an allergen (e.g. bee sting), often associated with endothelial disruption and capillary leak.
- **Neurogenic**: caused by major brain or spinal injury, which disrupts brainstem and neurogenic vasomotor control. High cervical cord trauma may result in disruption of the sympathetic outflow tracts, leading to inappropriate bradycardia due to a combination of loss of noradrenaline (norepinephrine)-mediated vasoconstriction and adrenaline (epinephrine)-mediated chronotropy. Guillain–Barré syndrome (p. 1224) involves the autonomic as well as the sensorimotor systems, which may result in periods of severe hypotension or hypertension.

**Clinical assessment and complications**

Clinical features depend on the primary pathophysiology (p. 180). Hypovolaemic, cardiogenic and obstructive causes of circulatory failure produce the ‘classical’ image of shock with cold peripheries, reduced or absent peripheral pulses, weak central pulses and evidence of a low cardiac output. In early haemorrhagic shock, a narrowed pulse pressure, i.e. a raised diastolic (DBP) and reduced systolic (SBP) blood pressure, such as 105/95 mmHg, indicates the combination of hypovolaemia (reduced stroke volume, hence SBP) and activation of the sympathetic nervous system, with noradrenaline (norepinephrine) inducing vasoconstriction and so raising the DBP.

In contrast, sepsis/SIRS and anaphylactic shock are usually associated with warm peripheries, bounding pulses and features of a high cardiac output. The BP pattern is again distinctive (e.g. 115/42 mmHg), with a low DBP in the early stages due to peripheral vasodilatation, but a normal systolic BP, as the left ventricular afterload is reduced and stroke volume thus maintained. These patients are usually warm peripherally, but in more advanced septic or anaphylactic shock, SBP falls and the peripheries become cool. This is usually due to the hypovolaemia associated with capillary leak and will respond to fluid resuscitation. If there is no improvement with this, myocardial depression may be present.

Neurogenic shock often results in vasodilated hypotension with a paradoxically slow heart rate. All forms of shock require early identification and treatment because, if inadequate regional tissue perfusion and cellular dysoxia persist, MOF will develop. Early institution of invasive haemodynamic monitoring is required.

**Circulatory support**

The primary goals (Box 8.7) are to:

- Restore global oxygen delivery (DO2) by ensuring adequate cardiac output.
- Maintain an MAP that ensures adequate perfusion of vital organs. The target pressure will be patient-specific, depending on pre-morbid factors (e.g. hypertension or coronary artery disease), and may range from 60 to 90 mmHg.

The first objective is to ensure that an ‘appropriate’ ventricular preload is restored, initially by adequate volume resuscitation. Vasoactive drugs may then have to be considered.
8.7 Immediate management of circulatory collapse

Correct hypoxaemia
- Oxygen therapy
- Consider ventilation
  - Intractable hypoxaemia
    - Hypercapnia: $\text{PaCO}_2 > 6.7$ kPa (50 mmHg)
    - Respiratory distress
    - Impaired conscious level

Assess circulation
- Heart rate
- BP: direct arterial pressure
- Peripheral perfusion

Optimise volume status
- Fluid challenge(s):
  - CVP < 6 mmHg: 250 mL 0.9% saline or colloid
  - CVP > 6 mmHg or poor ventricular function suspected:
    - 100 mL boluses and consider measuring cardiac output by PA catheter or oesophageal Doppler

Optimise haemoglobin concentration
- Transfuse red cells to maintain Hb at 70–90 g/L (or 100 g/L if ischaemic heart disease)
- Septic patients can become profoundly anaemic with crystalloid/colloid resuscitation due to haemodilution

Achieve target BP
- Use vasopressor/inotrope once hypovolaemia is corrected

Achieve adequate CO and $D_O_2$
- Inotrope if fluid alone is inadequate

Other measures
- Establish monitoring, including invasive measures, at once
- Trends in haemodynamics, ABG, $H^+$, base deficit and lactate guide further treatment

Therapeutic options to optimise cardiac function

If the cardiac output is inadequate and myocardial contractility is poor, the available treatment options are to:
- **Reduce afterload.** Reduction can be achieved by using an arteriolar dilator (e.g. nitrates), but this may be limited by the consequent fall in systemic pressure. A counterpulsation intra-aortic balloon pump offers the ideal physiological treatment because it reduces left ventricular afterload while increasing cardiac output, diastolic pressure and coronary perfusion. It is particularly valuable in treating myocardial ischaemia.
- **Increase preload.** If there is significant impairment of myocardial contractility, giving fluids to increase filling pressures will only produce a small increase in stroke volume and cardiac output, and risks precipitating pulmonary oedema.
- **Improve myocardial contractility.** An inotrope may be required to ensure adequate cardiac output and peripheral blood flow sufficient to secure adequate oxygen delivery. Box 8.8 lists some characteristics of the commonly used vasoactive agents.

8.8 Actions of commonly used vasoactive agents

<table>
<thead>
<tr>
<th>Drug</th>
<th>Vasoconstrictor</th>
<th>Inotrope</th>
<th>Chronotrope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenaline (epinephrine)</td>
<td>++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Noradrenaline (norepinephrine)</td>
<td>++++</td>
<td>+</td>
<td>(+)</td>
</tr>
<tr>
<td>Dobutamine</td>
<td>*</td>
<td>++++</td>
<td>++</td>
</tr>
</tbody>
</table>

In most patients dobutamine acts as a vasodilator but in some it causes vasoconstriction.

- **Control heart rate and rhythm.** The optimum heart rate is usually between 90 and 110 beats per minute. Correction of low serum potassium and magnesium concentrations should be the first step in treating tachyarrhythmias in the critically ill. Atrial fibrillation is particularly common; intravenous amiodarone (300 mg over 30–60 minutes, followed by 900 mg over 24 hours) can be successful in controlling ventricular rate and in restoring and maintaining sinus rhythm. Other anti-arrhythmic agents are described on page 573.

The management of cardiac tamponade and pulmonary embolism is described on pages 545 and 723, respectively. Circulatory support in the context of sepsis is described below.

Prognosis

If the precipitating cause and accompanying circulatory failure are dealt with promptly, before significant organ failure occurs (‘early’ shock), the prognosis is good. If not, there is progressive deterioration in organ function and MOF ensues (‘late’ shock). The mortality of MOF is high and increases with the number of organs that have failed, the duration of organ failure and the patient’s age. Failure of four or more organs is associated with a mortality of more than 80%.

Respiratory failure and acute respiratory distress syndrome

Respiratory failure may be the primary problem or could constitute a secondary complication (Box 8.9). The pattern of respiratory failure is classified using ABG analysis:
- **type 1:** hypoxaemia ($P_{a\text{O}_2} < 8$ kPa (< 60 mmHg) when breathing air) without hypercapnia
- **type 2:** hypoxaemia with hypercapnia ($P_{a\text{CO}_2} > 6.5$ kPa (> 49 mmHg)) due to alveolar hypoventilation

Acute hypoxaemia results from an increase in ventilation–perfusion mismatch within the lung and this can be caused by almost any pulmonary disease. The most extreme form of mismatch is pulmonary shunting, which occurs when an area of lung is not ventilated at all – for example, due to collapse or consolidation.
Acute or chronic hypercapnia usually results from alveolar hypoventilation. Causes include:
- central depression of respiratory drive
- impaired nerve transmission between the central nervous system and muscle (especially the diaphragm)
- reduced chest wall movements (including diaphragmatic movements)
- reduced alveolar ventilation due to pathology within the lungs.

The primary respiratory conditions causing acute respiratory failure are detailed in Chapter 19. Critically ill patients may have both Type 1 and 2 respiratory failure at some point, and the pattern and severity can change rapidly. Close monitoring and review are thus essential in order to decide which form of respiratory support is required, as this may change rapidly as the patient deteriorates and/or improves. Both the disease causing the illness and its effect on the patient’s physiology over time must be taken into consideration. A combination of clinical examination and investigation helps to determine the most appropriate interventions (Box 8.10). The best method for assessing hypoxaemia is the ratio of the PaO₂ (measured by blood gas) to the fractional inspired oxygen delivered (FiO₂). This ‘PF’ ratio is lower, the more severe the disease. For example, a patient receiving 60% oxygen with a FiO₂ of 10.0 kPa (75.2 mmHg) on blood gas has a PF ratio of 10.0/0.6 = 16.7 kPa (125.6 mmHg).

### Acute lung injury and the acute respiratory distress syndrome

A range of conditions (Box 8.11) can result in a diffuse acute inflammatory process in the lungs called acute lung injury (ALI); when severe (as defined by hypoxaemia), this is termed the acute respiratory distress syndrome (ARDS; Box 8.12). Inflammation throughout the lungs, affecting both endothelial and epithelial surfaces. Activated neutrophils are sequestered into the lungs and capillary permeability is increased, with damage to type I and II alveolar cells. This results in exudation and accumulation of protein-rich cellular fluid within alveoli and the formation of characteristic ‘hyaline membranes’. Local release of cytokines and chemokines by activated macrophages and neutrophils results in progressive recruitment of inflammatory cells. Secondary effects include loss of surfactant and impaired surfactant production.

The net effect is alveolar collapse and reduced lung compliance, which are most marked in dependent regions of the lung, where airspaces become fluid-filled (Fig. 8.7). The combination of loss of surfactant and fluid accumulation makes these areas difficult to ventilate, which results in hypoxaemia due to ventilation-perfusion mismatch and increased pulmonary shunt. ALI and ARDS can be difficult to distinguish from fluid overload or cardiac failure.
Non-invasive respiratory support

Non-invasive respiratory support includes techniques that do not require sedation or an endotracheal or tracheostomy tube. This helps preserve the patient’s respiratory muscle activity and reduces complications such as nosocomial infection. It can be used to support selected patients with type 1 or 2 respiratory failure, but the patient’s conscious level must be adequate to ensure airway protection from aspiration. Non-invasive respiratory support is classified as continuous positive airway pressure (CPAP) alone or CPAP plus additional support, in the form of pressure applied to the breathing circuit during inspiration (non-invasive ventilation, or NIV).

**CPAP therapy**

CPAP therapy involves the application of a continuous positive airway pressure throughout the patient’s breathing cycle, typically between 5 and 10 cmH₂O. CPAP recruits collapsed alveoli and can enhance clearance of alveolar fluid. It is particularly effective for treating pulmonary atelectasis (which may be post-operative) and pulmonary oedema, and helps correct hypoxaemia in some patients with pneumonia, especially the immunocompromised. CPAP therapy is most effective in correcting hypoxaemia in type 1 respiratory failure, but if it improves pulmonary compliance (by clearing fluid or improving lung volume), it can reduce the work of breathing and improve hypercapnia in type 2. However, many patients with the latter require NIV or invasive ventilation. CPAP therapy can be delivered using tight-fitting facial masks, high-flow nasal cannulae, and hoods (Fig. 8.8). Usually, a CPAP mask is tried first, but different systems can be trialled until the most comfortable for the patient is found. Patients must be cooperative,
able to protect their airway, and have the strength to breathe spontaneously and cough effectively. Failure to improve over 24–48 hours, or a further deterioration in conscious level or blood gases, indicates that invasive ventilation should be considered.

Non-invasive ventilation
Non-invasive ventilation (NIV) is ventilatory support by nasal or full facemask. It can be delivered by a simple bi-level (BiPAP) turbine ventilator, which delivers a higher pressure (approximately 15–25 cmH₂O) for inspiration and a lower pressure (4–10 cmH₂O) to allow expiration. If hypoxaemia is severe, a complex ICU ventilator is employed that allows higher oxygen concentrations to be administered. A simple breathing circuit with a leak rather than an expiratory valve is generally used, and ventilation can be spontaneous (triggered by the patient’s breaths) or timed (occurring at set intervals and/or frequency). Systems that synchronise with the patient’s efforts are better tolerated and more effective. NIV is the first-line therapy in patients with type 2 respiratory failure secondary to acute exacerbation of COPD because it reduces the work of breathing and offloads the diaphragm, allowing it to recover strength. It should be initiated early, especially when severe respiratory acidosis and/or decreased consciousness secondary to hypercapnia are present. Unless there is an improvement in acidosis within 4–6 hours, invasive ventilation is indicated. NIV can also be used to support selected patients with hypercapnia secondary to pulmonary oedema or pneumonia, or during weaning from invasive ventilation, but its effectiveness in these conditions is less certain. As with mask CPAP, NIV requires the patient to be conscious and cooperative.

Emergency endotracheal intubation and mechanical ventilation
Many patients admitted to ICU require endotracheal intubation and mechanical ventilation, mostly for respiratory failure (Boxes 8.13 and 8.14). The final decision to undertake these is based on clinical judgement rather than the results of ABGs in isolation. If possible, the patient’s relatives should be given the chance to visit prior to anaesthesia and intubation, as this may be the last opportunity they have to speak together.

In the conscious patient, intubation requires induction of anaesthesia and muscle relaxation, while in more obtunded patients, sedation alone may be adequate. Intubation can be hazardous in the critically ill patient, particularly if there is associated cardiovascular failure. Patients should be pre-oxygenated and cricoid pressure applied, with continuous monitoring of heart rate, ECG and BP (preferably invasively), together with capnography (and subsequently a chest X-ray) to confirm correct endotracheal tube placement. Complications are common, and intubation is ideally performed in a critical care environment, or with expert assistance, resuscitation facilities and appropriate medication immediately available. Hypotension may follow sedation or anaesthesia due to the direct cardiovascular effects of the anaesthetic agent and loss of sympathetic drive. Positive pressure ventilation may compound this by increasing intrathoracic pressure, thereby reducing venous return and thus cardiac output.

Tracheostomy is usually performed electively when endotracheal intubation is likely to be required for more than 7–10 days (Box 8.15). The timing is determined by individual patient factors and clinical judgement. Tracheostomy is usually carried out percutaneously in the ICU, to avoid transfer to an operating theatre. The passage of a smaller (4.5 mm internal diameter) ‘mini-tracheostomy’ tube through the cricothyroid membrane is a useful technique for clearing airway secretions in spontaneously breathing patients with a poor cough effort, particularly in the HDU and in post-operative patients.

General considerations in the management of the ventilated/intubated patient
Modern ventilators allow enormous flexibility in the way ventilator support is provided. The terminology used to describe ventilation modes can be confusing,
8.15 Advantages and disadvantages of tracheostomy

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Patient comfort</td>
<td>• Immediate complications: hypoxia, haemorrhage</td>
</tr>
<tr>
<td>• Improved oral hygiene</td>
<td>• Tracheal damage; late stenosis</td>
</tr>
<tr>
<td>• Access for tracheal toilet</td>
<td></td>
</tr>
<tr>
<td>• Enables speech with cuff deflated and a speaking valve attached</td>
<td></td>
</tr>
<tr>
<td>• Earlier weaning and ICU discharge</td>
<td></td>
</tr>
<tr>
<td>• Reduced sedation requirement</td>
<td></td>
</tr>
<tr>
<td>• Reduces vocal cord damage</td>
<td></td>
</tr>
</tbody>
</table>

Mechanical ventilation

**Invasive** (via ET or tracheostomy tube)

- Full support
- IPPV
- CMV
- PSV
- SIMV
- NIV ‘BiPAP’
- Cuirass
- Tank: ‘iron lung’
- Rocking bed

**Non-invasive**

- Pressure control (BiPAP)
- +ve pressure (via face or nasal mask)
- -ve pressure
- Partial support
- ‘BiPAP’
- ‘Iron lung’

**Fig. 8.9** Types of invasive and non-invasive ventilatory support. (BiPAP = bi-level positive airway pressure; CMV = controlled mandatory ventilation; ET = endotracheal; IPPV = intermittent positive pressure ventilation; NIV = non-invasive ventilation; PSV = pressure support ventilation; SIMV = synchronised intermittent mandatory ventilation)

because of subtle differences between modes, and the use of different names by different manufacturers. Figure 8.9 gives a classification of the different types of invasive ventilation support, and Box 8.16 outlines the terminology used and the parameters that are set on any ventilator. The inspired gas should always be humidified and warmed, usually achieved with a heat and moisture exchanger, but occasionally with hot water humidification systems.

**Initial settings**

Following intubation, the ventilator is set to deliver a safe mandatory mode of ventilation that will achieve oxygenation and carbon dioxide clearance in the majority of patients. Regular re-assessment of the patient’s parameters will show if modification is required. Hypoxia is avoided, as it is associated with adverse outcomes.

8.16 Ventilator parameter settings on initiating mechanical ventilation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>FiO₂</td>
<td>Initially set to achieve SpO₂ &gt; 92% based on pulse oximetry. Subsequently adjusted to achieve PaO₂ of 8–12 kPa (60.2–90.2 mmHg) on ABGs.</td>
</tr>
<tr>
<td>Positive end expiratory pressure (PEEP)</td>
<td>The pressure maintained by the ventilator at the end of expiration: important for recruiting and maintaining alveoli for gas exchange. Initially set at ~5 cmH₂O, except in patients with severe gas trapping (e.g. acute asthma), but increased in patients with severe hypoxaemia.</td>
</tr>
<tr>
<td>Tidal volume</td>
<td>Set low in ARDS to minimise pulmonary barotrauma and volutrauma. In patients with normal lungs, higher volumes are safe.</td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>Typically set at 12–15 breaths/min. Adjusted according to PaCO₂ on ABGs to achieve adequate CO₂ clearance.</td>
</tr>
<tr>
<td>Mode of ventilation</td>
<td>The safest initial mode is SIMV (see text): synchronises breaths with efforts made by patient, but ventilates at the set rate and tidal volume in the absence of respiratory effort.</td>
</tr>
<tr>
<td>Breath trigger</td>
<td>Initiates a breath in response to patient’s efforts, triggered by flow towards the patient on attempts to breathe. Set at the appropriate sensitivity for the patient.</td>
</tr>
<tr>
<td>Inspired to expired time ratio (I–E ratio)</td>
<td>In most patients, the I–E ratio should be 1:2 to 1:3. Sometimes altered in expiratory airway obstruction (e.g. lengthened in asthma) or difficult oxygenation (e.g. shortened or reversed in ARDS).</td>
</tr>
<tr>
<td>Alarms</td>
<td>Disconnection. Low or high tidal volumes. Low FiO₂. Excessively high or low airway pressures. A range of other parameters can also be selected. Adjust default settings to optimise safety for the individual patient.</td>
</tr>
</tbody>
</table>

**Mandatory modes of ventilation**

**Volume-controlled modes.** These are set to deliver a preset tidal volume at a set frequency to guarantee a specified minute ventilation. Synchronised intermittent mandatory ventilation (SIMV) is the most widely used, which also synchronises breaths with any efforts made by the patient. Volume-controlled modes will deliver the set volume, but the pressures generated in the patient’s lungs can be excessively high if the pulmonary compliance is low – for example, in ARDS – thus precipitating lung barotrauma or volutrauma, which could cause pneumothorax.

**Pressure-controlled modes.** These deliver a set pressure for a specified duration. Pressure-controlled ventilation (PCV) and bi-level positive airway pressure ventilation (BiPAP) are examples. The tidal and minute volumes achieved are determined by the pulmonary compliance; in patients with stiff lungs, only small tidal volumes may
be achieved, whereas in patients with normal compliance, excessive volumes may result. An advantage of these modes is that airflow pressures are controlled, but the effect on blood gases needs to be regularly assessed to identify changes in pulmonary compliance. Tidal volume is a useful safety alarm as it falls with increased resistance, such as with bronchospasm. In difficult-to-ventilate cases, several modes can be attempted sequentially to identify which is most effective.

**Weaning or spontaneously breathing modes.** Most modern ventilators can detect whether a patient is making breathing efforts, and use a flow trigger to augment each breath. Assisting breathing with additional pressure is more comfortable than fixed tidal volumes and allows sedation to be reduced. The most common mode applies additional pressure during inspiration, assisting the work of breathing and increasing tidal volume. The pressure is removed when the ventilator detects an expiratory effort. These modes are usually called pressure support ventilation (PSV) or assisted spontaneous breathing (ASB).

**Mixed modes.** Different modes can be applied simultaneously, tailored to meet individual requirements. For example, it is possible to specify a frequency of SIMV or PCV breaths, but also provide PSV for any additional efforts the patient makes.

**Advanced ventilation strategies**

In patients with severe acute lung disease, especially those resulting in reduced lung compliance such as ARDS, the ventilator can worsen lung injury as a result of overstretch and shearing forces in parts of the lung that continually open and collapse. The aim of advanced ventilation is to minimise further damage due to pressure (barotrauma), volume stretch (volutrauma), and the additional inflammatory mediators released into the body from ongoing lung injury (biotrauma) (Box 8.17). Tidal volumes and airway pressures are kept as low as possible while achieving adequate oxygenation.

In many cases, it is best to accept hypercapnia rather than apply higher pressure to clear CO₂. Often, higher levels of positive end expiratory pressure (PEEP) are required to achieve adequate alveolar recruitment and oxygenation.

Several other strategies can be used.

---

**EBM 8.17 Mechanical ventilation in ARDS**

- Low tidal volumes (4–6 mL/kg ideal body weight) reduce mortality from ARDS.¹
- High levels of PEEP should be avoided in patients with less severe ARDS, in whom it may be harmful. In severe ARDS, high level of PEEP to recruit the lungs and improve oxygenation may decrease mortality.²
- Early use of neuromuscular relaxing drugs to facilitate mechanical ventilation reduces mortality in severe ARDS.³
- Minimised fluid administration and accumulation reduce the duration of mechanical ventilation in ARDS.⁴

(PEEP = positive end-expiratory pressure)

---

²Briel M, et al. JAMA 2010; 303:865–873.

---

**Prone ventilation.** Oxygenation will often improve in patients turned on their front, as a result of improved lung recruitment and better ventilation–perfusion matching. Prone ventilation has not reduced mortality in controlled trials, but is a useful ‘rescue therapy’ in cases where oxygenation is difficult.

**High-frequency oscillatory ventilation (HFOV).** This uses a specialised ventilator to provide gas exchange with high-frequency oscillating gas movements (> 150/min). Conventional breaths and tidal volumes are not set, but effective oxygenation and CO₂ clearance are achieved by adjusting the frequency and power of the oscillations, and the mean airway pressure.

**Nitric oxide.** Nitric oxide is a very short-acting pulmonary vasodilator. When delivered to the airway, it improves blood flow to ventilated alveoli, thus improving ventilation-perfusion matching. Oxygenation can be improved markedly in some patients but there is evidence that this lasts for only 48 hours, and rebound effects can occur when it is withdrawn. No improvement in mortality has been shown in controlled trials. Its role is limited to rescue therapy when other interventions have failed, and it may be useful in patients with severe pulmonary hypertension.

**Extracorporeal membrane oxygenation therapy (ECMO).** ECMO involves connecting the patient to an external bypass circuit. Oxygenation and CO₂ clearance are achieved using a membrane oxygenator. The patient’s lungs are usually ‘rested’ and ventilation reduced to low levels. Advances in technology have dramatically improved the safety of these devices, although their use is restricted to specialised centres. Controlled trials indicate improved survival in appropriately selected cases, and patients with severe ARDS should be considered for treatment.

**Corticosteroids.** There is conflicting evidence regarding the use of steroids as anti-inflammatory agents in acute lung injury. Uncertainty remains about patient selection and the timing of therapy, and use of corticosteroids may be complicated by secondary infection and muscle weakness. However, they are often tried after 7–10 days of ARDS, if the patient remains severely unwell.

**Weaning from respiratory support**

Patients usually require most mechanical ventilation in the period following intubation when they are most unwell, following which support is gradually reduced as the underlying condition resolves and the patient is able to breathe with less assistance. This is the process of ‘weaning’ from ventilation. Sufficient support is provided to correct hypoxaemia and hypercapnia, but the level is decreased as quickly as possible to reduce the chance of secondary complications, such as infection and muscle weakness. Rapid weaning, often with reduction in sedation levels (see below), shortens length of ICU stay and improves patient outcomes. Patients who have required long-term ventilatory support for severe lung disease such as ARDS may be unable to sustain even a modest degree of respiratory work initially because of poor lung compliance, high work of breathing and respiratory muscle weakness. They require more prolonged weaning, until respiratory muscle strength improves.

Several criteria can be used to assess whether a patient is ready to start reducing respiratory support (Box 8.18). Approaches include:
Spontaneous breathing trials (SBTs). These involve removing all respiratory support, typically on a daily basis, and observing how long the patient is able to breathe unassisted. This is particularly effective when linked to sedation breaks. Signs of failure include rapid shallow breathing, hypoxaemia, rising PaCO₂, sweating and agitation. Patients who pass an SBT are assessed for extubation.

**Progressive reduction in pressure support ventilation.** Progressive reduction in the PSV is applied for each breath over a period of hours or days, according to patient response. When patients are strong enough to breathe over a period of hours or days, according to the severity of illness.

**Progressive reduction in pressure support ventilation.** Progressive reduction in the PSV is applied for each breath over a period of hours or days, according to patient response. When patients are strong enough to breathe over a period of hours or days, according to the severity of illness.

**Spontaneous breathing trials (SBTs).** These involve removing all respiratory support, typically on a daily basis, and observing how long the patient is able to breathe unassisted. This is particularly effective when linked to sedation breaks. Signs of failure include rapid shallow breathing, hypoxaemia, rising PaCO₂, sweating and agitation. Patients who pass an SBT are assessed for extubation.

**Weaning protocols.** The process of weaning is best undertaken as a continuous process. Protocols that empower nursing staff to initiate and progress weaning within agreed guidelines reduce ventilation times. Patients requiring prolonged mechanical ventilation typically require individualised weaning plans, with regular periods of training followed by rest, to enable respiratory muscles to regain strength.

The timing of extubation relies on clinical judgement. Patients must have stable ABGs with resolution of hypoxaemia and hypercapnia despite withdrawal of ventilator support. Conscious level must be adequate to protect the airway, comply with physiotherapy, and cough. The need for re-intubation following extubation is associated with poorer outcomes.

**Acute kidney injury**

Acute kidney injury (AKI) is defined as an abrupt and sustained decrease in kidney function (Box 8.19). AKI in the critically ill patient is often due to pre-renal problems such as hypovolaemia, hypotension and ischaemia resulting in reduced renal DO₂. However, it may also be due to acute tubular necrosis (ATN, p. 479), which may result from ischaemia, or nephrotoxicity caused by chemical or bacterial toxins, or a combination of these. Potentially nephrotoxic drugs include non-steroidal anti-inflammatory drugs (NSAIDs), angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor antagonists, radiological contrast media and some antibiotics.

Oliguria (< 0.5 mL/kg/hr for several hours) is an important early sign of systemic problems in critical illness. It requires investigation and early intervention to correct hypoxaemia, hypovolaemia, hypotension and renal hypoperfusion. Successful resuscitation is associated with restoration of good urine output, an improving acid–base balance and correction of plasma potassium, urea and creatinine.

Oliguria is an integral part of the normal stress response to major surgery, and care should be taken not to overfill the post-operative patient who has oliguria but is otherwise well from a cardiovascular and biochemical point of view.

**Renal support**

Sepsis is frequently implicated in the development of AKI, and the source must be promptly identified and adequately treated. Obstruction of the renal tract (including catheter blockage) should always be excluded and is most easily identified with abdominal ultrasound. It must be relieved at once. Acute glomerulonephritis and vasculitis must also be considered, and appropriate specialist referral, with investigations such as urine microscopy and immunopathological tests (p. 480), carried out early.

The mainstay of management is aggressive haemodynamic resuscitation to achieve normovolaemia, normotension and an appropriate cardiac output. There is little evidence that specific treatments aimed at inducing a diuresis, such as low-dose dopamine, furosemide or mannitol, have any renoprotective action or other benefit in restoring renal function.

If renal function cannot be restored following resuscitation, renal replacement therapy (p. 488) is indicated (Box 8.20). The preferred renal replacement therapy in ICU patients is pumped venovenous haemofiltration. This is associated with fewer osmotic fluid shifts and hence greater haemodynamic stability than haemodialysis. It is carried out using a double-lumen central venous catheter placed percutaneously. Haemofiltration should be initiated early in the septic patient.
be continuous in the early phase of treatment. Intermittent treatment may be used when the patient is recovering from the primary insult and return of normal renal function is expected. Provided the precipitating cause can be successfully treated, renal failure due to ATN usually recovers between 5 days and several weeks later.

Survival rates from MOF, including AKI, have been around 50% for many years, but modern haemofiltration techniques are being shown to produce better outcomes.

Gastrointestinal and hepatic disturbance

Gastrointestinal symptoms, such as nausea, vomiting and large nasogastric aspirates, may be the earliest signs of regional circulatory failure, and when associated with a tender, distended, silent abdomen, indicate that this is the probable site of the primary pathology. The gut has a rapid cell turnover rate and fasting alone can produce marked changes in mucosal structure and function. In hypovolaemia and frank shock states, splanchnic vasoconstriction produces gut mucosal ischaemia, damaging the mucosal barrier and allowing toxins to enter the portal circulation and lymphatics. Splanchnic ischaemia may contribute to the progression of MOF, possibly as a source of bacteraemia or systemic inflammation. Manifestations of MOF within the gastrointestinal tract include loss of gastric acid production, erosive gastritis, stress ulceration, bleeding, ischaemia, pancreatitis and acalculous cholecystitis. These occur less frequently when adequate circulatory resuscitation occurs early. Ischaemic bowel is difficult to diagnose in the critically ill patient, but in the context of otherwise unexplained lactic acidosis, hyperkalaemia and coagulopathy, abdominal imaging by contrast-enhanced computed tomography (CT) and laparotomy should be considered.

Three distinctive hepatic dysfunction syndromes can occur in the critically ill:

- **Shock liver or ischaemic hepatitis** results from extreme hepatic tissue hypoxia and is characterised by centrilobular hepatocellular necrosis. Transaminase levels are often massively raised (> 1000–5000 U/L) at an early stage, followed by moderate hyperbilirubinaemia (< 100 µmol/L or < 5.8 mg/dL). There is often associated hypoglycaemia, coagulopathy and lactic acidosis. Following successful resuscitation, hepatic function generally returns to normal.

- **Hyperbilirubinaemia (‘ICU jaundice’)** frequently develops following trauma or sepsis, particularly if there is inadequate control of the inflammatory process. There is a marked rise in bilirubin (predominantly conjugated), but only mild elevation of transaminase and alkaline phosphatase. This results from failure of bilirubin transport within the liver and produces the histological appearance of intrahepatic cholestasis. Extrahepatic cholestasis must be excluded by abdominal ultrasound and potentially hepatotoxic drugs should be stopped. Treatment is non-specific and should include early institution of enteral feeding. Therapy that compromises splanchnic blood flow, particularly high doses of vasoconstrictor agents, should be avoided.

- **Transaminitis** is most commonly due to drug toxicity: for example, antibiotics.

### Gastrointestinal and hepatic support

Early institution of enteral nutrition is the most effective strategy for protecting the gut mucosa and providing nutritional support. The optimum use of enteral nutrition involves simple protocols that initiate nutrition as early as possible, and progressively increase feeding volumes until nutritional targets are met. Current evidence supports early enteral nutrition using standard feeds, with the addition of prokinetic agents such as metoclopramide or low-dose erythromycin when gastric aspirates are high. The evidence for early supplementation with total parenteral nutrition (TPN) is weak, and it is not routinely indicated until enteral feeding attempts have been unsuccessful for approximately 7 days. Nutritional support is further considered on page 122.

Hyperglycaemia is common during critical illness and is associated with poor outcomes. Tight glycaemic control, using insulin infusions, has been studied in several controlled trials in the critically ill. The benefit was greatest in surgical patients at low risk of death, but the risk-to-benefit profile in the mixed critically ill population is uncertain. Inadvertent hypoglycaemia is associated with adverse patient outcomes. In most ICUs, the current target is for modestly elevated blood glucose concentrations: for example, 5.5–8 mmol/L (100–144 mg/dL).

Stress ulcer prophylaxis is best achieved with H₂-receptor antagonists (e.g. ranitidine), which are both safe and effective. Although stress ulcer bleeding is rare with modern resuscitation, evidence supports routine use in mechanically ventilated patients and those with renal failure or coagulopathy. H₂-receptor antagonists are associated with an increased incidence of nosocomial pneumonia, and treatment should be stopped following extubation in the absence of other indications. Withdrawal can also be considered when full enteral nutrition has been established, unless the patient has a history of peptic ulcer disease. Proton pump inhibitors are only required in upper gastrointestinal bleeding due to ulceration, and they should also be continued when the patient has been taking them long-term.

The management of liver failure is discussed on page 934.

### Neurological failure (coma)

Impaired consciousness or coma is often an early feature of severe systemic illness (Box 8.21). Prompt assessment of consciousness level and management of airway, breathing and circulation are essential to prevent further brain injury, to allow diagnosis and to permit definitive treatment to be instituted. Any patient with confusion or reduced conscious level should have blood sugar measured and hypoglycaemia corrected.

Impairment of conscious level is graded using the Glasgow Coma Scale (GCS, p. 1160), which is also used to monitor progress. A targeted neurological examination is very important. Pupil size and reaction to light,
Neurological support

A diverse range of neurological conditions require management in the ICU. These include not only the various causes of coma, but also spinal cord injury, peripheral neuromuscular disease and prolonged seizures. The goals are to:

- protect the airway, if necessary by endotracheal intubation
- provide respiratory support to correct hypoxaemia and hypercapnia
- treat circulatory problems, e.g. neurogenic pulmonary oedema in subarachnoid haemorrhage, autonomic disturbances in Guillain–Barré syndrome, and spinal shock following high spinal cord injuries
- manage acute brain injury with control of raised intracranial pressure (ICP)
- manage status epilepticus using anaesthetic agents such as thiopental or propofol.

The aim of management in acute brain injury is to optimise cerebral oxygen delivery by maintaining a normal arterial oxygen content and a cerebral perfusion pressure of more than 60 mmHg. Avoiding secondary insults to the brain, such as hypoxaemia and hypotension, improves outcome. ICP rises in acute brain injury as a result of haematomata, contusions, oedema or ischaemic swelling. Raised ICP causes direct damage to the cerebral cortex and, as a result of downward pressure on the brainstem, indirect damage by reducing cerebral perfusion pressure, thereby threatening cerebral blood flow and oxygen delivery:

**Cerebral perfusion pressure (CPP)**

\[
\text{CPP} = \text{mean BP} – \text{ICP}
\]

ICP is measured by pressure transducers that are inserted directly into the brain tissue. The normal upper limit for ICP is 15 mmHg and management should be directed at keeping it below 20 mmHg (Box 8.22). Sustained pressures of more than 30 mmHg are associated with a poor prognosis.

CPP should be maintained above 60 mmHg by ensuring adequate fluid replacement and, if necessary, by treating hypotension with a vasopressor such as noradrenaline (norepinephrine).

Complex neurological monitoring must be combined with frequent clinical assessment of GCS, pupil response to light, and focal neurological signs. The motor response to pain is an important prognostic sign. No response or extension of the upper limbs is associated with severe injury, and unless there is improvement within a few days, prognosis is very poor. A flexor response is encouraging and indicates that a good outcome is still possible.

**Neurological complications in intensive care**

Neurological complications may occur as a result of systemic critical illness. Sepsis may be associated both with an encephalopathy characterised by delirium, and with cerebral oedema and loss of vasoregulation. Hypotension and coagulopathy may provoke cerebral infarction or haemorrhage. Neurological examination is very
**Sepsis**

Sepsis can occur in many clinical situations. It may be due to a primary infection (e.g. pneumonia) or it may be the result of clinical interventions for other conditions (e.g. immunosuppressive drugs, chemotherapy, invasive lines). Patients who are in hospital are at increased risk of certain specific infections, such as meticillin-resistant *Staphylococcus aureus* (MRSA). Sepsis usually originates from a localised infection that progresses to an uncontrolled systemic response. It can rapidly lead to acute physiological deterioration with the risk of MOF and death. Early identification of sepsis and appropriate intervention with oxygen, fluids, antibiotics, and more advanced resuscitation where indicated, has been shown to improve survival.

The incidence of sepsis is thought to be increasing, possibly as a result of a growing elderly population, increased use of invasive surgery, higher bacterial resistance, and greater numbers of immunocompromised patients. Important comorbidities and risk factors for sepsis are shown in Box 8.23. These conditions not only increase risk of development of sepsis but also can exaggerate the severity of the process. However, sepsis can affect healthy people at any age.

Box 8.24 gives the common sites of infection in critically ill patients and appropriate investigations to consider. Any pathogen, including aerobic Gram-positive and negative bacteria, anaerobes and fungi, may cause sepsis but in nearly 45% of cases microbiological confirmation of the organism is lacking.

Any or all of the features of SIRS (see Box 8.2, p. 184) may be present, together with an obvious focus of infection, such as purulent sputum with chest X-ray shadowing, or erythema around an intravenous line. However, severe sepsis may present with unexplained hypotension (i.e. septic shock), and the speed of onset may mimic a major pulmonary embolus or myocardial infarction.

Nosocomial infections are an increasing problem in critical care units. Risk factors are similar to those in Box 8.23, but also include prolonged ICU stay, invasive ventilation and stress ulcer prophylaxis with H₂ antagonists. Cross-infection is a major concern, particularly with MRSA, multidrug-resistant Gram-negative organisms and *Clostridium difficile*. If cross-infection occurs frequently, it should prompt a review of the unit’s infection control practices.

---

**8.24 Sites of infection in critically ill patients**

<table>
<thead>
<tr>
<th>Sites of infection</th>
<th>Investigations and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major</strong></td>
<td></td>
</tr>
<tr>
<td>Intravenous lines</td>
<td>If the patient develops sepsis, replace any lines that have not been changed for &gt; 4 days</td>
</tr>
<tr>
<td>(particularly central)</td>
<td></td>
</tr>
<tr>
<td>Lungs</td>
<td>High risk of nosocomial pneumonia in intubated patients. After ICU stay &gt; 3–4 days, particularly if antibiotics are given, the nasopharynx becomes colonised with Gram-negative bacteria, which migrate to the lower respiratory tract. Prophylaxis with parenteral and enteral antibiotics (selective decontamination of the digestive tract) reduces the incidence of nosocomial pneumonia</td>
</tr>
<tr>
<td>Abdomen</td>
<td>Consider intra-abdominal abscess or necrotic gut in patients who have had abdominal surgery. Pancreatitis, acute cholecystitis or perforated peptic ulcer may develop as a complication of critical illness. Ultrasound, CT, aspiration of collections of fluid/pus and laparotomy may be required</td>
</tr>
<tr>
<td>Urinary tract</td>
<td>Urine culture (but this is a relatively unusual source in unexplained sepsis)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td>Heart valves</td>
<td>Transthoracic or transoesophageal echocardiogram</td>
</tr>
<tr>
<td>Meninges</td>
<td>Lumbar puncture after checking coagulation and platelet count</td>
</tr>
<tr>
<td>Joints and bones</td>
<td>X-ray, gallium or technetium white cell scan</td>
</tr>
<tr>
<td>Nasal sinuses, ears, parapharyngeal space</td>
<td>Clinical examination, plain X-ray, CT</td>
</tr>
<tr>
<td>Genitourinary tract (particularly post-partum)</td>
<td>PV examination, ultrasound</td>
</tr>
<tr>
<td>Gastrointestinal tract</td>
<td>PR examination, stool culture, <em>Clostridium difficile</em> toxin, sigmoidoscopy</td>
</tr>
</tbody>
</table>
control policies (p. 145). Limiting antibiotic use helps to prevent the emergence of multidrug-resistant bacteria.

Management

Prompt resuscitation, with early cultures, administration of appropriate antibiotics and eradication of the source of infection (if necessary by surgical drainage), is required (Box 8.25). Antibiotics should have a spectrum wide enough to cover probable causative organisms, based on the likely site of infection, whether community-acquired or nosocomial, previous antibiotic therapy and known local resistance patterns.

Other investigations required include:
- cultures of sputum, intravascular lines, urine and any wound discharge
- ABGs and coagulation profile
- urinalysis and chest X-ray.

Only 10% of ICU patients with a clinical diagnosis of ‘septic’ shock will have positive blood cultures, due to prior antibiotic treatment and the fact that an inflammatory state is not always due to infection. More specific investigations are driven by the history and examination (see Box 8.24).

The haemodynamic changes in septic shock are variable and not specific for the Gram status of the infecting organism. The first feature is often tachypnoea and the early stages are frequently dominated by hypotension with relative volume depletion due to vasodilatation. Sufficient intravenous fluid should be given to ensure that the intravascular volume is not the limiting factor in determining global oxygen delivery. The type of fluid that should be administered and what constitutes ‘adequate’ volume resuscitation remain controversial. The response to therapy is crucial when deciding what constitutes ‘adequate’ volume resuscitation remain controversial. The response to therapy is crucial when deciding this and is frequently unpredictable, so rigid protocols cannot be used. Depending on haemoglobin concentration, blood or synthetic colloid should be given as 100–200 mL boluses to assess BP response to volume (see Fig. 8.2, p. 183).

Although ventricular function is frequently impaired, the characteristically low systemic vascular resistance (SVR) usually ensures a high cardiac output (once the patient is adequately volume-resuscitated), albeit with low BP.

The choice of the most appropriate vasoactive drug to use should be based on a full assessment of the circulation and the different inotropic, dilating or constricting properties of these drugs (see Box 8.8, p. 191). In most cases, a vasoconstrictor such as noradrenaline (norepinephrine) is necessary to increase SVR and BP, while an inotrope (dobutamine) may be necessary to maintain cardiac output. In the later stages of severe sepsis, the fundamental problem is at the microcirculatory level. Oxygen uptake and utilisation are impaired due to failure of the regional distribution of flow and direct cellular toxicity despite adequate global oxygen delivery. Tissue oxygenation may be improved and aerobic metabolism sustained by reducing demand, i.e. metabolic rate (Box 8.26). This can be achieved with sedatives and muscle relaxants (see below).

Corticosteroids

Assessment of the pituitary–adrenal axis is difficult in the critically ill but in some series up to 30% of patients have adrenal insufficiency, as assessed by a short Synacthen test (p. 778). Corticosteroid replacement therapy is controversial. Recent evidence suggests that, although it is associated with earlier resolution of shock, it has no effect on survival.

Disseminated intravascular coagulation

Also known as consumptive coagulopathy, disseminated intravascular coagulation (DIC) is an acquired disorder of haemostasis (p. 1055); it is common in critically ill patients and often heralds the onset of MOF. It is characterised by an increase in prothrombin time, partial thromboplastin time and fibrin degradation products, and a fall in platelets and fibrinogen. The clinically dominant feature may be widespread bleeding from vascular access points, gastrointestinal tract, bronchial tree and surgical wound sites, or widespread microvascular and even macrovascular thrombosis. Management is supportive with infusions of fresh frozen plasma and platelets, while the underlying cause is treated.

General principles of critical care management

Essential aspects of the management of critically ill patients on admission to the ICU are shown in Box 8.27.
**Daily clinical management in the ICU**

Regular clinical examination is essential to identify any changes in a patient’s condition. Detailed clinical examination is performed at least daily, with additional focused and systematic assessment on ward rounds at least twice daily. Ward rounds are also an opportunity to ensure the reliable application of evidence-based measures to reduce complications; the mnemonic FAST HUG provides a useful checklist of feeding, analgesia, sedation, thromboprophylaxis, head of bed elevation, ulcer prophylaxis, glucose control.

Patient review should include:

- Review of progress reports from nursing and medical staff, and any specialist opinions.
- Review of 24-hour charts.
- Examination: general (including skin, line sites, wounds etc.).
- System reviews:
  - **Cardiovascular**: haemodynamics, fluids and inotropes
  - **Respiratory**: ventilator settings and ABGs
  - **Gastrointestinal**: nutrition (calorie, protein intake, route), nasogastric aspirate and bowel function
  - **Renal**: urine output, overall fluid balance, urea and electrolytes, and renal replacement therapy
  - **Neurological**: sedation level, GCS and pupil responses.
- Laboratory results: haematology, coagulation and biochemistry.
- Microbiology: temperature, white blood count, line sites and other possible sources of infection, results of cultures, antibiotic therapy.
- Drug therapy: review with pharmacist, consider adverse effects and interactions, and identify drugs that can be discontinued. Medicines required for long-term conditions should be continued in the context of the acute illness. An accurate record of the patient’s usual medicines must be obtained.
- Imaging: review X-rays and other specialist investigations with radiologists.
- Monitoring: are all measures still required? In particular, remove central venous catheters, arterial lines and peripheral venous catheters as soon as no longer needed, in order to avoid infection.
- Management plan: formulate an integrated plan, with specific goals for each organ system and goals for the patient, e.g. mobilising out of bed. Involve the family in the patient’s care.

**Sedation and analgesia**

Intensive care is an extremely stressful experience for the patient, with pain, discomfort and anxiety related to endotracheal intubation, invasive monitoring and other procedures.

Most patients require sedation and analgesia to ensure comfort, relieve anxiety, and allow tolerance of an endotracheal tube, mechanical ventilation and invasive procedures. Some conditions, especially severe neurological conditions that cause brain swelling and raised ICP, require deep sedation to reduce tissue oxygen requirements and protect organs from ischaemic damage. Excessive sedation is common in ICU patients, and is associated with longer ICU stays, a higher prevalence of delirium, prolonged requirement for mechanical ventilation, and more ICU-acquired infections. The optimally sedated patient is comfortable and tolerates treatments, but is awake and lucid.

Sedation and analgesia are usually provided via continuous infusions of sedative and/or analgesic drugs. As many critically ill patients have impaired liver and renal function, the potential for drugs to accumulate is high and the patient’s sedation must be regularly monitored. The sedative agents used ideally have predictable short half-lives that are not affected by liver or renal impairment. Short-acting intravenous agents, such as propofol, are usually employed. Analgesia can be provided using morphine infusions, but in patients with MOF, especially renal failure, active metabolites can accumulate. Opiates such as fentanyl or alfentanil, which are not renally metabolised or excreted, are commonly chosen.

Sedation is monitored via clinical sedation scales (Box 8.28) that record responses to voice and physical stimulation. Regular use of these to adjust sedation is associated with a shorter ICU stay. Many ICUs also have a daily ‘sedation break’, when all sedation is stopped in appropriate cases for a period, in order to re-assess the patient. This approach reduces the chance of oversedation and shortens ICU stay.

### Box 8.28 Richmond Agitation Sedation Scale (RASS)

<table>
<thead>
<tr>
<th>Score</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4</td>
<td>Coma</td>
<td>Overtly combative, or violent or immediate danger to staff</td>
</tr>
<tr>
<td>+3</td>
<td>Very agitated</td>
<td>Pulls on/removes tubes or catheters, or aggressive to staff</td>
</tr>
<tr>
<td>+2</td>
<td>Agitated</td>
<td>Frequent non-purposeful movement or patient–ventilator dysynchrony</td>
</tr>
<tr>
<td>+1</td>
<td>Restless</td>
<td>Anxious or apprehensive but no aggressive or vigorous movements</td>
</tr>
<tr>
<td>0</td>
<td>Alert and calm</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>Drowsy</td>
<td>Not fully alert but sustained awakening (&gt;10 secs) with eye opening/contact to voice</td>
</tr>
<tr>
<td>-2</td>
<td>Light sedation</td>
<td>Brief awakening (&lt;10 secs) with eye contact to voice</td>
</tr>
<tr>
<td>-3</td>
<td>Moderate sedation</td>
<td>Movement but no eye contact to voice</td>
</tr>
<tr>
<td>-4</td>
<td>Deep sedation</td>
<td>Movement to physical stimulation but no response to voice</td>
</tr>
<tr>
<td>-5</td>
<td>Unrousable</td>
<td>No response to voice or physical stimulation</td>
</tr>
</tbody>
</table>

**Muscle relaxants**

Muscle relaxants are avoided whenever possible in ICU patients. Their use is associated with a higher prevalence of critical illness neuropathy and myopathy, resulting in muscle weakness. Sedation use also tends to be higher when muscle relaxants are employed. They are required to facilitate endotracheal intubation and to facilitate ventilation in patients with critical oxygenation and/or poor...
lung compliance. Patients with critically increased ICP often receive intravenous infusions of muscle relaxant drugs such as atracurium, to help control it and to prevent coughing and high intrathoracic pressures, which increase ICP.

**Delirium**

Delirium is extremely common in critically ill patients. It often becomes apparent as sedation is reduced and stopped. About 60–80% of patients have hypoactive delirium, which is often missed unless formal testing is undertaken. Between 5 and 10% of patients have agitated delirium, and 10–20% a mixed pattern. Delirium of any type is associated with poorer outcome. Management is focused on reducing or avoiding precipitating factors, such as benzodiazepines and metabolic disturbances. Patients with agitated delirium should be managed with haloperidol in 2.5 mg increments, rather than additional sedatives. Some sedative drugs are associated with a lower incidence of delirium, such as α-2-adrenergic agonists (clonidine and dexmedetomidine). These have a central sedative action and can be useful in difficult cases. There is no evidence that pharmacological interventions are useful as prophylaxis or in hypoactive delirium. Additional information about diagnosis and management of delirium is given on page 1161.

**DISCHARGE FROM INTENSIVE CARE**

Discharge is appropriate when the original indication for admission has resolved and the patient has sufficient physiological reserve to continue recovery without the facilities of intensive care. Many ICUs and HDUs function as combined units, which allows stepdown to HDU care without a change of clinical team. Critically ill patients often have complex medical histories, multiple ongoing medical problems, and family and social issues. Many also have the emotional problems associated with survival from a life-threatening event or illness. Discharge from the ICU is stressful for patients and families, and communication with the clinical team accepting responsibility is vital. A key issue is that nursing care changes from one-to-one or one-to-two to much lower staffing levels. Discharges from ICU/HDU to standard wards should take place within normal working hours to ensure adequate medical and nursing support and detailed handover, as discharge outside normal working hours (and early discharge) are associated with higher ICU re-admission rates and increased mortality.

The receiving team should be provided with a written summary, including relevant recent investigations, and the critical care team should remain available for advice. Many ICU teams provide an outreach service to furnish advice and ensure continuity. Irrespective of the reason for admission to the ICU, many patients suffer problems (Box 8.29). These can last from weeks to many months and require ongoing support and rehabilitation.

**Withdrawal of intensive support**

Withdrawal of support must be considered when it is clear that the patient has no realistic prospect of recovery or of surviving with a quality of life that he or she would value. In these situations, intensive care will only prolong the dying process and is therefore futile. When intensive support is withdrawn, management remains active and is aimed at allowing the patient to die with dignity and as free from distress as possible (p. 290). Patients’ views are paramount and increasing use is being made of advance directives or ‘living wills’. Communication with the patient, if possible, and with the family, the referring clinicians, and between members of the critical care team is crucial (p. 165). Failure in this area causes stress and unrealistic expectations, damages working relations and leads to subsequent unhappiness, anger and litigation.

**Brainstem death**

Recent advances in the resuscitation and intensive care management of brain-injured patients have invariably increased the survival of patients who remain ventilated on ICU, in whom progression of brain injury results in brainstem death. The preconditions for considering brainstem death and the criteria for establishing the diagnosis are listed on page 1160.

When the formal criteria for brainstem death are met, it is clearly inappropriate to continue supporting life with mechanical ventilation, and the possibility of organ donation should be considered. All intensive care clinicians have a responsibility to approach relatives to seek consent for organ donation, provided there is no contra-indication to their use. This can be very difficult, but is easier if the patient carried an organ donor card or was registered with an organisation such as the UK Organ Donor Register. In the UK, each region has a team of specialist nurses in organ donation who provide help with the process and with care of the potential organ donor.

**8.29 Common problems after ICU discharge**

**Physical**

- Fatigue: almost universal and can last for many weeks
- Breathlessness
- Muscle weakness: a combination of muscle wasting from inadequate nutritional intake, muscle breakdown associated with inflammation/infection, and damage to nerves and muscle associated with severe systemic illness (critical illness polyneuropathy and polymyopathy)
- Altered taste and poor appetite: can result in weight loss
- Joint stiffness
- Itch
- Hair loss

**Psychological**

- Anxiety
- Depression
- Traumatic memories: including delusional memories (often unpleasant), dreams, flashbacks and hyper-arousal. Worse in more severe and prolonged illness, more prolonged deep sedation and after delirium. Recovery occurs over weeks/months, but can persist and affect quality of life

**Brainstem death**

Recent advances in the resuscitation and intensive care management of brain-injured patients have inevitably increased the survival of patients who remain ventilated on ICU, in whom progression of brain injury results in brainstem death. The preconditions for considering brainstem death and the criteria for establishing the diagnosis are listed on page 1160.

When the formal criteria for brainstem death are met, it is clearly inappropriate to continue supporting life with mechanical ventilation, and the possibility of organ donation should be considered. All intensive care clinicians have a responsibility to approach relatives to seek consent for organ donation, provided there is no contra-indication to their use. This can be very difficult, but is easier if the patient carried an organ donor card or was registered with an organisation such as the UK Organ Donor Register. In the UK, each region has a team of specialist nurses in organ donation who provide help with the process and with care of the potential organ donor.
OUTCOME OF INTENSIVE CARE

The measure used most widely to assess outcome from intensive care is mortality. Mortality is strongly influenced by the case mix of an ICU (the type of patients and their illness severity at admission). Typically, about 20% of ICU patients will die during their ICU stay despite treatment, and about 30% will die before leaving hospital. Some factors associated with higher mortality are shown in Box 8.30. Mortality in patients requiring HDU care is much lower. Many patients have pre-existing illnesses prior to ICU admission, which, combined with the effect of the illness that resulted in ICU admission and the subsequent complications, means that many who survive the ICU admission have reduced life expectancy compared to those of similar age in the general population. The long-term physical and psychological effects of critical illness can mean that surviving patients have a reduced quality of life for many months or years. Families often carry a heavy burden of care after critical illness. Many patients do not regain pre-illness health and may be unable to work, resulting in economic and social hardship.

Scoring systems

Admission and discharge criteria vary between ICUs, so it is important to define the characteristics of the patients admitted (case mix) in order to assess the effects of the care provided on the outcome achieved. Two systems are widely used to measure severity of illness:

- ‘APACHE’ II: Acute Physiology Assessment and Chronic Health Evaluation
- ‘SAPS’ 2: Simplified Acute Physiology Score.

These scores include assessment of certain admission characteristics (e.g. age and pre-existing organ dysfunction) and a variety of routine physiological measurements (e.g. temperature, BP, GCS) that reflect the response of the patient to his or her illness (Box 8.31). Patient age is included in many scoring systems (Box 8.32). When combined with the admission diagnosis, scoring systems have been shown to correlate well with the risk of hospital death. Such outcome predictions can never be 100% accurate and should be viewed as only one of many factors to be considered when deciding whether or not further intervention is appropriate.

Predicted mortality figures by diagnosis have been calculated from large databases generated from a range of ICUs. These allow a particular unit to evaluate its performance compared to the reference ICUs by calculating standardised mortality ratios (SMRs) for each diagnostic group, by dividing observed mortality by predicted mortality.

A value of unity indicates the same performance as the reference ICUs, while a value less than 1 indicates a better than predicted outcome. If a unit has a high SMR in a certain diagnostic category, it should prompt investigation into the management of patients with that diagnosis, in order to identify aspects of care that could be improved.

Further information

Websites

www.adqi.net Evidence-based appraisal and consensus recommendations for diagnosis, treatment and research in acute kidney injury.
www.icudelirium.org Information on delirium and sedation in intensive care patients.
www.scottishintensivecare.org.uk/education/index.htm On-line tutorials, educational materials for learners and teaching support materials for educators.
www.survivingsepsis.org Surviving Sepsis website.
Comprehensive evaluation of the poisoned patient 206

Evaluation of the envenomed patient 207

General approach to the poisoned patient 208
  Triage and resuscitation 208
  Clinical assessment and investigations 209
  Psychiatric assessment 210
  General management 210

Poisoning by specific pharmaceutical agents 212
  Analgesics 212
  Antidepressants 213
  Cardiovascular medications 214
  Antimalarials 215
  Iron 216
  Antipsychotic drugs 216
  Antidiabetic agents 216

Drugs of misuse 216
  Cannabis 216
  Benzodiazepines 217
  Stimulants and entactogens 217
  Gamma-hydroxybutyrate and gamma-butyrolactone 218
  d-Lysergic acid diethylamide 218
  Opioids 219
  Body packers and body stuffers 219

Chemicals and pesticides 219
  Carbon monoxide 219
  Organophosphorus insecticides and nerve agents 220
  Carbamate insecticides 222
  Methanol and ethylene glycol 222
  Aluminium and zinc phosphide 223

Environmental poisoning and illness 223
  Arsenism 223
  Fluorosis 223

Substances less commonly taken in overdose 224

Envenoming 224
  Venom 225
  Venemous animals 225
  Clinical effects 226
  Management 226
Taking a history in poisoning

- What toxin(s) have been taken and how much?
- What time were they taken and by what route?
- Has alcohol or any drug of misuse been taken as well?

- Obtain details from witnesses of the circumstances of the overdose (e.g., family, friends, ambulance personnel)
- Ask the general practitioner for background and details of prescribed medication
- Assess suicide risk (full psychiatric evaluation when patient has physically recovered)
- Assess capacity to make decisions about accepting or refusing treatment
- Establish past medical history, drug history and allergies, social and family history
- Record all information carefully
Evaluation of the envenomed patient

Taking a history in envenoming

- When was the patient exposed to a bite/sting?
- Was the organism causing it seen and what did it look like (size, colour)?
- What were the circumstances (on land, in water etc.)?
- Was there more than one bite/sting?
- What first aid was used, when and for how long?
- What symptoms has the patient had (local and systemic)?
- Are there symptoms suggesting systemic envenoming (paralysis, myolysis, coagulopathy etc.)?
- Past medical history and medications?
- Past exposure to antivenom/venom and allergies?

Bites showing puncture marks, blistering, bruising and bleeding.
POISONING

Acute poisoning is common, accounting for about 1% of hospital admissions in the UK. Common or otherwise important substances involved are shown in Box 9.1. In developed countries, the most frequent cause is intentional drug overdose in the context of self-harm and usually involves prescribed or ‘over-the-counter’ medicines. Accidental poisoning is also common, especially in children and the elderly (Box 9.2). Toxicity also may occur as a result of alcohol or recreational substance use, or following occupational or environmental exposure. Poisoning is a major cause of death in young adults, but most deaths occur before patients reach medical attention, and mortality is much lower than 1% in those admitted to hospital.

In developing countries, the frequency of self-harm is more difficult to estimate. Household and agricultural products, such as pesticides and herbicides, are more freely available, are common sources of poisoning and are associated with a much higher case fatality. In China and South-east Asia, pesticides account for about 300000 suicides each year.

GENERAL APPROACH TO THE POISONED PATIENT

A general approach is shown on pages 206–207.

9.1 Important substances involved in poisoning

In the UK

- Analgesics: paracetamol and non-steroidal anti-inflammatory drugs (NSAIDs)
- Antidepressants: tricyclic antidepressants (TCAs), selective serotonin re-uptake inhibitors (SSRIs) and lithium
- Cardiovascular agents: β-blockers, calcium channel blockers and cardiac glycosides
- Drugs of misuse: opiates, benzodiazepines, stimulants (e.g. amphetamines, MDMA, cocaine)
- Carbon monoxide
- Alcohol

In South and South-east Asia

- Organophosphorus and carbamate insecticides
- Aluminium and zinc phosphide
- Oleander
- Snake venoms
- Antimalarial drugs: chloroquine
- Antidiabetic medication

9.2 Poisoning in old age

- **Aetiology:** commonly results from accidental poisoning (e.g. due to confusion or dementia) or drug toxicity as a consequence of impaired renal or hepatic function or drug interaction. Toxic prescription medicines are more likely to be available.
- **Psychiatric illness:** self-harm is less common than in younger adults but more frequently associated with depression and other psychiatric illness, as well as chronic illness and pain. There is a higher risk of subsequent suicide.
- **Severity of poisoning:** increased morbidity and mortality result from reduced renal and hepatic function, reduced functional reserve, increased sensitivity to sedative agents and frequent comorbidity.

Triage and resuscitation

Patients who are seriously poisoned must be identified early so that appropriate management is not delayed. Triage involves:
- immediate assessment of vital signs
- identifying the poison(s) involved and obtaining adequate information about them
- identifying patients at risk of further attempts at self-harm and removing any remaining hazards.

Those with possible external contamination with chemical or environmental toxins should undergo appropriate decontamination (Fig. 9.1). Critically ill patients must be resuscitated (p. 180).

The Glasgow Coma Scale (GCS) is commonly employed to assess conscious level, although it has not been specifically validated in poisoned patients. The AVPU (alert/verbal/painful/unresponsive) scale is also a rapid and simple method. An electrocardiogram (ECG) should be performed and cardiac monitoring instituted in all patients with cardiovascular features or where exposure to potentially cardiotoxic substances is suspected. Patients who may need antidotes should be weighed when this is feasible, so that appropriate weight-related doses can be prescribed.

Substances that are unlikely to be toxic in humans should be identified so that inappropriate admission and intervention are avoided (Box 9.3).

![Fig. 9.1 Methods of external decontamination.](image-url)
9.3 Substances of very low toxicity

- Writing/educational materials
- Decorating products
- Cleaning/bathroom products (except dishwasher tablets, which are corrosive)
- Pharmaceuticals: oral contraceptives, most antibiotics (but not tetracyclines and antituberculous drugs), H₂-blockers, proton pump inhibitors, emollients and other skin creams, baby lotion
- Miscellaneous: plasticine, silica gel, household plants, plant food

9.4 Causes of acidosis in the poisoned patient

<table>
<thead>
<tr>
<th>Cause</th>
<th>Normal lactate*</th>
<th>High lactate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toxic</td>
<td>Salicylates</td>
<td>Metformin</td>
</tr>
<tr>
<td></td>
<td>Methanol</td>
<td>Iron</td>
</tr>
<tr>
<td></td>
<td>Ethylene glycol</td>
<td>Cyanide</td>
</tr>
<tr>
<td></td>
<td>Paraldehyde</td>
<td>Sodium valproate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carbon monoxide</td>
</tr>
<tr>
<td>Other</td>
<td>Renal failure</td>
<td>Shock</td>
</tr>
<tr>
<td></td>
<td>Ketoadiabetes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Severe diarrhoea</td>
<td></td>
</tr>
</tbody>
</table>

*Unless circulatory shock is present, when it will be high in any case.

9.5 Anion and osmolar gaps in poisoning

<table>
<thead>
<tr>
<th>Anion gap</th>
<th>Osmolar gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Na⁺ + K⁺] – [Cl⁻ + HCO₃⁻]</td>
<td>[Osm (measured)] [2 × Na + Urea + Glucose (all mmol/L)]</td>
</tr>
</tbody>
</table>

Reference range: 12–16 mmol/L; < 10 mOsm/kg

1Osm (measured) stands for measured osmolality. For non-SI units, the corresponding formula is [Osm (measured)] = [2 × Na (meq/L)] + Urea/2.8 (mg/dL) + Glucose/18 (mg/dL).

2Box 16.19 (p. 445) gives non-toxic causes.

9.6 Laboratory analysis in poisoning

Organophosphates
- Plasma cholinesterase is reduced more rapidly but is less specific than red cell cholinesterase (p. 222)
- Antidote use should not be delayed pending results

Carboxyhaemoglobin
- > 20% indicates significant carbon monoxide exposure

Digoxin
- Therapeutic range usually 1–2 ng/mL (1.28–2.46 mmol/L)
- Concentrations > 4 ng/mL (5.12 mmol/L) usually associated with toxicity, especially with chronic poisoning

Ethanol
- Toxicity at concentrations > 1.8 g/L
- Take sample ≥ 4 hrs after overdose or if clinical signs of toxicity
- Concentrations > 5 mg/L suggest severe toxicity

Lithium
- Take sample ≥ 6 hrs after overdose or if clinical signs of toxicity
- Usual therapeutic range 0.4–1.0 mmol/L

Methaemoglobin
- Poisoning with nitrates, benzoicaine, dapsone, chloroquine and aniline dyes is associated with methaemoglobinemia
- Concentrations > 20% may require treatment with methylythioninium chloride (methylene blue)

Paracetamol
- Take sample ≥ 4 hrs after overdose
- Use nomogram to determine need for antidotal treatment (see Fig. 9.2, p. 212)

Salicylate
- Take sample ≥ 2 hrs (symptomatic patients) or 4 hrs (asymptomatic patients) after overdose
- Concentrations > 500 mg/L suggest serious toxicity
- Repeat after 2 hrs if severe toxicity is suspected

Theophylline
- Take sample ≥ 4 hrs after overdose or if clinical signs of toxicity
- Repeat after 2 hrs if severe toxicity is suspected
- Concentrations > 60 mg/L suggest severe toxicity

Clinical assessment and investigations

History and examination are described on page 206. Occasionally, patients may be unaware or confused about what they have taken, or may exaggerate (or less commonly underestimate) the size of the overdose, but rarely mislead medical staff deliberately. In regions of the world where self-poisoning is illegal, patients may be reticent about giving a history.

Toxic causes of abnormal physical signs are shown on page 207. The patient may have a cluster of clinical features (‘toxicmorphic’) suggestive of poisoning with a particular drug type, e.g. anticholinergic, serotoninergic (see Box 9.11, p. 213), stimulant, sedative, opioid (see Box 9.12, p. 217) or cholinergic (see Box 9.14, p. 221) feature clusters. Poisoning is a common cause of coma, especially in younger people, but it is important to exclude other potential causes (p. 1159), unless the aetiology is certain.

Urea, electrolytes and creatine should be measured in all patients with suspected systemic poisoning. Arterial blood gases should be checked in those with significant respiratory or circulatory compromise, or when poisoning with substances likely to affect acid–base status is suspected (Box 9.4). Calculation of anion and osmolar gaps may help to inform diagnosis and management (Box 9.5).

For a limited number of specific substances, management may be facilitated by measurement of the amount of toxin in the blood (Box 9.6). Qualitative urine screens for potential toxins, including near-patient testing kits, have a limited clinical role.
Psychiatric assessment

All patients presenting with deliberate drug overdose should undergo psychiatric evaluation by a health professional with appropriate training prior to discharge (p. 238). This should take place once the patient has recovered from any features of poisoning, unless there is an urgent issue, such as uncertainty about their capacity to decline medical treatment.

General management

Patients presenting with eye/skin contamination should undergo local decontamination procedures (see Fig. 9.1).

Gastrointestinal decontamination

Patients who have ingested potentially life-threatening quantities of toxins may be considered for gastrointestinal decontamination if poisoning has been recent (Box 9.7). Induction of emesis using ipecacuanha is no longer recommended.

Activated charcoal

Given orally as slurry, activated charcoal absorbs toxins in the bowel as a result of its large surface area. If given sufficiently early, it can prevent absorption of an important proportion of the ingested dose of toxic. Efficacy decreases with time and current guidelines do not advocate use more than 1 hour after overdose in most circumstances (see Box 9.7). However, use after a longer interval may be reasonable when a delayed-release preparation has been taken or when gastric emptying may be delayed. Some toxins do not bind to activated charcoal (Box 9.8) so it will not affect their absorption. In patients with impaired swallowing or a reduced level of consciousness, activated charcoal, even via a nasogastric tube, carries a risk of aspiration pneumonitis, which can be reduced (but not eliminated) by protecting the airway with a cuffed endotracheal tube.

9.8 Substances poorly adsorbed by activated charcoal

<table>
<thead>
<tr>
<th>Medicines</th>
<th>Chemicals</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Iron</td>
<td>• Lithium</td>
</tr>
<tr>
<td>• Acids*</td>
<td>• Mercury</td>
</tr>
<tr>
<td>• Alkalis*</td>
<td>• Methanol</td>
</tr>
<tr>
<td>• Ethanol</td>
<td>• Petroleum distillates*</td>
</tr>
</tbody>
</table>

*Gastric lavage contraindicated.

Multiple doses of oral activated charcoal (50 g 6 times daily in an adult) may enhance the elimination of some drugs at any time after poisoning and are recommended for serious poisoning with some substances (see Box 9.7). This interrupts enterohepatic circulation or reduces the concentration of free drug in the gut lumen, to the extent that drug diffuses from the blood back into the bowel to be absorbed on to the charcoal: so-called ‘gastrointestinal dialysis’. A laxative is generally given with the charcoal to reduce the risk of constipation or intestinal obstruction by charcoal ‘briquette’ formation in the gut lumen.

Evidence suggests that single or multiple doses of activated charcoal do not improve clinical outcomes after poisoning with pesticides or oleander.

Gastric aspiration and lavage

Gastric aspiration and/or lavage is very infrequently indicated in acute poisoning, as it is no more effective than activated charcoal and complications are common, especially aspiration. Use may be justified for life-threatening overdoses of some substances that are not absorbed by activated charcoal (see Box 9.8).

Whole bowel irrigation

This is occasionally indicated to enhance the elimination of ingested packets of illicit drugs or slow-release tablets such as iron and lithium that are not absorbed by activated charcoal. It involves the administration of large quantities of osmotically balanced polyethylene glycol and electrolyte solution (1–2 L/hr for an adult), usually by a nasogastric tube, until the rectal effluent is clear. Contraindications include inadequate airway protection, haemodynamic instability, gastrointestinal haemorrhage, obstruction or ileus. Whole bowel irrigation may precipitate nausea and vomiting, abdominal pain and electrolyte disturbances.

Urinary alkalisation

Urinary excretion of weak acids and bases is affected by urinary pH, which changes the extent to which they are ionised. Highly ionised molecules pass poorly through lipid membranes and therefore little tubular reabsorption occurs and urinary excretion is increased. If the urine is alkalised (pH > 7.5) by the administration of sodium bicarbonate (e.g. 1.5 L of 1.26% sodium bicarbonate over 2 hrs), weak acids (e.g. salicylates, methotrexate and the herbicides 2,4-dichlorophenoxyacetic acid and mecoprop) are highly ionised, resulting in enhanced urinary excretion.
Urinary alkalisation is currently recommended for patients with clinically significant salicylate poisoning when the criteria for haemodialysis are not met (see below). It is also sometimes used for poisoning with methotrexate. Complications include alkalaeemia, hypokalaemia and occasionally alkalotic tetany (p. 447). Hypocalcaemia may occur but is rare.

Haemodialysis and haemoperfusion

These techniques can enhance the elimination of poisons that have a small volume of distribution and a long half-life after overdose, and are appropriate when poisoning is sufficiently severe to justify invasive elimination methods. The toxin must be small enough to cross the dialysis membrane (haemodialysis) or must bind to activated charcoal (haemoperfusion) (Box 9.9). Haemodialysis may also correct acid–base and metabolic disturbances associated with poisoning (p. 209).

Lipid emulsion therapy

Lipid emulsion therapy, or ‘lipid rescue’, is being used increasingly for the management of poisoning with lipid-soluble agents, such as local anaesthetics, tricyclic antidepressants, calcium channel blockers and lipid-soluble β-blockers such as propranolol. It involves intravenous infusion of 20% lipid emulsion (e.g. Intralipid®) at an initial dose of 1.5 mL/kg, followed by a continued infusion of 0.25 mL/kg/min until there is clinical improvement. It is thought that lipid-soluble toxins partition into the intravenous lipid, reducing target tissue concentrations. The elevated myocardial concentration of free fatty acid induced by Intralipid administration may also have beneficial effects on myocardial metabolism and performance by counteracting the inhibition of myocardial fatty acid oxidation produced by local anaesthetics and some other cardiotoxins. This reverses cardiac depression by enabling increased ATP synthesis and energy production. Animal studies have suggested efficacy and case reports of use in human poisoning have also been encouraging, with recovery of circulatory collapse reported in cases where other treatment modalities have been unsuccessful. No controlled trials of this technique have been performed, however, and as a result, its efficacy remains uncertain.

9.9 Poisons effectively eliminated by haemodialysis or haemoperfusion

Haemodialysis

- Ethylene glycol
- Isopropanol
- Methanol

Haemoperfusion

- Theophylline
- Phenytoin
- Carbamazepine

- Salicylates
- Sodium valproate
- Lithium

- Phenobarbital
- Amobarbital

9.10 Complications of poisoning and their management

<table>
<thead>
<tr>
<th>Examples of causative agents</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coma</td>
<td>Sedative agents</td>
</tr>
<tr>
<td>Seizures</td>
<td>NSAIDs</td>
</tr>
<tr>
<td></td>
<td>Anticonvulsants</td>
</tr>
<tr>
<td></td>
<td>TCAs</td>
</tr>
<tr>
<td></td>
<td>Theophylline</td>
</tr>
<tr>
<td>Acute dystonias</td>
<td>Typical antipsychotics</td>
</tr>
<tr>
<td></td>
<td>Metoclopramide</td>
</tr>
<tr>
<td>Hypotension Due to vasodilatation</td>
<td>Vasodilator antihypertensives</td>
</tr>
<tr>
<td></td>
<td>Anticholinergic agents</td>
</tr>
<tr>
<td></td>
<td>TCAs</td>
</tr>
<tr>
<td></td>
<td>β-blockers</td>
</tr>
<tr>
<td></td>
<td>Calcium channel blockers</td>
</tr>
<tr>
<td></td>
<td>TCAs</td>
</tr>
<tr>
<td>Ventricular tachycardia Monomorphic, associated with QRS prolongation</td>
<td>Sodium channel blockers</td>
</tr>
<tr>
<td>Torsades de pointes, associated with QT interval prolongation</td>
<td>Anti-arrhythmic drugs (quinidine, amiodarone, sotalol)</td>
</tr>
<tr>
<td></td>
<td>Antimalarials</td>
</tr>
<tr>
<td></td>
<td>Organophosphate insecticides</td>
</tr>
<tr>
<td></td>
<td>Antipsychotic agents</td>
</tr>
<tr>
<td></td>
<td>Antidepressants</td>
</tr>
<tr>
<td></td>
<td>Antibiotics (erythromycin)</td>
</tr>
</tbody>
</table>

Appropriate airway protection and ventilatory support
Pressure area and bladder care
Identification and treatment of aspiration pneumonia

Appropriate airway and ventilatory support
IV benzodiazepine (e.g. diazepam 10–20 mg, lorazepam 2–4 mg)
Correction of hypoxia, acid–base and metabolic abnormalities

IV fluids
Vaspressors (rarely indicated; p. 191)
Optimisation of volume status
Inotropics (p. 191)

Correction of electrolyte and acid–base abnormalities and hypoxia
Sodium bicarbonate
Magnesium sulphate, 2 g IV over 1–2 mins, repeated if necessary
POISONING

Supportive care
For most poisons, antidotes and methods to accelerate elimination are inappropriate, unavailable or incompletely effective. Outcome is dependent on appropriate nursing and supportive care, and on treatment of complications (Box 9.10). Patients should be monitored carefully until the effects of any toxins have dissipated.

Antidotes
Antidotes are available for some poisons and work by a variety of mechanisms: for example, by specific antagonism (isoprenaline for β-blockers), chelation (desferroxamine for iron) or reduction (methylene blue for dapsone). The use of some antidotes is described in the management of specific poisons below.

POISONING BY SPECIFIC PHARMACEUTICAL AGENTS

Analgesics

Paracetamol
Paracetamol (acetaminophen) is the drug most commonly used in overdose in the UK. Toxicity results from formation of an intermediate reactive metabolite that binds covalently to cellular proteins, causing cell death. This results in hepatic and occasionally renal failure. In therapeutic doses, the toxic intermediate metabolite is detoxified in reactions requiring glutathione, but in overdose, glutathione reserves become exhausted.

Management
Management is summarised in Figure 9.2. Activated charcoal may be used in patients presenting within 1 hour. Antidotes for paracetamol act by replenishing hepatic glutathione and should be administered to all patients with paracetamol concentrations above the ‘treatment line’ provided on paracetamol poisoning nomograms. Acetylcysteine given intravenously (or orally in some countries) is highly efficacious if administered within 8 hours of the overdose. However, since efficacy declines thereafter, administration should not be delayed in patients presenting after 8 hours to await a paracetamol blood concentration result. The antidote can be stopped if the paracetamol concentration is shown to be below the nomogram treatment line.

The most important adverse effect of acetylcysteine is related to dose-related histamine release, the ‘anaphylactoid’ reaction, which causes itching and urticaria, and in occasional severe cases, bronchospasm and hypotension. Most cases can be managed by temporary discontinuation of acetylcysteine and administration of an antihistamine.

An alternative antidote is methionine 2.5 g orally (adult dose) every 4 hours to a total of 4 doses, but this is less effective, especially after delayed presentation. If a patient presents more than 15 hours after ingestion, liver function tests, prothrombin time (or international normalised ratio - INR), renal function tests and a venous bicarbonate should be measured, the antidote started, and a poisons information centre or local liver unit contacted for advice if results are abnormal. An arterial blood gas sample should be taken in patients with severe liver function abnormalities; metabolic acidosis indicates severe poisoning. Liver transplantation should be considered in individuals who develop life-threatening liver failure due to paracetamol poisoning (p. 932).

If multiple ingestions of paracetamol have taken place over several hours or days (i.e. a staggered overdose), acetylcysteine may be indicated. Recommended thresholds for treatment vary between countries.

Salicylates (aspirin)

Clinical features
Salicylate overdose commonly causes nausea, vomiting, sweating, tinnitus and deafness. Direct stimulation of the respiratory centre produces hyperventilation and respiratory alkalosis. Peripheral vasodilatation with bounding pulses and profuse sweating occurs in moderately severe poisoning. Serious salicylate poisoning is associated with metabolic acidosis, hypoprothrombinaemia, hyperglycaemia, hyperpyrexia, renal failure, pulmonary oedema, shock and cerebral oedema. Agitation, confusion, coma and fits may occur, especially in children. Toxicity is enhanced by acidosis, which increases salicylate transfer across the blood-brain barrier.
Management
Activated charcoal should be administered if the patient presents within 1 hour. Multiple doses of activated charcoal may enhance salicylate elimination but currently are not routinely recommended.

The plasma salicylate concentration should be measured at least 2 (in symptomatic patients) or 4 hours (asymptomatic patients) after overdose and repeated in suspected serious poisoning, since concentrations may continue to rise some hours after overdose. In adults, concentrations above 500 mg/L and 700 mg/L suggest serious and life-threatening poisoning respectively, although clinical status is more important than the salicylate concentration in assessing severity.

Dehydration should be corrected carefully, as there is a risk of pulmonary oedema, and metabolic acidosis should be identified and treated with intravenous sodium bicarbonate (8.4%), once plasma potassium has been corrected. Urinary alkalisation is indicated for adults with salicylate concentrations above 500 mg/L.

Haemodialysis is very effective at removing salicylate and correcting acid–base and fluid balance abnormalities, and should be considered when serum concentrations are above 700 mg/L in adults with severe toxic features, or when there is renal failure, pulmonary oedema, coma, convulsions or refractory acidosis.

Non-steroidal anti-inflammatory drugs
Clinical features
Overdose of most non-steroidal anti-inflammatory drugs (NSAIDs) usually causes little more than minor abdominal discomfort, vomiting and/or diarrhoea, but convulsions can occur occasionally, especially with mfenamic acid. Coma, prolonged seizures, apnoea, liver dysfunction and renal failure can occur after substantial overdose but are rare. Features of toxicity are unlikely to develop in patients who are asymptomatic more than 6 hours after overdose.

Management
Electrolytes, liver function tests and a full blood count should be checked in all but the most trivial cases. Activated charcoal may be given if the patient presents sufficiently early. Symptomatic treatment for nausea and gastrointestinal irritation may be necessary.

Antidepressants
Tricyclic antidepressants
Tricyclic antidepressants (TCAs) are used frequently in overdose and carry a high morbidity and mortality relating to their sodium channel-blocking, anticholinergic and α-adrenoceptor-blocking effects.

Clinical features
Anticholinergic effects are common (Box 9.11). Life-threatening complications are frequent, including convulsions, coma, arrhythmias (ventricular tachycardia, ventricular fibrillation and, less commonly, heart block) and hypotension, which results from inappropriate vasodilatation or impaired myocardial contractility. Serious complications appear to be more common with dosulepin and amitriptyline.

<table>
<thead>
<tr>
<th>Anticholinergic</th>
<th>Serotonin syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common causes</td>
<td></td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td>SSRIs</td>
</tr>
<tr>
<td>Antipsychotics</td>
<td>Monoamine oxidase inhibitors (MAOIs)</td>
</tr>
<tr>
<td>TCAs</td>
<td>TCAs</td>
</tr>
<tr>
<td>Antihistamines</td>
<td>Amphetamines</td>
</tr>
<tr>
<td>Scopolamine</td>
<td></td>
</tr>
<tr>
<td>Benztropine</td>
<td>Buspironne</td>
</tr>
<tr>
<td>Belladonna</td>
<td>Bupropion (especially in combination)</td>
</tr>
<tr>
<td>Jimson weed</td>
<td></td>
</tr>
<tr>
<td>Mushrooms (some)</td>
<td></td>
</tr>
</tbody>
</table>

Clinical features
Cardiovascular
- Tachycardia, hypertension
- Tachycardia, hyper- or hypotension

CNS
- Confusion, hallucinations, sedation
- Confusion, delirium, hallucinations, sedation, coma

Muscle
- Myoclonus
- Shivering, tremor, myoclonus, raised creatine kinase

Temperature
- Fever
- Normal pupil size

Eyes
- Diplopia, mydriasis
- Diarrhoea, vomiting

Abdomen
- Ileus, palpable bladder
- Diarrhoea, vomiting

 Mouth
- Dry
- Flushing, hot, dry
- Flushing, sweating

Complications
- Seizures
- Seizures
- Rhabdomyolysis
- Renal failure
- Metabolic acidosis
- Coagulopathies

Fig 9.3 ECG in severe tricyclic antidepressant poisoning. This rhythm strip shows a broad QRS complex due to impaired conduction.

Management
Activated charcoal should be administered if the patient presents within 1 hour. All patients with possible TCA overdose should have a 12-lead ECG and ongoing cardiac monitoring for at least 6 hours. Prolongation of the QRS interval (especially if > 0.16 s) indicates severe sodium channel blockade and is associated with an increased risk of arrhythmia (Fig. 9.3). QT interval prolongation may also occur. Arterial blood gases should be measured in suspected severe poisoning.

In patients with arrhythmias, significant QRS or QT prolongation or acidosis, intravenous sodium bicarbonate (50 mL of 8.4% solution) should be administered and repeated to correct pH. The correction of the acidosis and the sodium loading that result is often associated with rapid improvement in ECG features and arrhythmias. Hypoxia and electrolyte abnormalities should also be corrected. Anti-arrhythmic drugs should only be given on specialist advice. Prolonged convulsions should be treated with intravenous benzodiazepines (see Box 9.10). There is anecdotal evidence of benefit
from lipid emulsion therapy in severe intractable poisoning.

**Selective serotonin and noradrenaline re-uptake inhibitors**

Selective serotonin re-uptake inhibitors (SSRIs) are a group of antidepressants that include fluoxetine, paroxetine, fluvoxamine, sertraline, citalopram and escitalopram. They are increasingly used to treat depression, partly because they are less toxic in overdose than TCAs. A related group of compounds termed serotonin–noradrenaline reuptake inhibitors (SNRIs), such as venlafaxine and duloxetine, are also in common use and are sometimes taken in overdose.

**Clinical features and management**

Overdose of SSRIs may produce nausea and vomiting, tremor, insomnia and sinus tachycardia. Agitation, drowsiness and convulsions occur infrequently and may be delayed for several hours after ingestion. Occasionally, features of serotonin syndrome may develop (see Box 9.11), especially if SSRIs are taken in combination or with other serotonergic agents. Cardiac arrhythmias occur infrequently and most patients require supportive care only. The toxic effects of SNRIs are similar but tachycardia, hypertension or hypotension and ECG changes (QRS and QT prolongation) may be more prominent and hypoglycaemia can also occur.

**Lithium**

Severe lithium toxicity is uncommon after intentional overdose and is more often encountered in patients taking therapeutic doses as the result of interactions with drugs such as diuretics or NSAIDs that can cause dehydration or renal impairment, or because an excessive dose has been prescribed. Severe toxicity is more common with acute overdose in patients taking chronic therapy ‘acute on chronic’ poisoning).

**Clinical features**

Nausea, diarrhoea, polyuria, dizziness and tremor may progress to muscular weakness, drowsiness, confusion, myoclonus, fasciculations, chorea and renal failure. Coma, convulsions, ataxia, cardiac dysrhythmias such as heart block, blood pressure disturbances and renal failure may all occur in severe poisoning.

**Management**

Activated charcoal is ineffective. Gastric lavage is of theoretical benefit if used early after overdose, but lithium tablets are likely to remain intact in the stomach and may be too large for aspiration via a lavage tube. Some advocate whole bowel irrigation after substantial overdose but efficacy is unknown.

Lithium concentrations should be measured immediately in symptomatic patients or after at least 6 hours in asymptomatic patients following acute overdose. Adequate hydration should be maintained with intravenous fluids. Convulsions should be treated as in Box 9.10.

In patients with features suggesting severe toxicity associated with high lithium concentrations (e.g. > 4.0 mmol/L after chronic or ‘acute on chronic’ poisoning, or > 7.5 mmol/L after acute poisoning), haemodialysis should be considered. Lithium concentrations are reduced substantially during dialysis but rebound increases occur after discontinuation, and multiple sessions may be required.

**Cardiovascular medications**

**Beta-adrenoceptor blockers**

These have negative inotropic and chronotropic effects. Some have additional properties that may increase toxicity, such as blockade of sodium channels with propranolol, acebutolol and carvedilol, and blockade of potassium channels with sotalol.

**Clinical features**

The major features of toxicity are bradycardia and hypotension. Heart block, pulmonary oedema and cardiogenic shock occur in severe poisoning. Beta-blockers with sodium channel-blocking effects may cause seizures, confusion and coma, while sotalol may be associated with repolarisation abnormalities (including QT prolongation) and torsades de pointes (p. 570).

**Management**

Intravenous fluids may reverse hypotension but care is required to avoid pulmonary oedema. Bradycardia and hypotension may respond to high doses of atropine (up to 3 mg in an adult). The adrenergic agonist isoprenaline may also be effective but high doses are often needed. Glucagon (5–10 mg over 10 mins, then 1–5 mg/hr by infusion), which counteracts the effect of β-blockers by stimulating intracellular production of cyclic adenosine monophosphate (cAMP), is now more commonly used. In severe cases, ‘hyperinsulinaemia euglycaemic therapy’ has been used, as described under calcium channel blockers. Lipid emulsion therapy may have a role in severe poisoning with lipid-soluble agents such as propranolol, carvedilol and oxprenolol.

**Calcium channel blockers**

Calcium channel blockers are highly toxic in overdose because of their inhibitory effects on L-type calcium channels. Dihydropyridines, such as nifedipine or amloidipine, affect vascular smooth muscle in particular, resulting in vasodilatation, whereas diltiazem and verapamil, which are used in the treatment of arrhythmias, have predominantly cardiac effects, including bradycardia and reduced myocardial contractility.

**Clinical features**

The usual presentation is with hypotension due to vasodilatation or myocardial depression. Bradycardias and heart block may also occur, especially with verapamil and diltiazem. Gastrointestinal disturbances, confusion, metabolic acidosis, hyperglycaemia and hyperkalaemia may also be present.

**Management**

Hypotension should be corrected with intravenous fluids, taking care to avoid pulmonary oedema. Persistent hypotension may respond to intravenous calcium gluconate (10 mg IV over 5 mins, repeated as required). Isoprenaline and glucagon may also be useful. Successful use of intravenous insulin with glucose (10–20%
dextrose with insulin initially at 0.5–2.0 U/kg/hr, increasing to 5–10 U/kg/hr according to clinical response), so-called ‘hyperinsulinaemia euglycaemic therapy’, has been reported in patients unresponsive to other strategies. The mechanism of action remains to be fully elucidated, but in states of shock myocardial metabolism switches from use of free fatty acids to glucose. Calcium channel blocker poisoning is also associated with hypoinsulininaemia and insulin resistance, impeding glucose uptake by myocytes. High doses of insulin inhibit lipolysis and increase glucose uptake and the efficiency of glucose utilisation. Cardiac pacing may be needed for severe unresponsive bradycardias or heart block. Lipid emulsion therapy has been used in severe poisoning with apparent benefit, although evidence is largely anecdotal.

**Digoxin and oleander**

Poisoning with digoxin is usually accidental, arising from prescription of an excessive dose, impairment of renal function or drug interactions. In South Asia, deliberate self-poisoning with yellow oleander (*Thevetia peruviana*), which contains cardiac glycosides, is common.

**Clinical features**

Characteristic cardiac effects of toxicity are tachyarrhythmias (either atrial or ventricular) and bradyarrhythmias, with or without atrioventricular block. Ventricular bigeminy is common and atrial tachycardia with evidence of atrioventricular block is highly suggestive of the diagnosis. Severe poisoning is associated with hyperkalaemia. Non-cardiac features include confusion, headache, nausea, vomiting, diarrhoea and (rarely) altered colour vision.

**Management**

Activated charcoal is commonly administered to patients presenting within 1 hour of ingestion of an acute overdose, although evidence of benefit is lacking. Urea, electrolytes and creatinine should be measured, a 12-lead ECG performed and cardiac monitoring instituted. Hypoxia, hypokalaemia (sometimes associated with concurrent diuretic use), hypomagnesaemia and acidosis increase the risk of arrhythmias and should be corrected. Significant bradyarrhythmias may respond to atropine, although temporary pacing is sometimes needed. Ventricular arrhythmias may respond to intravenous magnesium (see Box 9.10). If available, digoxin-specific antibody fragments should be administered when there are severe ventricular arrhythmias or unresponsive bradycardias. This antidote has been shown to be effective for both digitalis and yellow oleander poisoning.

**Antimalarials**

**Chloroquine**

Chloroquine is highly toxic in overdose and quantities of 5 g or more of chloroquine base are likely to be fatal in an adult.

**Clinical features**

Features of toxicity occur within 1 hour of ingestion and include nausea, vomiting, agitation, drowsiness, hypokalaemia, acidosis, headaches and blurred vision. Coma, convulsions and hypotension may occur in severe poisoning. ECG changes indicating conduction and repolarisation delay (prolonged QRS and QT intervals) occur and are associated with ventricular tachycardia (including torsades de pointes), ventricular fibrillation and sudden death.

**Management**

Activated charcoal should be given to all patients presenting within 1 hour of ingestion of chloroquine in amounts greater than 15 mg/kg. Cardiac rhythm should be monitored and dysrhythmias managed as outlined in Box 9.10. The arterial pH should be corrected, but hypokalaemia is thought to have a protective effect and should not be corrected in the first 8 hours after poisoning. High-dose diazepam (2 mg/kg body weight IV over 30 mins followed by an infusion of 2 mg/kg/hr) has been suggested to have a protective effect, especially if given in the early stages of severe chloroquine poisoning, but evidence is limited as yet. One controlled trial did not show beneficial effects on the ECG. Diazepam therapy requires intubation and mechanical ventilation to avoid pulmonary aspiration.

**Quinine**

Quinine salts are widely used for treating malaria and leg cramps. Deaths have been reported with ingestion of as little as 1.5 g in an adult and 900 mg in a child.

**Clinical features**

Features of toxicity include nausea, vomiting, tremor, tinnitus and deafness. Hypotension, haemolysis, renal failure, ataxia, convulsions and coma are features of serious poisoning. Conduction and repolarisation delay results in prolonged QRS and QT intervals on the ECG, and ventricular tachycardia (including torsades de pointes), ventricular fibrillation and sudden death may occur. Quinine-induced retinal photoreceptor cell toxicity may result in blurred vision and impaired colour perception. This usually develops a few hours after overdose and progresses to constriction of the visual field, scotoma and complete blindness associated with pupillary dilatation and unresponsiveness to light. Fundoscopy may show retinal artery spasm, disc pallor and retinal oedema. Although visual loss can be permanent, some degree of recovery, especially of central vision, often occurs over several weeks.

**Management**

Multiple-dose activated charcoal should be commenced in patients who have taken quinine in amounts greater than 15 mg/kg. Gastric lavage should be considered in patients who have taken a substantial overdose who present within 1 hour. All patients should have a 12-lead ECG and cardiac monitoring, and their urea, electrolytes and glucose checked. Dysrhythmias, hypotension, seizures and coma should be managed as outlined in Box 9.10.

There are no effective treatments for the visual effects of quinine. Stellate ganglion block and retrobulbar or intravenous injections of vasodilators such as nitrates were previously used but are ineffective, as are haemodialysis and haemoperfusion.
**Iron**

Overdose with iron can cause severe and sometimes fatal poisoning. The toxicity of individual iron preparations is related to their elemental iron content.

**Clinical features**

Early clinical features include gastrointestinal disturbance with the passage of grey or black stools. Hyperglycaemia and leucocytosis may occur. Haematemeses, rectal bleeding, drowsiness, convulsions, coma, metabolic acidosis and cardiovascular collapse may occur in severe poisoning.

Early symptoms may improve or even resolve within 6-12 hours, but hepatocellular necrosis may develop 12-24 hours after overdose and occasionally progresses to hepatic failure. Gastrointestinal strictures are late complications of iron poisoning.

**Management**

Gastric lavage may be considered in patients presenting within 1 hour of life-threatening overdose but efficacy has not been established. Activated charcoal is ineffective since iron is not bound. Serum iron concentration should be measured (see Box 9.6, p. 209). The antidote desferrioxamine chelates iron and should be administered immediately in patients with severe features, without waiting for serum iron concentrations to be available. Symptomatic patients with high serum iron concentrations (e.g. > 5 mg/L) should also receive desferrioxamine. Desferrioxamine may cause hypotension, allergic reactions and occasionally pulmonary oedema. Otherwise, treatment is supportive and directed at complications.

**Antipsychotic drugs**

Antipsychotic drugs (p. 248) are often prescribed for patients at high risk of self-harm or suicide, and are commonly encountered in overdose.

**Clinical features**

Drowsiness, tachycardia and hypotension are frequently found. Anticholinergic features (see Box 9.11) and acute dystonias, such as oculogyric crisis, torticollis and trismus, may occur after overdose with typical antipsychotics like haloperidol or chlorpromazine. QT interval prolongation and torsades de pointes can occur with typical antipsychotics such as thioridazine and haloperidol, as well as atypical antipsychotics like quetiapine, amisulpride and ziprasidone. Convulsions may occur with both groups of agent.

**Management**

Activated charcoal may be of benefit if given within 1 hour of overdose. Cardiac monitoring should be undertaken for at least 6 hours. Management is largely supportive, with treatment directed at complications (see Box 9.10, p. 211).

**Antidiabetic agents**

Antidiabetic agents commonly causing toxicity in overdose include sulphonylureas such as chlorpropamide, glibenclamide, gliclazide, glipizide and tolbutamide; biguanides like metformin and phenformin; and insulin. Overdose may also be encountered with some of the newer antidiabetic drugs, such as thiazolidinediones (pioglitazone), meglinitides (nateglinide, repaglinide) and dipeptidyl peptidase (DPP)-IV inhibitors (sitagliptin).

**Clinical features**

Sulphonylureas, meglinitides and parenteral insulin cause hypoglycaemia when taken in overdose, although insulin is non-toxic if ingested by mouth. The duration of hypoglycaemia depends on the half-life or release characteristics of the preparation and may be prolonged over several days with long-acting agents such as chlorpropamide, insulin zinc suspension or insulin glargine.

Features of hypoglycaemia include nausea, agitation, sweating, aggression and behavioural disturbances, confusion, tachycardia, hypothermia, drowsiness, coma or convulsions (p. 814). Permanent neurological damage can occur if hypoglycaemia is prolonged. Hypoglycaemia can be diagnosed using bedside glucose strips but venous blood should also be sent for laboratory confirmation.

Metformin is uncommonly associated with hypoglycaemia. Its major toxic effect in overdose is lactic acidosis, which can have a high mortality, and is particularly common in older patients and those with renal or hepatic impairment, or when ethanol is co-ingested. Other features of metformin overdose are nausea and vomiting, diarrhoea, abdominal pain, drowsiness, coma, hypotension and cardiovascular collapse.

There is limited experience of overdose involving thiazolidinediones and DPP-IV inhibitors, but significant hypoglycaemia is unlikely.

**Management**

Activated charcoal should be considered for all patients who present within 1 hour of ingestion of a substantial overdose of an oral hypoglycaemic agent. Venous blood glucose and urea and electrolytes should be measured and tests repeated regularly. Hypoglycaemia should be corrected using oral or intravenous glucose (50 mL of 50% dextrose); an infusion of 10–20% dextrose may be required to prevent recurrence. Intramuscular glucagon can be used as an alternative, especially if intravenous access is unavailable. Failure to regain consciousness within a few minutes of normalisation of the blood glucose can indicate that a central nervous system (CNS) depressant has also been ingested, the hypoglycaemia has been prolonged, or that the coma has another cause (e.g. cerebral haemorrhage or oedema).

Arterial blood gases should be taken after metformin overdose to assess the extent of acidosis. If present, plasma lactate should be measured and acidosis should be corrected with intravenous sodium bicarbonate (250 mL 1.26% solution or 50 mL 8.4% solution, repeated as necessary). In severe cases, haemodialysis or haemodiafiltration is used.

**DRUGS OF MISUSE**

**Cannabis**

Cannabis is derived from the dried leaves and flowers of Cannabis sativa. When it is smoked, the onset of effect occurs within 10–30 minutes, whereas after ingestion
the onset is 1–3 hours later. The duration of effect is 4–8 hours. Cannabis produces euphoria, perceptual alterations and conjunctival injection, followed by enhanced appetite, relaxation and occasionally hypertension, tachycardia, slurred speech and ataxia. High doses may produce anxiety, confusion, hallucinations and psychosis (Box 9.12). Psychological dependence is common but tolerance and withdrawal symptoms are unusual. Long-term use is thought to increase the lifetime risk of developing schizophrenia. Ingestion or smoking of cannabis rarely results in serious poisoning and supportive treatment is all that is required.

**Benzodiazepines**

Benzodiazepines may be prescribed or used illicitly. They are of low toxicity when taken alone in overdose but can enhance CNS depression when taken with other sedative agents, including alcohol. They may also cause significant toxicity in the elderly and those with chronic lung or neuromuscular disease.

**Clinical features**

Clinical features of toxicity include drowsiness, ataxia and confusion (see Box 9.12). Respiratory depression and hypotension may occur with severe poisoning in susceptible groups, especially after intravenous administration of short-acting agents.

**Management**

Activated charcoal may be useful after ingestion in susceptible patients or after mixed overdose, if given within 1 hour. Conscious level, respiratory rate and oxygen saturation should be monitored for at least 6 hours after substantial overdose.

The specific benzodiazepine antagonist flumazenil increases conscious level in patients with overdose but carries a risk of seizures, and is contraindicated in patients co-ingesting proconvulsant agents such as TCAs and in those with a history of seizures.

**Stimulants and entactogens**

This group includes amphetamines, ecstasy, cathinones such as mephedrone, piperazines and cocaine. These are sympathomimetic and serotonergic amines; as a result, they have clinical features of poisoning that overlap (see Box 9.12).

---

**9.12 Stimulant, sedative and opioid feature clusters**

<table>
<thead>
<tr>
<th>Stimulant</th>
<th>Sedative hypnotic</th>
<th>Opioid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common causes</strong></td>
<td>Amphetamines</td>
<td>Barbiturates</td>
</tr>
<tr>
<td></td>
<td>MDMA (‘ecstasy’)</td>
<td>Ethanol</td>
</tr>
<tr>
<td></td>
<td>Ephedrine</td>
<td>Gamma-hydroxybutyrate (GHB)</td>
</tr>
<tr>
<td></td>
<td>Pseudoephedrine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cocaine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cannabis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phenylcyclidine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cathinones (e.g. mephedrone)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benzylpiperazine</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Clinical features**

<table>
<thead>
<tr>
<th>Respiratory</th>
<th>Tachypnoea</th>
<th>Reduced respiratory rate and ventilation¹</th>
<th>Reduced respiratory rate and ventilation¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>Tachycardia, hypertension</td>
<td>Hypotension¹</td>
<td>Hypotension, relative bradycardia</td>
</tr>
<tr>
<td>CNS</td>
<td>Restlessness, anxiety, anorexia,</td>
<td>Confusion, hallucinations, slurred speech</td>
<td>Confusion, hallucinations, slurred speech</td>
</tr>
<tr>
<td></td>
<td>insomnia</td>
<td>Sedation, coma¹</td>
<td>Sedation, coma²</td>
</tr>
<tr>
<td>Muscle</td>
<td>Tremor</td>
<td>Ataxia, reduced muscle tone</td>
<td>Ataxia, reduced muscle tone</td>
</tr>
<tr>
<td>Temperature</td>
<td>Fever</td>
<td>Hypothermia</td>
<td>Hypothermia</td>
</tr>
<tr>
<td>Eyes</td>
<td>Mydriasis</td>
<td>Diplopia, strabismus, nystagmus</td>
<td>Miosis</td>
</tr>
<tr>
<td>Abdomen</td>
<td>Abdominal pain, diarrhoea</td>
<td>Normal pupil size</td>
<td>Ileus</td>
</tr>
<tr>
<td>Mouth</td>
<td>Dry</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Skin</td>
<td>Piloerection</td>
<td>Blisters, pressure sores</td>
<td>Needle tracks²</td>
</tr>
</tbody>
</table>

**Complications**

- Seizures
- Myocardial infarction
- Dysrhythmias
- Rhabdomyolysis
- Renal failure
- Intracerebral haemorrhage or infarction
- Respiratory failure¹
- Respiratory failure²

¹ Especially barbiturates. ²IV use.
**Cocaine**

Cocaine is available as a water-soluble hydrochloride salt suitable for nasal inhalation (‘snorting’), or as an insoluble free base (‘crack’ cocaine) that, unlike the hydrochloride salt, vaporises at high temperature and can be smoked, giving a more rapid and intense effect. Cocaine hydrochloride is usually purchased as a white crystalline powder, and crack cocaine in ‘rocks’.

**Clinical features**

Effects appear rapidly after inhalation and especially after smoking. sympathomimetic stimulant effects are common (see Box 9.12). Serious complications usually occur within 3 hours of use and include coronary artery spasm, which may result in myocardial ischaemia or infarction, even in patients with normal coronary arteries. This may lead to hypotension, cyanosis and ventricular arrhythmias. Cocaine toxicity should be considered in young adults who present with ischaemic chest pain. Hyperpyrexia may be associated with rhabdomyolysis, acute renal failure and disseminated intravascular coagulation.

**Management**

All patients should be observed with ECG monitoring for a minimum of 4 hours. A 12-lead ECG should be performed. Abnormalities are common, including ST-segment elevation, which may occur even in the absence of myocardial infarction. Troponin T estimations are the most sensitive and specific markers of myocardial damage. Benzodiazepines and intravenous nitrates are useful for managing patients with chest pain or hypertension, but β-blockers are best avoided because of the risk of unopposed α-adrenoceptor stimulation. Coronary angiography should be considered in patients with myocardial infarction or acute coronary syndromes. Acidosis should be corrected. Physical cooling measures may be required for hyperthermia (p. 105).

**Amphetamines and cathinones**

These include amphetamine sulphate (‘speed’), methamphetamine (‘crystal meth’), 3,4-methylenedioxyamphetamine (MDMA, ‘ecstasy’) and methadone. Tolerance is common, leading regular users to seek progressively higher doses.

**Clinical features**

Toxic features usually appear within a few minutes of use and last 4-6 hours, or substantially longer after a large overdose. sympathomimetic stimulant and serotonergic effects are common (see Boxes 9.11 and 9.12). A proportion of ecstasy users develop hyponatraemia as a result of excessive water drinking and inappropriate antidiuretic hormone secretion. Muscle rigidity, pain and bruxism (clenching of the jaw) may occur. Hyperpyrexia, rhabdomyolysis, metabolic acidosis, acute renal failure, disseminated intravascular coagulation, hepatocellular necrosis, acute respiratory distress syndrome (ARDS) and cardiovascular collapse have all been described following MDMA use but are rare. Cerebral infarction and haemorrhage have been reported, especially after intravenous amphetamine use.

**Management**

Management is supportive and directed at complications (see Box 9.10, p. 211).

**Gamma-hydroxybutyrate and gamma butyrolactone**

Gamma-hydroxybuterate (GHB) and gamma butyrolactone (GBL) are sedative agents with psychedelic and body-building effects. They are easily manufactured from commonly available industrial chemicals, including 1,4 butanediol, which is metabolised to GHB in vivo and has similar effects after ingestion. GHB solution is drunk by users, who titrate the dose until the desired effects are achieved.

**Clinical features**

Toxic features are those of a sedative hypnotic (see Box 9.12). Nausea, diarrhoea, vertigo, tremor, mydriasis, extrapyramidal signs, euphoria, bradycardia, convulsions, metabolic acidosis, hypokalaemia and hyperglycaemia may also occur. As the drug may be produced in batches and shared amongst a number of individuals, several patients may present with coma at the same time. The sedative effects are potentiated by other CNS depressants, such as alcohol, benzodiazepines, opioids and neuroleptics. Coma usually resolves spontaneously and abruptly within a few hours but may occasionally persist for several days. Dependence may develop in regular users, who may experience severe prolonged withdrawal effects if use is discontinued abruptly.

**Management**

Activated charcoal is recommended within 1 hour for ingestion of GHB in amounts greater than 20 mg/kg. Urea, electrolytes and glucose should be measured in all but the most trivial of cases. All patients should be observed for a minimum of 2 hours, with monitoring of blood pressure, heart rate, respiratory rate and oxygenation. Patients who remain symptomatic should be observed in hospital until symptoms resolve, but require supportive care only. Withdrawal may require treatment with high doses of benzodiazepines.

**d-Lysergic acid diethylamide**

d-Lysergic acid diethylamide (LSD) is a synthetic hallucinogen usually ingested as small squares of impregnated absorbent paper (often printed with a distinctive design) or as ‘microdots’. The drug causes perceptual effects, such as heightened visual awareness of colours, distortion of images, and hallucinations that may be pleasurable or terrifying (‘bad trip’) and associated with panic, confusion, agitation or aggression. Dilated pupils, hypertension, pyrexia and metabolic acidosis may occur and psychosis may sometimes last several days.

Patients with psychotic reactions or CNS depression should be observed in hospital, preferably in a quiet, dimly lit room to minimise external stimulation. Where sedation is required, diazepam is the drug of choice. Antipsychotics should be avoided if possible, as they may precipitate cardiovascular collapse or convulsions.
Opioids

Commonly encountered opioids are shown in Box 9.12. Toxicity may result from misuse of illicit drugs such as heroin or after overdose of medicinal opiates such as dextropropoxyphene. Intravenous use of heroin or morphine gives a rapid, intensely pleasurable experience, often accompanied by heightened sexual arousal. Physical dependence occurs within a few weeks of regular high-dose injection; as a result, the dose is escalated and the user’s life becomes increasingly centred on obtaining and taking the drug. Withdrawal, which can start within 12 hours, presents with intense craving, rhinorrhoea, lacrimation, yawning, perspiration, shivering, piloerection, vomiting, diarrhoea and abdominal cramps. Examination reveals tachycardia, hypertension, mydriasis and facial flushing.

Accidental overdose with prescribed strong opioid preparations is common, especially in the elderly.

Clinical features

These are shown in Box 9.12. Needle tracks may be visible in intravenous drug users and drug-related paraphernalia may be found amongst their possessions. Severe poisoning results in respiratory depression, hypotension, non-cardiogenic pulmonary oedema and hypothermia, leading to respiratory arrest or aspiration of gastric contents. Dextropropoxyphene (the opioid component of co-proxamol) may cause cardiac conduction effects, particularly QRS prolongation, ventricular arrhythmias and heart block, and has been withdrawn in the UK and other countries. Methadone may cause QT prolongation and torsades de pointes.

Symptoms of opioid poisoning can be prolonged for up to 48 hours after use of long-acting agents such as methadone, dextropropoxyphene and oxycodone.

Management

The airway should be cleared and, if necessary, respiratory support and oxygen given. Oxygen saturation monitoring and measurement of arterial blood gases should be performed. Prompt use of the specific opioid antagonist naloxone (0.4–2 mg IV in an adult, repeated if necessary) may obviate the need for intubation, although excessive doses may precipitate acute withdrawal in chronic opiate users. An infusion may be required in some cases because the half-life of the antidote is short compared to that of most opiates, especially those with prolonged elimination. Patients must be monitored for at least 6 hours after the last naloxone dose. Other complications of naloxone therapy include fits and ventricular arrhythmias, although these are rare. Box 9.10 (p. 211) describes the management of coma, fits and hypotension. Non-cardiogenic pulmonary oedema does not usually respond to diuretic therapy, and continuous positive airways pressure (CPAP) or positive end-expiratory pressure (PEEP) ventilatory support (p. 193) may be required.

Body packers and body stuffers

Body packers (‘mules’) attempt to smuggle illicit drugs (usually cocaine, heroin or amphetamines) by ingesting multiple small packages wrapped in several layers of clingfilm or in condoms. Body stuffers are those who have ingested unpackaged or poorly wrapped substances, often to avoid arrest. Both groups are at risk of severe toxicity if the packages rupture. This is more likely for body stuffers, who may start to develop symptoms of poisoning within 8 hours of ingestion. The risk of poisoning depends on the quality of the wrapping, and the amount and type of drug ingested. Cocaine, for example, presents a much higher risk than heroin because of its high toxicity and lack of a specific antidote.

Patients suspected of body packing or stuffing should be admitted for observation. A careful history taken in private is important, but for obvious reasons patients may withhold details of the drugs involved. The mouth, rectum and vagina should be examined as possible sites for concealed drugs. A urine toxicology screen performed at intervals may provide evidence of leakage, although positive results may reflect earlier recreational drug use. Packages may be visible on plain abdominal films (Fig. 9.4), but this is not always the case, and ultrasound and computed tomography (CT) are more sensitive methods of visualisation. One of these (preferably CT) should be performed in all suspected body packers.

Antimotility agents are often used by body packers to prevent premature passage of packages, so it can take a number of days for packages to pass spontaneously; during this period the carrier is at risk from package rupture. Whole bowel irrigation is commonly used to accelerate passage and is continued until all packages have passed. Surgery may be required for mechanical bowel obstruction or when evolving clinical features suggest package rupture, especially with cocaine.

CHEMICALS AND PESTICIDES

Carbon monoxide

Carbon monoxide (CO) is a colourless, odourless gas produced by faulty appliances burning organic fuels. It is also present in vehicle exhaust fumes and sometimes in smoke from house fires. It causes toxicity by binding with haemoglobin and cytochrome oxidase, which
POISONING

reduces tissue oxygen delivery and inhibits cellular respiration. It is a common cause of death by poisoning and most patients who die of CO poisoning do so before reaching hospital.

Clinical features

Early clinical features of acute severe carbon monoxide poisoning include headache, nausea, irritability, weakness and tachypnoea. Because these are non-specific, the correct diagnosis will not be obvious if the exposure is occult, such as from a faulty domestic appliance. Subsequently, ataxia, nystagmus, drowsiness and hyper-reflexia may develop, progressing to coma, convulsions, hypotension, respiratory depression, cardiovascular collapse and death. Myocardial ischaemia may occur and results in arrhythmias or myocardial infarction. Cerebral oedema is common and rhabdomyolysis may lead to myoglobinuria and renal failure. In those who recover from acute toxicity, longer-term neuropsychiatric effects are common, such as personality change, memory loss and concentration impairment. Extrapyramidal effects, urinary or faecal incontinence, and gait disturbance may also occur. Poisoning during pregnancy may cause fetal hypoxia and intrauterine death.

Management

Patients should be removed from exposure as soon as possible and resuscitated as necessary. Oxygen should be administered in as high a concentration as possible via a tightly fitting facemask, as this reduces the half-life of carboxyhaemoglobin from 4–6 hours to about 40 minutes. Measurement of carboxyhaemoglobin is useful for confirming exposure (see Box 9.6, p. 209) but levels do not correlate well with the severity of poisoning, partly because concentrations fall rapidly after removal of the patient from exposure, especially if supplemental oxygen has been given.

An ECG should be performed in all patients with acute poisoning, especially those with pre-existing heart disease. Arterial blood gas analysis should be checked in those with serious poisoning. Oxygen saturation readings by pulse oximetry are misleading since both carboxyhaemoglobin and oxyhaemoglobin are measured. Excessive intravenous fluid administration should be avoided, particularly in the elderly, because of the risk of pulmonary and cerebral oedema. Convulsions should be controlled with diazepam.

Hyperbaric oxygen therapy is controversial. At 2.5 atmospheres it reduces the half-life of carboxyhaemoglobin to about 20 minutes and increases the amount of dissolved oxygen by a factor of 10, but systematic reviews have shown no improvement in clinical outcomes. The logistical difficulties of transporting sick patients to hyperbaric chambers and managing them therein should not be underestimated.

Organophosphorus insecticides and nerve agents

Organophosphorus (OP) compounds (Box 9.13) are widely used as pesticides, especially in developing countries. The case fatality rate following deliberate ingestion of OP pesticides in developing countries in Asia is 5–20%.

<table>
<thead>
<tr>
<th>9.13 Organophosphorus compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nerve agents</strong></td>
</tr>
<tr>
<td>G agents: sarin, tabun, soman</td>
</tr>
<tr>
<td>V agents: VX, VE</td>
</tr>
<tr>
<td><strong>Insecticides</strong></td>
</tr>
<tr>
<td>Dimethyl compounds</td>
</tr>
<tr>
<td>Dichlorvos</td>
</tr>
<tr>
<td>Fenthion</td>
</tr>
<tr>
<td>Malathtin</td>
</tr>
<tr>
<td>Methamidophos</td>
</tr>
<tr>
<td>Diethyl compounds</td>
</tr>
<tr>
<td>Chlorpyrifos</td>
</tr>
<tr>
<td>Diazinon</td>
</tr>
<tr>
<td>Parathion-ethyl</td>
</tr>
<tr>
<td>Quinalphos</td>
</tr>
</tbody>
</table>

Nerve agents developed for chemical warfare are derived from OP insecticides but are much more toxic. They are commonly classified as G (originally synthesised in Germany) or V (‘venomous’) agents. The ‘G’ agents, such as tabun, sarin and soman, are volatile, absorbed by inhalation or via the skin, and dissipate rapidly after use. ‘V’ agents, such as VX, are contact poisons unless aerosolised, and contaminate ground for weeks or months.

The toxicology and management of nerve agent and pesticide poisoning are similar.

Mechanism of toxicity

OP compounds phosphorylate the active site of acetylcholinesterase (AChE), inactivating the enzyme and leading to the accumulation of acetylcholine (ACh) in cholinergic synapses (Fig. 9.3). Spontaneous hydrolysis of the OP–enzyme complex allows reactivation of the enzyme. However, loss of a chemical group from the OP–enzyme complex prevents further enzyme reactivation, a process termed ‘ageing’. After ageing has taken place, new enzyme needs to be synthesised before function can be restored. The rate of ageing is an important determinant of toxicity and is more rapid with dimethyl compounds (3.7 hours) than diethyl (31 hours), and especially rapid after exposure to nerve agents (soman in particular), which cause ageing within minutes.

Clinical features and management

OP poisoning causes an acute cholinergic phase, which may occasionally be followed by the intermediate syndrome or organophosphate-induced delayed polyneuropathy (OPIDN). The onset, severity and duration of poisoning depend on the route of exposure and agent involved. Cholinergic features may be prolonged over several weeks with some lipid-soluble agents.

Acute cholinergic syndrome

The acute cholinergic syndrome usually starts within a few minutes of exposure. Nicotinic or muscarinic features may be present (Box 9.14). Vomiting and profuse diarrhoea are typical following oral ingestion. Bronchoconstriction, bronchorrhea and salivation may cause severe respiratory compromise. Miosis is characteristic and the presence of muscle fasciculations strongly suggests the diagnosis, although this feature is often absent, even in serious poisoning. Subsequently, the patient may develop generalised flaccid paralysis which can affect respiratory and ocular muscles and result in respiratory failure. Ataxia, coma and convulsions may
9.14 Cholinergic features in poisoning*

<table>
<thead>
<tr>
<th>Cholinergic muscarinic</th>
<th>Cholinergic nicotinic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respiratory</strong></td>
<td></td>
</tr>
<tr>
<td>Bronchorrhoea, bronchoconstriction</td>
<td>Reduced ventilation</td>
</tr>
<tr>
<td><strong>Cardiovascular</strong></td>
<td></td>
</tr>
<tr>
<td>Bradycardia, hypotension</td>
<td>Tachycardia, hypertension</td>
</tr>
<tr>
<td><strong>CNS</strong></td>
<td></td>
</tr>
<tr>
<td>Confusion</td>
<td>Fasciculation, paralysis</td>
</tr>
<tr>
<td><strong>Temperature</strong></td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td></td>
</tr>
<tr>
<td><strong>Eyes</strong></td>
<td></td>
</tr>
<tr>
<td>Diplopia, mydriasis</td>
<td>Lacrimation, miosis</td>
</tr>
<tr>
<td><strong>Abdomen</strong></td>
<td></td>
</tr>
<tr>
<td>Ileus, palpable bladder</td>
<td>Vomiting, profuse diarrhoea</td>
</tr>
<tr>
<td><strong>Mouth</strong></td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>Salivation</td>
</tr>
<tr>
<td><strong>Skin</strong></td>
<td></td>
</tr>
<tr>
<td>Flushing, hot, dry</td>
<td>Sweating</td>
</tr>
<tr>
<td><strong>Complications</strong></td>
<td>Seizures</td>
</tr>
</tbody>
</table>

*Both muscarinic and nicotinic features occur in OP poisoning. Nicotinic features occur in nicotine poisoning and black widow spider bites. Cholinergic features are sometimes seen with some mushrooms.

Fig. 9.5 Mechanism of toxicity of organophosphorus compounds and treatment with oxime.

In severe poisoning, cardiac repolarisation abnormalities and torsades de pointes may occur. Other early complications of OP poisoning include extrapyramidal features, pancreatitis, hepatic dysfunction and pyrexia.

**Management**

The airway should be cleared of excessive secretions, breathing and circulation assessed, high-flow oxygen administered and intravenous access obtained. In the event of external contamination, further exposure should be prevented, contaminated clothing and contact lenses removed, the skin washed with soap and water, and the eyes irrigated. Gastric lavage or activated charcoal may be considered if the patient presents within 1 hour of ingestion. Convulsions should be treated as described in Box 9.10 (p. 211). The ECG, oxygen saturation, blood gases, temperature, urea and electrolytes, amylase and glucose should be monitored closely.

Early use of sufficient doses of atropine is potentially life-saving in patients with severe toxicity. Atropine reverses ACh-induced bronchospasm, bronchorrhoea, bradycardia and hypotension. When the diagnosis is uncertain, a marked increase in heart rate associated with skin flushing after a 1 mg intravenous dose makes OP poisoning unlikely. In OP poisoning, atropine should be administered in doses of 0.6–2 mg IV, repeated every 10–25 mins until secretions are controlled, the skin is dry and there is a sinus tachycardia. Large doses may be needed but excessive doses may cause anticholinergic effects (see Box 9.11, p. 213).

In patients requiring atropine, an oxime such as pralidoxime chloride (or obidoxime), if available, should also be administered, as this may reverse or prevent muscle weakness, convulsions or coma, especially if given rapidly after exposure. The pralidoxime dose for an adult is 2 g IV over 4 mins, repeated 4–6 times daily. Oximes work by reactivating AChE that has not undergone ‘ageing’ and are therefore less effective with dimethyl compounds and nerve agents, especially soman. Oximes may provoke hypotension, especially if administered rapidly.

Ventilatory support should be instituted before the patient develops respiratory failure (p. 193).
POISONING

Benzodiazepines may be used to reduce agitation and fasciculations, treat convulsions and sedate patients during mechanical ventilation.

Exposure is confirmed by measurement of plasma (butyrylcholinesterase) or red blood cell cholinesterase activity. These correlate poorly with the severity of clinical features, although values are usually less than 10% in severe poisoning, 20–50% in moderate poisoning and > 50% in subclinical poisoning.

The acute cholinergic phase usually lasts 48–72 hours, with most patients requiring intensive cardiorespiratory support and monitoring.

The intermediate syndrome

About 20% of patients with OP poisoning develop weakness that spreads rapidly from the ocular muscles to those of the head and neck, proximal limbs and the muscles of respiration, resulting in ventilatory failure. This ‘intermediate syndrome’ (IMS) generally develops quite rapidly between 1 and 4 days after exposure, often after resolution of the acute cholinergic syndrome, and may last 2–3 weeks. There is no specific treatment but supportive care, including maintenance of airway and ventilation, should be provided if necessary.

Organophosphate-induced delayed polyneuropathy

Organophosphate-induced delayed polyneuropathy (OPIDN) is a rare complication that usually occurs 2–3 weeks after acute exposure. It is a mixed sensory/motor polyneuropathy, especially affecting long myelinated neurons, and appears to result from inhibition of enzymes other than AChE. It is a feature of poisoning with some OPs such as trichlorocresylphosphate, but is less common with nerve agents,.. Early clinical features are muscle cramps followed by numbness and paraesthesia, proceeding to flaccid paralysis of the lower and subsequently the upper limbs. Paralysis of the lower limbs is associated with foot drop and a high-stepping gait, progressing to paraplegia. Paralysis of the arms leads to wrist drop. Sensory loss may also be present but is variable. Initially, tendon reflexes are reduced or lost but mild spasticity may develop later.

There is no specific therapy for OPIDN. Regular physiotherapy may limit deformity caused by muscle-wasting. Recovery is often incomplete and may be limited to the hands and feet, although substantial functional recovery after 1–2 years may occur, especially in younger patients.

Carbamate insecticides

Carbamate insecticides such as aldicarb, carbofuran and methomyl inhibit a number of tissue esterases, including AChE. The mechanism of action, clinical features and management are similar to those of OP compounds. However, clinical features tend to be less severe and of shorter duration, because the carbamate–AChE complex dissociates quickly, with a half-life of 30–40 minutes, and does not undergo ageing. Also, carbamates penetrate the CNS poorly. OPIDN and IMS are not common features of carbamate poisoning. Pancreatitis has been reported as a sequel, and deaths have occurred.

Atropine may be given intravenously in frequent small doses (0.6–2.0 mg IV for an adult) until signs of atropinisation develop. Diazepam may be used to relieve anxiety. The use of oximes is unnecessary.

**Methanol and ethylene glycol**

Ethylene glycol (1,2-ethanediol) is found in antifreeze, brake fluids and, in lower concentrations, windscreen washes. Methanol is present in some antifreeze products and commercially available industrial solvents, and in low concentrations in some screen washes and methylated spirits. It may also be an adulterant of illicitly produced alcohol. Both are rapidly absorbed after ingestion. Although methanol and ethylene glycol are not of high intrinsic toxicity, they are converted via alcohol dehydrogenase to toxic metabolites that are largely responsible for their clinical effects (Fig. 9.6).

**Clinical features**

Early features with either methanol or ethylene glycol include ataxia, drowsiness, dysarthria and nystagmus, often associated with vomiting. As the toxic metabolites are formed, metabolic acidosis, tachypnoea, coma and seizures may develop.

Toxic effects of ethylene glycol include ophthalmoplegia, cranial nerve palsies, hyporeflexia and myclonus. Renal pain and acute tubular necrosis occur because of precipitation of calcium oxalate in the kidneys. Hypocalcaemia, hypomagnesaemia and hyperkalaemia are common.

Features of methanol poisoning include headache, confusion and vertigo. Visual impairment and photophobia develop, associated with optic disc and retinal oedema and impaired pupil reflexes. Blindness may be permanent, although some recovery may occur over several months. Pancreatitis and abnormal liver function have also been reported.

**Management**

Urea, electrolytes, chloride, bicarbonate, glucose, calcium, magnesium, albumin and plasma osmolarity and arterial blood gases should be measured in all

---

![Fig. 9.6 Metabolism of methanol and ethylene glycol.](image-url)
patients with suspected methanol or ethylene glycol toxicity. The osmolar and anion gaps should be calculated (see Box 9.5, p. 209). Initially, poisoning is associated with an increased osmolar gap, but as toxic metabolites are produced, an increased anion gap associated with metabolic acidosis will develop. The diagnosis can be confirmed by measurement of ethylene glycol or methanol concentrations, but assays are not widely available.

An antidote, either ethanol or fomepizole, should be administered to all patients with suspected significant exposure while awaiting the results of laboratory investigations. These block alcohol dehydrogenase and delay the formation of toxic metabolites until the parent drug is eliminated in the urine or by dialysis. The antidote should be continued until ethylene glycol or methanol concentrations are undetectable. Metabolic acidosis should be corrected with sodium bicarbonate (e.g. 250 mL of 1.26% solution, repeated as necessary). Convulsions should be treated with an intravenous benzodiazepine. In ethylene glycol poisoning, hypocalcaemia should only be corrected if there are severe ECG features or seizures occur, since this may increase calcium oxalate crystal formation.

Haemodialysis or haemodiafiltration should be used in severe poisoning, especially if renal failure is present or there is visual loss in the context of methanol poisoning. It should be continued until acute toxic features are no longer present and ethylene glycol or methanol concentrations are no longer detectable.

### Aluminium and zinc phosphide

These rodenticides and fumigants are a common means of self-poisoning in northern India. The mortality rate for aluminium phosphide ingestion has been estimated at 60%; zinc phosphide ingestion appears less toxic, at about 2%. When ingested, both compounds react with gastric acid to form phosphine, a potent pulmonary and gastrointestinal toxicant. Clinical features include severe gastrointestinal disturbances, chest tightness, cough and breathlessness progressing to ARDS and respiratory failure, tremor, paraesthesiae, convulsions, coma, tachycardia, metabolic acidosis, electrolyte disturbances, hypoglycaemia, myocarditis, liver and renal failure, and leucopenia. Just a few tablets can be fatal.

Detection of phosphine in the exhaled air or stomach aspirate using either a silver nitrate-impregnated strip or specific phosphine detector tube is diagnostic, but gas chromatography provides the most specific indicator. Treatment is supportive and directed at correcting electrolyte abnormalities and treating complications; there is no specific antidote. Early gastric lavage is sometimes used, often with vegetable oil to reduce the release of toxic phosphine, but the benefit is uncertain.

### ENVIRONMENTAL POISONING AND ILLNESS

#### Arsenism

Chronic arsenic exposure from drinking water has been reported in many countries, especially India, Bangladesh, Nepal, Thailand, Taiwan, China, Mexico and South America, where a large proportion of the drinking water (ground water) has a high arsenic content, placing large population groups at risk. The World Health Organization (WHO) guideline value for arsenic content in tube well water is 10 µg/L.

Health effects associated with chronic exposure to arsenic in drinking water are shown in Box 9.15. In exposed individuals, high concentrations of arsenic are present in bone, hair and nails. Specific treatments are of no benefit in chronic arsenic toxicity and recovery from the peripheral neuropathy may never be complete. The emphasis should be on the prevention of exposure to arsenic in drinking water.

#### Fluorosis

Fluoride poisoning can result either from exposure to excessive quantities of fluoride (> 10 ppm) in drinking water or from industrial exposure to fluoride dust and consumption of brick tea. Clinical features include yellow staining and pitting of permanent teeth, osteosclerosis, soft tissue calcification, deformities (e.g. kyphosis) and joint ankylosis. Changes in the bones of the thoracic cage may lead to rigidity that causes dyspnoea on exertion. Very high doses of fluoride may cause abdominal pain, nausea, vomiting, seizures and muscle spasm. In calcium-deficient children, the toxic effects of fluoride manifest even at marginally high exposures to fluoride.

In endemic areas, such as Jordan, Turkey, Chile, India, Bangladesh, China and Tibet, fluorosis is a major public health problem. The maximum impact is seen in communities engaged in physically strenuous agricultural or industrial activities. Dental fluorosis is endemic in East Africa and some West African countries.
### SUBSTANCES LESS COMMONLY TAKEN IN OVERDOSE

Boxes 9.16 and 9.17 respectively give an overview of the clinical features and management for some substances that are less often encountered in overdose.

#### Table 9.16 Clinical features associated with substances taken less commonly in overdose

<table>
<thead>
<tr>
<th>Substances</th>
<th>Anticonvulsants</th>
<th>Sodium valproate</th>
<th>Isoniazid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamazepine, phenytoin</td>
<td>• Cerebellar signs</td>
<td>• Coma</td>
<td>• Peripheral neuropathy</td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td>• Metabolic acidosis</td>
<td>• Convulsions</td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>• Cardiac arrhythmias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coma</td>
<td>• Cardiac arrhythmias</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoniazid</td>
<td>• Peripheral neuropathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theophylline</td>
<td>• Cardiac arrhythmias</td>
<td>• Coma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrosives and bleach</td>
<td>• Abdominal pain</td>
<td>• GI perforation and late strictures</td>
<td>• Abdominal pain</td>
</tr>
<tr>
<td></td>
<td>• Microcytic anaemia with basophilic stippling</td>
<td>• Aspiration pneumonitis</td>
<td>• Headache</td>
</tr>
<tr>
<td></td>
<td>• Headache and encephalopathy</td>
<td></td>
<td>• Nausea, vomiting</td>
</tr>
<tr>
<td>Petroleum distillates, white spirit, kerosene</td>
<td>• Abdominal pain</td>
<td>• Motor neuropathy</td>
<td>• Respiratory depression</td>
</tr>
<tr>
<td></td>
<td>• Microcytic anaemia with basophilic stippling</td>
<td>• Nephrotoxicity</td>
<td>• Cardiac arrhythmias</td>
</tr>
<tr>
<td></td>
<td>• Headache and encephalopathy</td>
<td>• Hypertension</td>
<td>• Respiratory depression</td>
</tr>
<tr>
<td></td>
<td>• Abdominal pain</td>
<td>• Hypocalcaemia</td>
<td>• Hypertension</td>
</tr>
<tr>
<td>Lead, e.g. chronic occupational exposure, lead paint, water contaminated by lead pipes, use of kohl cosmetics</td>
<td>• Abdominal pain</td>
<td>• Motor neuropathy</td>
<td>• Hypocalcaemia</td>
</tr>
<tr>
<td></td>
<td>• Microcytic anaemia with basophilic stippling</td>
<td>• Nephrotoxicity</td>
<td>• Hypocalcaemia</td>
</tr>
<tr>
<td></td>
<td>• Headache and encephalopathy</td>
<td>• Hypertension</td>
<td>• Hypocalcaemia</td>
</tr>
<tr>
<td></td>
<td>• Abdominal pain</td>
<td>• Hypocalcaemia</td>
<td>• Hypocalcaemia</td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td>• Aspiration pneumonitis</td>
<td>• Hypocalcaemia</td>
</tr>
<tr>
<td>Paraquat</td>
<td>• Progressive respiratory fibrosis with respiratory failure</td>
<td>• Buccal ulceration</td>
<td>• Abdominal pain</td>
</tr>
<tr>
<td></td>
<td>• Abdominal pain</td>
<td>• Renal failure</td>
<td>• Nausea, vomiting</td>
</tr>
<tr>
<td>Organochlorines, e.g. DDT, lindane, dieldrin, endosulfan</td>
<td>• Microcytic anaemia with basophilic stippling</td>
<td>• Respiratory depression</td>
<td>• Respiratory depression</td>
</tr>
<tr>
<td></td>
<td>• Headache and encephalopathy</td>
<td>• Cardiac arrhythmias</td>
<td>• Cardiac arrhythmias</td>
</tr>
<tr>
<td></td>
<td>• Convulsions</td>
<td>• Hypertension</td>
<td>• Hypertension</td>
</tr>
<tr>
<td>Pyrethroid insecticides, e.g. cypermethrin, permethrin, imiprothrin</td>
<td>• Abnormal bleeding (prolonged)</td>
<td>• Respiratory depression</td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Skin contact: dermatitis, skin parasthesiae</td>
<td>• Cardiac arrhythmias</td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Eye contact: lacrimation, photophobia and oedema of the eyelids</td>
<td>• Hyperthermia</td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Inhalation: dyspnoea, nausea, headache</td>
<td>• Rhabdomyolysis</td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Ingestion: epigastric pain, nausea, vomiting, headache, coma, convulsions, pulmonary oedema</td>
<td>• Pulmonary oedema</td>
<td>• Hypertension</td>
</tr>
<tr>
<td>Anticoagulant rodenticides, (e.g. brodifacoum, bromodialone) and warfarin</td>
<td>• Abnormal bleeding (prolonged)</td>
<td>• Disseminated intravascular coagulation</td>
<td>• Hypertension</td>
</tr>
<tr>
<td></td>
<td>• Abnormal bleeding (prolonged)</td>
<td></td>
<td>• Hypertension</td>
</tr>
</tbody>
</table>

#### Table 9.17 Specific management of poisoning by substances taken less commonly in overdose

<table>
<thead>
<tr>
<th>Anticonvulsants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamazepine, phenytoin</td>
</tr>
<tr>
<td>• Multiple-dose activated charcoal (carbamazepine)</td>
</tr>
<tr>
<td>Sodium valproate</td>
</tr>
<tr>
<td>• Haemodialysis for severe poisoning</td>
</tr>
<tr>
<td>Isoniazid</td>
</tr>
<tr>
<td>• Activated charcoal</td>
</tr>
<tr>
<td>• IV pyridoxine</td>
</tr>
<tr>
<td>Theophylline</td>
</tr>
<tr>
<td>• Multiple-dose activated charcoal</td>
</tr>
<tr>
<td>Corrosives and bleach</td>
</tr>
<tr>
<td>• Gastric lavage and neutralising chemicals are contraindicated</td>
</tr>
<tr>
<td>• Chest X-ray to exclude perforation</td>
</tr>
<tr>
<td>• Consider early endoscopy or gastrografin studies to assess extent of damage and need for surgery</td>
</tr>
<tr>
<td>Lead</td>
</tr>
<tr>
<td>• Prevent further exposure</td>
</tr>
<tr>
<td>• Measure blood lead concentration, full blood count and blood film, urea and electrolytes, liver function tests and calcium</td>
</tr>
<tr>
<td>• Abdominal X-ray in children to detect pica</td>
</tr>
<tr>
<td>• Bone X-ray for ‘lead lines’</td>
</tr>
<tr>
<td>• Chelation therapy with dimercaprol, DMSA, DMPS or sodium calcium edetate for severe poisoning (esp. in children)</td>
</tr>
<tr>
<td>Petroleum distillates, white spirit, kerosene</td>
</tr>
<tr>
<td>• Gastric lavage contraindicated</td>
</tr>
<tr>
<td>• Activated charcoal ineffective</td>
</tr>
<tr>
<td>• Oxygen and nebulised bronchodilators</td>
</tr>
<tr>
<td>• Chest X-ray to assess pulmonary effects</td>
</tr>
<tr>
<td>Paraquat</td>
</tr>
<tr>
<td>• Urine screen for paraquat</td>
</tr>
<tr>
<td>• Multiple-dose activated charcoal</td>
</tr>
<tr>
<td>• Check blood paraquat concentration and compare with survival curve for prognosis</td>
</tr>
<tr>
<td>Organochlorines</td>
</tr>
<tr>
<td>• Activated charcoal (with nasogastric aspiration for liquid preparations) within 1 hour of ingestion</td>
</tr>
<tr>
<td>• Cardiac monitoring</td>
</tr>
<tr>
<td>Pyrethroid insecticides</td>
</tr>
<tr>
<td>• Symptomatic and supportive care</td>
</tr>
<tr>
<td>• Washing contaminated skin makes irritation worse</td>
</tr>
<tr>
<td>Anticoagulant rodenticides and warfarin</td>
</tr>
<tr>
<td>• Monitor INR/prothrombin time</td>
</tr>
<tr>
<td>• Vitamin K, by slow IV injection if coagulopathy</td>
</tr>
<tr>
<td>• Fresh frozen plasma or specific clotting factors for bleeding</td>
</tr>
</tbody>
</table>

(1) DMPS = 2,3-dimercaptopropane sulphonate; DMSA = 2,3-dimercaptopropanesulphonic acid

### ENVENOMING

Envenoming occurs when a venomous animal injects sufficient venom by a bite or a sting into a prey item or perceived predator to cause deleterious local and/or systemic effects. Venomous animals generally use their venom to acquire and in some cases predigest prey, with defensive use a secondary function. Accidental encounters between venomous animals and humans are
frequent, particularly in the rural tropics, where millions of cases of venomous bites and stings occur annually. Globally, an increasing number of exotic venomous animals are kept privately, so cases of envenomation may present to hospitals where doctors have insufficient knowledge to manage potentially complex presentations. Doctors everywhere should thus be aware of the basic principles of management of envenomation and how to seek expert support.

### Venom

Venom is a complex mixture of diverse components, often with several separate toxins that can cause adverse effects in humans, and each potentially capable of multiple effects (Box 9.18). Venom is produced at considerable metabolic cost, so is used sparingly; thus only some bites/stings by venomous animals result in significant envenomation, the remainder being ‘dry bites’. The concept of dry bites is important in understanding approaches to management.

#### Venous animals

There are many animal groups that contain venomous species (Box 9.19). The epidemiology estimates shown reflect the importance of snakes and scorpions as causes of severe or lethal envenomation. Social insect stings from bees and wasps may also cause lethal anaphylaxis. Other venomous animals may commonly envenom humans but cause mostly non-lethal effects. A few rarely envenom humans, but have a high potential for severe or lethal envenomation. These include box jellyfish, cone shells, blue-ringed octopus, paralysis ticks and Australian funnel web spiders. Within any given group, particularly snakes, there may be a wide range of clinical presentations. Some are described here but for a more detailed discussion of the types of

### Key venom effects

<table>
<thead>
<tr>
<th>Venom component</th>
<th>Clinical effects</th>
<th>Type of venomous animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurotoxin</td>
<td>Paralytic</td>
<td>Flaccid paralysis</td>
</tr>
<tr>
<td></td>
<td>Excitatory</td>
<td>Neuroexcitation: autonomic storm, cardiotoxicity, pulmonary oedema</td>
</tr>
<tr>
<td>Myotoxins</td>
<td>Systemic or local myolysis</td>
<td>Some snakes</td>
</tr>
<tr>
<td>Cardiotoxins</td>
<td>Direct or indirect cardiotoxicity; cardiac collapse, shock</td>
<td>Some snakes, scorpions, spiders, and jellyfish</td>
</tr>
<tr>
<td>Haemostasis system toxins</td>
<td>Vary from rapid coagulopathy and bleeding to thrombosis, deep venous thrombosis and pulmonary emboli</td>
<td>Many snakes and a few scorpions Brazilian caterpillars</td>
</tr>
<tr>
<td>Nephrotoxins</td>
<td>Renal damage</td>
<td>Some snakes, multiple bee, wasp stings</td>
</tr>
<tr>
<td>Necrotoxins</td>
<td>Local tissue injury/necrosis, shock</td>
<td>Some snakes, a few scorpions, spiders, jellyfish and stings</td>
</tr>
<tr>
<td>Allergic toxins</td>
<td>Induce acute allergic response (direct and indirect)</td>
<td>Almost all venoms but particularly those of social insects (bees, wasps, ants)</td>
</tr>
</tbody>
</table>

*All venom components have lethal potential.*

### Venous animals and human envenoming

<table>
<thead>
<tr>
<th>Phyla</th>
<th>Principal venomous animal groups</th>
<th>Estimated number of human cases/year</th>
<th>Estimated number of human deaths/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chordata</td>
<td>Snakes</td>
<td>&gt; 2.5 million</td>
<td>&gt; 100 000</td>
</tr>
<tr>
<td></td>
<td>Spiny fish</td>
<td>&gt; 100 000</td>
<td>Close to 0</td>
</tr>
<tr>
<td></td>
<td>Stingrays</td>
<td>&gt; 100 000</td>
<td>? &lt; 10</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>Scorpions</td>
<td>&gt; 1 million</td>
<td>? &lt; 5000</td>
</tr>
<tr>
<td></td>
<td>Spiders</td>
<td>&gt; 100 000</td>
<td>? &lt; 100</td>
</tr>
<tr>
<td></td>
<td>Paralysis ticks</td>
<td>&gt; 1000</td>
<td>? &lt; 10</td>
</tr>
<tr>
<td></td>
<td>Insects</td>
<td>? &gt; 1 million</td>
<td>? &gt; 1000*</td>
</tr>
<tr>
<td>Mollusca</td>
<td>Cone shells</td>
<td>? &lt; 1000</td>
<td>? &lt; 10</td>
</tr>
<tr>
<td></td>
<td>Blue-ringed octopus</td>
<td>? &lt; 100</td>
<td>? &lt; 10</td>
</tr>
<tr>
<td>Coelenterata</td>
<td>Jellyfish</td>
<td>? &gt; 1 million</td>
<td>? &lt; 10</td>
</tr>
</tbody>
</table>

*Social insect stings cause death by anaphylaxis rather than primary venom toxicity, except for massive multiple sting attacks.
venomous animal, their venoms and effects on humans see www.toxinology.com.

**Clinical effects**

With the exception of dry bites where no significant effects occur, venomous bites/stings can result in three broad classes of effect.

**Local effects**

These vary from trivial to severe (Box 9.20). There may be minimal or no local effects with some snakebites (not even pain), yet lethal systemic envenoming may still be present. For other species, local effects predominate over systemic, and for some species such as snakes, both are important.

**General systemic effects**

By definition, these are non-specific (see Box 9.20). Shock is an important complication of major local envenoming by some snake species and, if inadequately treated, can prove lethal, especially in children.

**Specific systemic effects**

These are important in both diagnosis and treatment.

- *Neurotoxic flaccid paralysis* can develop very rapidly, progressing from mild weakness to full respiratory paralysis in less than 30 minutes (blue-ringed octopus bite, cone shell sting), or may develop far more slowly, over hours (some snakes) to days (paralysis tick). For neurotoxic snakes, the cranial nerves are usually involved first, with ptosis a common initial sign (Fig. 9.7). From this, paralysis may extend to the limbs, with weakness and loss of deep tendon reflexes, then respiratory paralysis.
- *Excitatory neurotoxins* cause an ‘autonomic storm’, with profuse sweating, variable cardiac effects and cardiac failure, sometimes with pulmonary oedema (notably Australian funnel web spider bite, some scorpions such as Indian red scorpion). This type of envenoming can be rapidly fatal (many scorpions, funnel web spiders), or may cause distressing symptoms but constitute a lesser risk of death (widow spiders, banana spiders).

![Fig. 9.7 Ptosis following neurotoxin envenomation.](image)

- *Myotoxicity* can initially be silent, then present with generalised muscle pain, tenderness, myoglobinuria and huge rises in serum creatine kinase (CK). Secondary renal failure can precipitate potentially lethal hyperkalaemic cardiotoxicity.
- *Cardiotoxicity* is often secondary, but symptoms and signs are non-specific in most cases.
- *Haemostasis system toxins* cause a variety of effects, depending on the type of toxin, and the specific features can be diagnostic. Coagulopathy may present as bruising and bleeding from the bite site, gums and intravenous sites. Surgical interventions are high-risk in such cases. Other venoms cause thrombosis, usually presenting as deep venous thrombosis (DVT), pulmonary embolus or stroke (particularly Caribbean/Martinique vipers).
- *Renal damage* in envenoming is mostly secondary, although some species such as Russell’s vipers can cause primary renal damage. The presentation is similar in both cases, with changes in urine output (polyuria, oliguria or anuria) or rises in creatinine and urea. In cases with intravascular haemolysis, secondary renal damage is likely. The clinical effects of specific animals in different regions of the world are shown in Boxes 9.21–9.23.

### Management

It is important to determine an accurate diagnosis and the degree of risk, so that severe and potentially lethal cases are identified quickly and managed as a priority. With correct care, even severe cases are treatable, but delays in initiating effective treatment can severely compromise outcome. Expert advice should thus be sought at the earliest opportunity.

### First aid

Pre-hospital first aid (Box 9.24) can be critical in major envenoming. It depends on the type of envenoming, but the key principles are to:

- support vital systems
- delay or prevent the onset of envenoming
- avoid harmful ‘treatments’ such as electric shock, cut and suck, tourniquets, and cryotherapy in snakebite.

Many preventable deaths occur prior to hospital transfer when ineffective cardiopulmonary resuscitation is given to patients with respiratory paralysis or cardiac arrest/failure, which can occur due to either primary envenoming or an anaphylactic reaction (p. 91).
<table>
<thead>
<tr>
<th>Scientific name¹</th>
<th>Common name</th>
<th>Clinical effects</th>
<th>Antivenom/antidote/treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indian subcontinent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bungarus</em> spp. (E)</td>
<td>Kraits</td>
<td>Flaccid paralysis²³, myolysis⁴, hyponatraemia⁴</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Naja</em> spp. (E)</td>
<td>Cobras</td>
<td>Flaccid paralysis¹, local necrosis/ blistering, shock</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Ophiophagus hannah</em> (E)</td>
<td>King cobra</td>
<td>Flaccid paralysis¹, local necrosis, shock</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Echis</em> spp. (Nv)</td>
<td>Saw-scaled vipers</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, renal failure</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Daboia russelli</em> (Nv)</td>
<td>Russell’s viper</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, myolysis, renal failure, shock, flaccid paralysis²</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Hypnale</em> spp. (Vc)</td>
<td>Hump-nosed vipers</td>
<td>Procoagulant coagulopathy, shock, renal failure</td>
<td>Try Indian PV</td>
</tr>
<tr>
<td><em>Trimeresurus</em>² spp. (Vc)</td>
<td>Green pit vipers</td>
<td>Procoagulant coagulopathy, local necrosis, shock</td>
<td>Indian PV or specific</td>
</tr>
<tr>
<td><em>Mesobuthus</em> spp. (Sc)</td>
<td>Indian scorpions</td>
<td>Neuroexcitation, cardiotoxicity</td>
<td>Indian specific AV, prazosin</td>
</tr>
<tr>
<td><strong>East Asia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bungarus</em> spp. (E)</td>
<td>Kraits</td>
<td>Flaccid paralysis²³</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td><em>Naja</em> spp. (E)</td>
<td>Cobras (some spitters)</td>
<td>Flaccid paralysis¹, local necrosis/ blistering, shock</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td><em>Ophiophagus hannah</em> (E)</td>
<td>King cobra</td>
<td>Flaccid paralysis¹, local necrosis, shock</td>
<td>King cobra AV</td>
</tr>
<tr>
<td><em>Calloselasma rhodostoma</em> (Vc)</td>
<td>Malayan pit viper</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, renal failure, shock</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td><em>Daboia russelli</em> (Nv)</td>
<td>Russell’s viper</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, renal failure, shock</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td><em>Gloydius</em> (Vc)</td>
<td>Mamushis, pit vipers</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, shock, renal failure, flaccid paralysis²</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td><em>Trimeresurus</em>² (Vc)</td>
<td>Green pit vipers, habus</td>
<td>Procoagulant coagulopathy, local necrosis/blistering, shock</td>
<td>Specific AV from country</td>
</tr>
</tbody>
</table>

¹Family names: A = Atractaspidae; C = “Coburidae” (mostly ‘non-venomous’; family subject to major taxonomic revisions); E = Elapidae (all venomous); Sc = scorpions; Vc = Viperidae crotalinae (New World and Asian vipers); Vv = Viperidae viperae (Old World vipers). ²Pre-synaptic. ³Post-synaptic. ⁴Only reported so far for *B. candidus* and *B. multicinctus*. ⁵Genus is subject to major taxonomic change (split into at least eight genera). (AV = antivenom; PV = polyvalent)


## Diagnosis

Envenoming is usually obvious but might not be on some occasions. Humans may be bitten or stung by an unseen organism, or may not be aware of a bite or sting having occurred at all. In such cases the patient may present with a variety of symptoms but with no linking history to indicate envenoming. Accordingly, envenoming should be considered as a possible diagnosis in cases of unexplained paralysis, myotoxicity, coagulopathy, nephrotoxicity, cardiotoxicity, pulmonary oedema, necrosis, collapse and convulsions.

History, examination and laboratory findings help to confirm or exclude a diagnosis of envenoming and to determine its extent. It is also important to obtain a description of the organism if possible. Multiple bites or stings are more likely to cause major envenoming. Ask for specific symptoms and search for specific signs that may indicate the type and extent of envenoming (p. 207).

Specific tests for venom are currently only commercially available for Australian snakebite but are likely to be developed for snakebite in other regions. They are not available for other types of envenoming, where venom concentrations are low. For snakebite, a screen for envenoming includes full blood count, coagulation screen, urea and electrolytes, creatinine, CK and ECG. Lung function tests, peripheral oximetry or arterial blood gases may be indicated in cases with potential or established respiratory failure. In areas without access to routine laboratory tests, the whole-blood clotting time (using a glass test tube) is a valuable test for coagulopathy. A derivative of this, the 20-minute whole-blood clotting test is useful (a few millilitres of venous blood are placed in a glass vessel and checked for clotting at 20 minutes).

If patients state that they have been bitten by a particular species, ensure this is accurate. Private keepers of venomous animals may not have accurate
POISONING

9.22 Important venomous animals in the Americas and Australia

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>Common name</th>
<th>Clinical effects</th>
<th>Antivenom/antidote/treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>North America</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crotalus spp. (Vc)</td>
<td>Rattlesnakes</td>
<td>Procoagulant coagulopathy, local necrosis/ blistering (flaccid paralysis r)</td>
<td>CroFab AV® or Bioclon</td>
</tr>
<tr>
<td>Sistrurus spp. (Vc)</td>
<td>Massasaugas</td>
<td>Procoagulant coagulopathy, local necrosis/ blistering, shock</td>
<td>AntiVipmyn AV®</td>
</tr>
<tr>
<td>Agkistrodon spp. (Vc)</td>
<td>Copperheads and moccasins</td>
<td>Procoagulant coagulopathy, local necrosis/ blistering, shock</td>
<td>AntiVipmyn AV®</td>
</tr>
<tr>
<td>Micrurus spp. (E)</td>
<td>Coral snakes</td>
<td>Flaccid paralysis 2, myolysis, renal failure</td>
<td>Bioclon Coralymn AV®</td>
</tr>
<tr>
<td>Latrodectus mactans</td>
<td>Widow spider</td>
<td>Neuroexcitation</td>
<td>MSD Widow spider AV®</td>
</tr>
<tr>
<td>Centruroides sculpturatus</td>
<td>Arizona bark scorpion</td>
<td>Neuroexcitation</td>
<td>Bioclon Anascorp AV®</td>
</tr>
<tr>
<td><strong>Central and South America</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crotalus spp. (Vc)</td>
<td>Rattlesnakes</td>
<td>Flaccid paralysis 2, myolysis, procoagulant coagulopathy, shock, renal failure</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td>Bothrops spp. (Vc)</td>
<td>Lancehead vipers</td>
<td>Procoagulant coagulopathy, local necrosis/ blistering, shock, renal failure</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td>Bothriechis spp. (Vc)</td>
<td>Eyelash pit vipers</td>
<td>Shock, pain and swelling</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td>Lachesis spp. (Vc)</td>
<td>Bushmasters</td>
<td>Procoagulant coagulopathy, shock, renal failure, local necrosis/blistering</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td>Micrurus spp. (E)</td>
<td>Coral snakes</td>
<td>Flaccid paralysis 2, myolysis, renal failure</td>
<td>Specific AV from country</td>
</tr>
<tr>
<td>Tityus serrulatus</td>
<td>Brazilian scorpion</td>
<td>Neuroexcitation, shock</td>
<td>Instituto Butantan scorpion AV®</td>
</tr>
<tr>
<td>Loxosceles spp.</td>
<td>Recluse spiders</td>
<td>Local necrosis</td>
<td>Instituto Butantan spider AV®</td>
</tr>
<tr>
<td>Phoneutria nigriventer</td>
<td>Banana spider</td>
<td>Neuroexcitation, shock</td>
<td>Instituto Butantan spider AV®</td>
</tr>
<tr>
<td>Potamotrygon, Dasyatis spp.</td>
<td>Freshwater stingrays</td>
<td>Necrosis of bite area, shock, severe pain and oedema</td>
<td>No available AV; good wound care</td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudechis spp.</td>
<td>Black and mulga snakes</td>
<td>Anticoagulant coagulopathy, myolysis, renal failure</td>
<td>CSL black snake AV® or PVAV</td>
</tr>
<tr>
<td>Enhydrina schistosa</td>
<td>Sea snakes (all species globally)</td>
<td>Flaccid paralysis and/or myolysis</td>
<td>CSL sea snake AV®</td>
</tr>
<tr>
<td>Notcheis spp. (E)</td>
<td>Tiger snakes</td>
<td>Procoagulant coagulopathy, myolysis, flaccid paralysis r, renal failure</td>
<td>CSL tiger snake AV® or PVAV</td>
</tr>
<tr>
<td>Oxyuranus spp. (E)</td>
<td>Taipans</td>
<td>Procoagulant coagulopathy, flaccid paralysis r, myolysis, renal failure</td>
<td>CSL taipan AV® or PVAV</td>
</tr>
<tr>
<td>Acanthophis spp. (E)</td>
<td>Death adders</td>
<td>Flaccid paralysis 2</td>
<td>CSL death adder AV® or PVAV</td>
</tr>
<tr>
<td>Pseudechis spp.</td>
<td>Black and mulga snakes</td>
<td>Anticoagulant coagulopathy, myolysis, renal failure</td>
<td>CSL black snake AV® or PVAV</td>
</tr>
<tr>
<td>Enhydrina schistosa</td>
<td>Sea snakes (all species globally)</td>
<td>Flaccid paralysis and/or myolysis</td>
<td>CSL sea snake AV®</td>
</tr>
<tr>
<td>Atrax, Hadronyche spp.</td>
<td>Funnel web spiders</td>
<td>Neuroexcitation, shock</td>
<td>CSL funnel web spider AV®</td>
</tr>
<tr>
<td>Latrodectus hasseltii</td>
<td>Red back spider</td>
<td>Neuroexcitation, pain and sweating</td>
<td>CSL red back spider AV®</td>
</tr>
<tr>
<td>Chironex fleckeri</td>
<td>Box jellyfish</td>
<td>Neuroexcitation, cardiotoxicity, local necrosis</td>
<td>CSL box jellyfish AV®</td>
</tr>
<tr>
<td>Synancea spp.</td>
<td>Stonefish</td>
<td>Severe local pain</td>
<td>CSL stonefish AV®</td>
</tr>
</tbody>
</table>

*For family name, see Box 9.21. *Pre-synaptic. *Post-synaptic. (CSL = CSL Ltd, Melbourne, producer of Australian antivenoms)

knowledge of what they are keeping, and misidentification of a snake, scorpion or spider can have dire consequences if the wrong antivenom is used.

**Treatment**

Envenoming is managed on two levels, which must be delivered in tandem:

- supportive management of the organ systems affected and of the whole patient
- treating the effects with specific treatments/antidotes (usually antivenom).

For a snakebite by a potentially lethal species such as Russell’s viper, the patient might have local effects with oedema, blistering, necrosis, and resultant fluid shifts causing shock, and at the same time have systemic effects such as intractable vomiting, coagulopathy, paralysis and secondary renal failure. Specific treatment with antivenom will be required to reverse
### 9.23 Important venomous animals in Africa and Europe

<table>
<thead>
<tr>
<th>Scientific name¹</th>
<th>Common name</th>
<th>Clinical effects</th>
<th>Antivenom/antidote/treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Africa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Naja</em> spp. (E)</td>
<td>Cobras</td>
<td>Flaccid paralysis² ± local necrosis/blistering</td>
<td>South African PV or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>South African PV or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td></td>
<td>Non-spitters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spitters</td>
<td>Local necrosis/blistering (flaccid paralysis³ uncommon)</td>
<td>South African PV or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td><em>Dendroaspis</em> spp. (E)</td>
<td>Mambas</td>
<td>Mamba neurotoxin flaccid paralysis and muscle fasciculation, shock, necrosis (uncommon)</td>
<td>South African PV or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td><em>Hemachatus haemachatus</em> (E)</td>
<td>Rinkhals</td>
<td>Flaccid paralysis³, local necrosis, shock</td>
<td>South African PV</td>
</tr>
<tr>
<td><em>Atheris</em> spp. (Vv)</td>
<td>Bush vipers</td>
<td>Procoagulant coagulopathy, shock, pain and swelling</td>
<td>No available AV (can try South African AV)</td>
</tr>
<tr>
<td><em>Bitis</em> spp. (Vv)</td>
<td>Puff adders etc.</td>
<td>Procoagulant coagulopathy, shock, cardiotoxicity, local necrosis/blistering</td>
<td>South African PV or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td><em>Causus</em> spp. (Vv)</td>
<td>Night adders</td>
<td>Pain and swelling</td>
<td>No available AV</td>
</tr>
<tr>
<td><em>Echis</em> spp. (Vv)</td>
<td>Carpet vipers</td>
<td>Procoagulant coagulopathy, shock, renal failure, local necrosis/blistering</td>
<td>Specific anti-<em>Echis</em> AV for species/geographic region or Sanofi Pasteur FavAfrica AV®</td>
</tr>
<tr>
<td><em>Cerastes</em> spp. (Vv)</td>
<td>Horned desert vipers</td>
<td>Procoagulant coagulopathy, local necrosis, shock</td>
<td>Specific or polyspecific AV covering <em>Cerastes</em> from country of origin</td>
</tr>
<tr>
<td><em>Dispholidus</em> typus (C)</td>
<td>Boomslang</td>
<td>Procoagulant coagulopathy, shock</td>
<td>Boomslang AV</td>
</tr>
<tr>
<td><em>Androctonus</em> spp.</td>
<td>North African scorpions</td>
<td>Neuroexcitation</td>
<td>Specific scorpion AV (Algeria, Tunisia, Sanofi Pasteur Scorpifav®)</td>
</tr>
<tr>
<td><em>Leiurus</em> quinquestriatus</td>
<td>Yellow scorpion</td>
<td>Neuroexcitation, shock</td>
<td>Specific scorpion AV (Algeria, Tunisia, Sanofi Pasteur Scorpifav®)</td>
</tr>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Vipera</em> spp. (Vv)</td>
<td>Vipers and adders</td>
<td>Shock, local necrosis/blistering, procoagulant coagulopathy (flaccid paralysis² rare)</td>
<td>ViperaTab AV® or Zagreb AV® or SanofiPasteur Viperfav AV®</td>
</tr>
</tbody>
</table>

¹For family name, see Box 9.21. ²Pre-synaptic. ³Post-synaptic.

### 9.24 First aid for envenoming

<table>
<thead>
<tr>
<th>Method</th>
<th>Situations where indicated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Immobilisation of bitten limb</strong></td>
<td>All snakebites</td>
</tr>
<tr>
<td><strong>Pressure bandage and immobilisation</strong></td>
<td>Non-necrotic snakebites (Australian snakes, sea snakes, some cobras, king cobra, kraits, coral snakes, mambas, a few vipers, Australian funnel web spiders, blue-ringed octopus, cone shell)</td>
</tr>
<tr>
<td><strong>Local heat (hot water immersion or</strong></td>
<td>Venomous fish stings, stingray injuries, jellyfish stings (except possibly box jellyfish)</td>
</tr>
<tr>
<td><strong>shower, to 45°C)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Stauning local bleeding</strong></td>
<td>Traumatic injury with significant bleeding (some stingray injuries)</td>
</tr>
<tr>
<td><strong>Cardiorespiratory support</strong></td>
<td>Cardiac or respiratory impairment and particularly if respiratory paralysis is developing (some snakes, paralysis ticks, blue-ringed octopus, cone shells)</td>
</tr>
<tr>
<td><strong>No specific first aid</strong></td>
<td>Widow and recluso spider bites</td>
</tr>
</tbody>
</table>
the coagulopathy, and may prevent worsening of the paralysis and reduce the vomiting, but will not greatly affect the local tissue damage or the renal failure or shock. The latter will require intravenous fluid therapy, possibly respiratory support, renal dialysis and local wound care, possibly including antibiotics.

Each animal will cause a particular pattern of envenoming, requiring a tailored response. Listing all of these is beyond the scope of this chapter (see Further information below). Pulse, blood pressure, pulse oximetry and urine output should be monitored in all cases.

Antivenom

This is the most important tool in treating envenoming. It is made by hyperimmunising an animal, usually a horse, to produce antibodies against venom. Once refined, these bind to venom toxins and render them inactive or allow their rapid clearance. Antivenom is only available for certain venomous animals and cannot reverse all types of envenoming. With a few exceptions, it should be given intravenously, with adrenaline (epinephrine) ready in case of anaphylaxis. It should only be used when clearly indicated, and indications will vary between animals. It is critical that the correct antivenom is used at the appropriate dose. Doses vary widely between antivenoms; those recommended for North American antivenoms are not applicable to those elsewhere. In some situations (such as the Indian sub-continent), pre-treatment with subcutaneous adrenaline may reduce the chance of anaphylaxis to antivenom.

Antivenom can sometimes reverse post-synaptic neurotoxic paralysis (α-bungarotoxin-like neurotoxins) but will not usually reverse established pre-synaptic paralysis (β-bungarotoxin-like neurotoxins), so should be given before major paralysis has occurred. Coagulopathy is best reversed by antivenom, but even after all venom is neutralised, there may be a delay of hours before normal coagulation is restored. More antivenom should not be given because coagulopathy has failed to normalise fully in the first 1–3 hours (except in very particular circumstances). Thrombocytopenia may persist for days, despite antivenom. The role of antivenom in reversing established myolysis and renal failure is uncertain. Antivenom may help limit local tissue effects/injury in the bitten limb, but this is quite variable and time-dependent. Neuroexcitatory envenomation can respond very well to antivenom (Australian funnel web spider bites; Mexican, South American, Indian scorpion stings), but there is controversy about the effectiveness of antivenom for some species (some North African and Middle Eastern scorpions). The role of antivenom in limiting local venom effects, including necrosis, is also controversial; it is most likely to be effective when given early.

All patients receiving antivenom are at risk of both early and late adverse reactions, including anaphylaxis (early; not always IgE-related) and serum sickness (late).

Other treatments

Anticholinesterases are used as an adjunctive treatment for post-synaptic paralysis.

Prazosin (an α-adrenoceptor antagonist) is used in the management of hypertension or pulmonary oedema in scorpion sting cardiotoxicity, particularly for Indian red scorpion stings, though antivenom is now the preferred treatment.

Antibiotics are not routinely required for most bites/stings, though a few animals regularly cause significant wound infection/abscess, such as some South American pit vipers and stingrays. Tetanus is a risk in some bites or stings, such as snakebite, but intramuscular toxoid should not be given until any coagulopathy is reversed.

Mechanical ventilation (p. 194) is vital for established respiratory paralysis that will not reverse with antivenom, and may be required for prolonged periods – up to several months in some cases.

Follow-up

Cases with significant envenoming and those receiving antivenom should be followed up to ensure that any complications have resolved and to identify any delayed envenoming.

Further information and acknowledgements

Books and journal articles


WHO-SEARO. Guidelines for the management of snakebites (for SE Asia); 2010.

The last three documents can be accessed via www.toxinfo.org/Links.htm.

Websites

http://curriculum.toxicology.wikispaces.net/ Free access to educational material related to poisoning.

www.toxbase.org Toxbase, the clinical toxicology database of the UK National Poisons Information Service. Free for UK health professionals but registration is required. Access for overseas users by special arrangement.

www.toxinology.com Women’s and Children’s Hospital Adelaide Toxintoxology Department.


www.who.int/gho/peh/chemicalsafety/poisonscentres/en/ World directory of poisons centres held by the WHO, including interactive map and contact details.

Figure acknowledgements

Medical psychiatry

Classification of psychiatric disorders 232
Epidemiology of psychiatric disorders 232
Aetiology of psychiatric disorders 232
Diagnosing psychiatric disorders 233
Presenting problems in psychiatric illness 234
Anxiety symptoms 234
Depressed mood 235
Elated mood 235
Medically unexplained somatic symptoms 236
Delusions and hallucinations 236
Disturbed and aggressive behaviour 237
Confusion 238
Self-harm 238
Alcohol misuse 240
Substance misuse 240
Psychological factors affecting medical conditions 240

Treating psychiatric disorders 240
Biological treatments 240
Psychological treatments 241
Social interventions 242

Psychiatric disorders 242
Stress-related disorders 242
Anxiety disorders 242
Mood disorders 243
Somatoform disorders 245
Factitious disorders and malingering 247
Schizophrenia 247
Delirium, dementia and other organic disorders 250
Alcohol misuse and dependence 252
Substance misuse disorder 254
Personality disorders 255
Eating disorders 255
Puerperal disorders 256

Psychiatry and the law 257
Psychiatric disorders have traditionally been considered as ‘mental’ rather than as ‘physical’ illnesses. This is because they manifest with disordered functioning in the areas of emotion, perception, thinking and memory, and/or have had no clearly established biological basis. However, as research identifies abnormalities of the brain in an increasing number of psychiatric disorders and an important role for psychological and behavioural factors in many medical illnesses, a clear distinction between mental and physical illness has become increasingly questionable. We therefore refer to psychiatric disorders simply to mean those conditions traditionally regarded as the province of psychiatry.

**CLASSIFICATION OF PSYCHIATRIC DISORDERS**

There are two main classifications of psychiatric disorders in current use:
- the American Psychiatric Association’s Diagnostic and Statistical Manual (4th edition), or DSM-IV

The two systems are similar; here we use the ICD-10 classification (Box 10.1).

**EPIDEMIOLOGY OF PSYCHIATRIC DISORDERS**

Psychiatric disorders are amongst the most common of all human illnesses. The relative frequency of each varies with the setting (Box 10.2). In the general population, depression, anxiety disorders and adjustment disorders are most common (10%); in acute medical wards of general hospitals, organic disorders such as delirium (20–30%) are prevalent; in specialist general psychiatric services, psychoses are the most common disorders.

**AETIOLOGY OF PSYCHIATRIC DISORDERS**

The aetiology of psychiatric disorders is multifactorial, with a combination of biological, psychological and social causes. Each of these factors may play a role in predisposing to, precipitating or perpetuating the disorder (Box 10.3).

**Biological factors**

**Genetic**

Genetic factors play a predisposing role in many psychiatric disorders, including schizophrenia and bipolar...
Affective disorder. However, whilst some disorders such as Huntington’s disease are due to mutations in a single gene, the genetic contribution to most psychiatric disorders is polygenic in nature and mediated by the combined effects of several genetic variants, each with modest effects and modulated by environmental factors.

**Brain structure and function**

Brain structure is grossly normal in most psychiatric disorders, although abnormalities may be observed in some conditions, such as generalised atrophy in Alzheimer’s disease and enlarged ventricles with a slight decrease in brain size in schizophrenia. The functioning of the brain, however, is commonly altered with, for example, changes in neurotransmitters such as dopamine, noradrenaline (norepinephrine) and 5-hydroxytryptamine (5-HT, serotonin), and differences in activity of specific areas of the brain, as seen on functional brain scans.

**Psychological and behavioural factors**

**Early environment**

Early childhood adversity, such as emotional deprivation or abuse, predisposes to psychiatric disorders such as depression and eating disorders in adulthood.

**Personality**

The relationship between personality and psychiatric disorder can be difficult to assess because the development of psychiatric disorder can change a patient’s personality. However, some personality types predispose the individual to develop psychiatric disorder; for example, a depressive personality increases the risk of depression. A disordered personality may also perpetuate a psychiatric disorder once it is established, leading to a poorer prognosis.

**Behaviour**

A person’s behaviour may predispose to the development of a disorder (e.g. excess alcohol intake leading to dependence, and dieting to anorexia) or perpetuate it.

**Social and environmental factors**

**Social isolation**

The lack of a close, confiding relationship predisposes to some psychiatric disorders such as depression. The reduced social support resulting from having a psychiatric disorder may also act to perpetuate it.

**Stressors**

Social and environmental stressors often play an important role in precipitating psychiatric disorder in those who are predisposed. For example, trauma in post-traumatic stress disorder, losses (such as bereavement) in depression, and events perceived as threatening (such as potential loss of employment) in anxiety.

**Diagnosing psychiatric disorders**

Psychiatric assessment differs from a standard medical assessment in the following ways:

- There is greater emphasis on the history.
- It includes a systematic examination of the patient’s thinking, emotion and behaviour (mental state).
- It commonly includes the routine interviewing of an informant (usually a relative or friend who knows the patient), especially when the illness affects the patient’s ability to give an accurate history.

Because of its greater complexity, a full psychiatric history (Box 10.4) and detailed mental state examination (MSE) may take an hour or more. However, a brief mental state examination, usually taking no more than a few minutes (see below), should be part of the assessment of all patients, not merely those deemed to be ‘psychiatric’.

**Psychiatric interview**

The aims of the interview are to:

- establish a therapeutic relationship with the patient
- elicit the symptoms, history and background information (see Box 10.4)
- examine the mental state
- provide information, reassurance and advice.

Whilst some aspects of the patient’s mental state may be observed whilst the history is being taken, specific enquiries for important features should always be made.

**Mental state examination**

**General appearance and behaviour**

Any abnormalities of alertness or motor behaviour, such as restlessness or retardation, are noted. The level of consciousness should be determined, especially in the assessment of possible delirium.

**Speech**

Speed and fluency should be observed, including slow (retarded) speech and word-finding difficulty. ‘Pressure of speech’ describes rapid speech that is difficult to interrupt.
Abnormal beliefs
A delusion is a false belief, out of keeping with a patient’s cultural background, which is held with conviction despite evidence to the contrary (p. 236).

Abnormal perceptions
Illusions are abnormal perceptions of real stimuli. Hallucinations are sensory perceptions which occur in the absence of external stimuli: for example, hearing voices when no one is present (p. 237).

Cognitive function
The Mini-Mental State Examination (MMSE) is a useful screening questionnaire to detect cognitive impairment. A score of less than 24 out of 30 typically suggests cognitive impairment. The Addenbrooke’s Cognitive Examination—Revised (ACE-R) provides a more comprehensive assessment. A brief clinical assessment is as follows:

• Memory. Registration of memories is tested by asking the patient to repeat simple new information, such as a name and address, immediately after hearing it. Short-term memory is assessed by asking him or her to repeat it after an interval of 1–2 minutes, during which time the patient’s attention should be diverted elsewhere. Long-term memory is assessed by gauging the recall of previous events.

• Concentration. Serial 7s is a test in which the patient is asked to subtract 7 from 100 and then 7 from the answer, and so on.

• Orientation. This is assessed by asking the patient about place – his or her exact location; time – what day, date, month and year it is now; and person – details of personal identity, such as name, date of birth, marital status and address.

• Intellectual ability. This can be gauged from the history of the patient’s educational background and attainments but can also be assessed during the interview from the patient’s speech, vocabulary and grasp of the interviewer’s questions.

Note that the degree of cognitive impairment in delirium typically fluctuates over time, and consequently may be missed by a single assessment.

Patients’ own understanding of their symptoms (‘insight’)
Patients should be asked what they think their symptoms are due to, and whether they warrant treatment. Lack of insight refers to a failure to accept that one is ill and/or in need of treatment, and is characteristic of acute psychosis.

Mood
This can be judged by facial expression, posture and movements. Patients should also be asked if they feel sad or depressed and if they lack ability to experience pleasure (anhedonia). Are they anxious, worried or tense? Is mood elevated with excess energy and a reduced need for sleep, as in mania?

Thoughts
The content of thought can be elicited by asking ‘What are your main concerns?’ Is thinking negative, guilty or hopeless, suggesting depression? Are there thoughts of self-harm? If so, enquiry should be made about plans. Is he or she excessively worried about many things, suggesting anxiety? Does the patient think that he or she is especially powerful, important or gifted (grandiose thoughts), suggesting mania?

The form of thinking may also be abnormal. In schizophrenia, patients may display loosened associations between ideas, making it difficult to follow their train of thought. There may also be abnormalities of thought possession, when patients experience the intrusion of alien thoughts into their mind or the broadcasting of their own thoughts to other people (p. 247).
Depressed mood

Depressive disorder is common, with a prevalence of approximately 5% in the general population. Depression is at least twice as common in the medically ill. It is important to note that depression has physical as well as mental symptoms (Box 10.7). The diagnosis of depression in the medically ill, who may have physical symptoms of disease, relies on detection of the core psychological symptoms of low mood and anhedonia.

Differential diagnosis

Depressive disorder must be differentiated from an adjustment disorder with depressed mood (p. 242). Adjustment disorders are common, self-limiting reactions to adversity, including physical illness, which are transient and require only general support. Depressive disorders (p. 243) are characterised by a more severe and persistent disturbance of mood and require specific treatment. In some cases, depression may occur as a result of a direct effect of a medical condition or its treatment on the brain, when it is referred to as an ‘organic mood disorder’ (Box 10.8).

Suicide

Depression is the major risk factor for suicide. Other risk factors are shown in Box 10.9. When depression is suspected, tactful enquiry should always be made into suicidal thoughts and plans. Asking about suicide does not increase the risk of it occurring, whereas failure to enquire denies the opportunity to prevent it.

Elated mood

Elation, or euphoria, is the converse of depression and is characteristic of mania. It may manifest as infectious joviality, over-activity, lack of sleep and appetite, undue optimism, over-talkativeness, irritability, and recklessness in spending and sexual behaviour. When elated mood is severe, psychotic symptoms are often evident, such as delusions of grandeur (e.g. believing erroneously that one is royalty). Elevated mood is much less

<table>
<thead>
<tr>
<th>10.5 Symptoms of anxiety disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Psychological</strong></td>
</tr>
<tr>
<td>- Apprehension</td>
</tr>
<tr>
<td>- Irritability</td>
</tr>
<tr>
<td>- Worry</td>
</tr>
<tr>
<td><strong>Somatic</strong></td>
</tr>
<tr>
<td>- Palpitations</td>
</tr>
<tr>
<td>- Fatigue</td>
</tr>
<tr>
<td>- Tremor</td>
</tr>
<tr>
<td>- Dizziness</td>
</tr>
<tr>
<td>- Sweating</td>
</tr>
<tr>
<td>- Diarrhoea</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10.7 Symptoms of depressive disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Psychological</strong></td>
</tr>
<tr>
<td>- Depressed mood</td>
</tr>
<tr>
<td>- Reduced self-esteem</td>
</tr>
<tr>
<td>- Pessimism</td>
</tr>
<tr>
<td>- Guilt</td>
</tr>
<tr>
<td><strong>Somatic</strong></td>
</tr>
<tr>
<td>- Reduced appetite</td>
</tr>
<tr>
<td>- Weight change</td>
</tr>
<tr>
<td>- Disturbed sleep</td>
</tr>
<tr>
<td>- Fatigue</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10.8 Organic mood disorders*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neurological</strong></td>
</tr>
<tr>
<td>- Cerebrovascular disease</td>
</tr>
<tr>
<td>- Cerebral tumour</td>
</tr>
<tr>
<td>- Multiple sclerosis</td>
</tr>
<tr>
<td>- Parkinson’s disease</td>
</tr>
<tr>
<td>- Huntington’s disease</td>
</tr>
<tr>
<td>- Alzheimer’s disease</td>
</tr>
<tr>
<td>- epilepsies</td>
</tr>
<tr>
<td><strong>Endocrine</strong></td>
</tr>
<tr>
<td>- Hypothyroidism</td>
</tr>
<tr>
<td>- Hyperthyroidism</td>
</tr>
<tr>
<td>- Cushing’s syndrome</td>
</tr>
<tr>
<td>- Addison’s disease</td>
</tr>
<tr>
<td>- Hyperparathyroidism</td>
</tr>
<tr>
<td><strong>Malignant disease</strong></td>
</tr>
</tbody>
</table>

*Diseases that may cause organic affective disorders by direct action on the brain.
common than depressed mood, and in medical settings is often secondary to drug or alcohol misuse, an organic disorder or medical treatment. Where none of these applies, the patient may have a bipolar disorder (p. 244).

**Medically unexplained somatic symptoms**

Patients commonly present to doctors with physical symptoms. Whilst these symptoms may be an expression of a medical condition, they often are not. They may then be referred to as ‘medically unexplained symptoms’ (MUS). MUS are very common in patients attending general medical outpatient clinics. Almost any symptom can be medically unexplained, e.g.:

- pain (including back, chest, abdominal and headache)
- fatigue
- dizziness
- fits, ‘funny turns’ and feelings of weakness.

Patients with MUS may receive a medical diagnosis of a so-called functional somatic syndrome, such as irritable bowel syndrome (Box 10.10), and may also merit a psychiatric diagnosis on the basis of the same symptoms. The most frequent psychiatric diagnoses associated with MUS are anxiety or depressive disorders. When these are absent, a diagnosis of somatoform disorder may be appropriate (Box 10.11).

**Differential diagnosis**

The main medical differential diagnosis for MUS is from symptoms of a medical disease. Diagnostic difficulties are most likely with unusual presentations of common diseases and with rare diseases. MUS are commonly an expression of depression and anxiety. A medical and psychiatric assessment should be completed in all cases (Fig. 10.1).

**Delusions and hallucinations**

**Delusions**

Various types of delusion are identified on the basis of their content. They may be:

- persecutory, such as a conviction that others are out to get me
• hypochondriacal, such as an unfounded conviction that one has cancer
• grandiose, such as a belief that one has special powers or status
• nihilistic, e.g. ‘My head is missing’, ‘I have no body’, ‘I am dead’.

Delusions should be differentiated from over-valued ideas, which are strongly held but not fixed.

**Hallucinations**

These are perceptions without external stimuli. They can occur in any sensory modality, most commonly visual or auditory. Typical examples are hearing voices when no one else is present, or seeing ‘visions’. Hallucinations have the quality of ordinary perceptions and are perceived as originating in the external world, not in the patient’s own mind (when they are termed pseudo-hallucinations). Those occurring when falling asleep (‘hypnagogic’) and on waking (‘hypnopompic’) are not pathological. Hallucinations should be distinguished from illusions, which are misperceptions of real external stimuli (such as mistaking a shrub for a person in poor light).

**Differential diagnosis**

Agitation, terror or the fear of being thought ‘mad’ may make patients unable or unwilling to volunteer or describe their abnormal beliefs or experiences. Careful and tactful enquiry is therefore required. The nature of hallucinations can be important diagnostically; for example, ‘running commentary’ voices that discuss the patient are strongly associated with schizophrenia. In general, auditory hallucinations suggest schizophrenia, while hallucinations in other sensory modalities, especially vision but also taste and smell, suggest an ‘organic psychosis’ such as delirium or temporal lobe epilepsy.

Hallucinations and delusions often co-occur; if their content is consistent with coexisting emotional symptoms, they are described as ‘mood-congruent’. Thus, patients with severely depressed mood may believe themselves responsible for all the evils in the world, and hear voices saying ‘You’re worthless. Go and kill yourself’. In this case, the diagnosis of depressive psychosis is made on the basis of the congruence of different phenomena (mood, delusion and hallucination). Incongruence between hallucinations, delusions and mood suggests schizophrenia.

Where hallucinations and delusions arise within disturbed consciousness and impaired cognition, the diagnosis is usually an organic disorder, most commonly delirium and/or dementia (p. 244). This differential diagnosis is made by assessing the nature, extent and time course of any cognitive disturbances, and by investigating for underlying causes.

**Disturbed and aggressive behaviour**

Disturbed and aggressive behaviour is common in general hospitals, especially in emergency departments. Most behavioural disturbance arises not from medical or psychiatric illness, but from alcohol intoxication, reaction to the situation and personality characteristics. The key principles of management are, first, to establish control of the situation rapidly and thereby ensure the safety of the patient and others, and, second, to assess the cause of the disturbance in order to remedy it. Establishing control requires the presence of an adequate number of trained staff, an appropriate physical environment and sometimes sedation (Fig. 10.2). Hospital security staff and sometimes the assistance of the police may be required. In all cases, the staff approach is important; a calm, non-threatening manner expressing understanding of the patient’s concerns is often all that is required to defuse potential aggression (Box 10.12).

If sedating drugs are required, antipsychotic drugs, such as haloperidol, and benzodiazepines, such as diazepam, are commonly used. The choice of drug, dose, route and rate of administration will depend on the

---

**Fig. 10.2 Acute management of disturbed behaviour.**
patient’s age, sex and physical health, as well as the likely cause of the disturbed behaviour. The benefits of sedation must be balanced against the associated risks, however. Haloperidol can cause acute dystonias, including oculogyric crises, while the benzodiazepines can precipitate respiratory depression in patients with lung disease, and encephalopathy in those with liver disease. Thus, for a frail elderly woman with emphysema and delirium, sedation may be achieved with a low dose (0.5 mg) of oral haloperidol, while for a strong young man with an acute psychotic episode, at least 10 mg of intravenous diazepam and a similar dose of haloperidol may be needed. A parenterally administered anticholinergic agent, such as procyclidine, should be available to treat extrapyramidal effects arising from haloperidol, and flumazenil (p. 217) to reverse respiratory depression if large doses of benzodiazepines are used.

**Differential diagnosis**

Many factors may contribute to disturbed behaviour. When the patient is cooperative, these are best determined at interview. Other sources of information about the patient include medical and psychiatric records, and discussion with nursing staff, family members and other informants, including the patient’s general practitioner. The following information should be sought:

- psychiatric, medical (especially neurological) and criminal history
- current psychiatric and medical treatment
- alcohol and drug misuse
- recent stressors
- the time course and accompaniments of the current episode in terms of mood, belief and behaviour.

Observation of the patient’s behaviour may also yield useful clues. Do they appear to be responding to hallucinations? Are they alert or variably drowsy and confused? Are there physical features suggestive of drug or alcohol misuse or withdrawal? Are there new injuries or old scars, especially on the head? Do they smell of alcohol or solvents? Do they bear the marks of drug injection? Are they unwashed and unkempt, suggesting a gradual development of their condition?

If the person has an acute psychiatric disorder, then admission to a psychiatric facility may be indicated. If a medical cause is likely, psychiatric transfer is usually inappropriate and the patient should be managed in a medical setting, with whatever nursing and security support is required. Where it is clear that there is no medical or psychiatric illness, the person should be removed from the hospital, to police custody if necessary.

Measures such as restraint, sedation, the investigation and treatment of medical problems, and psychiatric transfer all raise legal as well as medical issues (p. 257). In most countries, including the UK, common law confers upon doctors the right, and indeed the duty, to intervene against a patient’s wishes in cases of acute behavioural disturbance, if this is necessary to protect the patient or other people. Many countries, such as the UK, also have specific mental health legislation that may be used to detain patients.

**Confusion**

This is a vague term used to describe a range of primarily cognitive problems, including disturbances in perception, belief and behaviour. ‘Confusion’ usually presents as a problem when it becomes clear that the patient cannot comply with medical care; they may repeatedly wander off the ward, pull out essential cannulae and catheters, and hit nurses. The methods of assessment of cognitive function range from simple screening questions to detailed psychometric testing. All doctors should be able to undertake a brief cognitive assessment, as outlined above (p. 233).

**Differential diagnosis**

A history from the patient and informants is essential to establish the time course, variability and functional consequences of any cognitive deficit. Mental state examination is necessary to seek evidence of associated mood disorder, hallucinations, delusions or behavioural abnormalities, and physical examination to identify any relevant medical conditions. The assessment should seek to distinguish between:

- organic disorders such as delirium, dementia, and focal deficits secondary to brain lesions
- psychiatric disorders such as depressive pseudodementia and dissociative disorder
- malingering (p. 247).

Further investigation will usually be needed to identify the specific causes of any cognitive impairment identified (see Box 10.32, p. 250, and p. 209).

**Self-harm**

Self-harm (SH) is a common reason for presentation to medical services. The term ‘attempted suicide’ is
potentially misleading, as most such patients are not unequivocally trying to kill themselves. Most cases of SH involve overdose, of either prescribed or non-prescribed drugs (Ch. 9). Less common methods include asphyxiation, drowning, hanging, jumping from a height or in front of a moving vehicle, and the use of firearms. Methods that carry a high chance of being fatal are more likely to be associated with serious psychiatric disorder. Self-cutting is common and often repetitive, but rarely leads to contact with medical services.

The incidence of SH varies over time and between countries. In the UK, the lifetime prevalence of suicidal ideation is 15% and that of acts of SH is 4%. SH is more common in women than men, and in young adults than the elderly. (In contrast, completed suicide is more common in men and the elderly (see Box 10.9).) There is a higher incidence of self-harm among lower socioeconomic groups, particularly those living in crowded, socially deprived urban areas. There is also an association with alcohol misuse, child abuse, unemployment and recently broken relationships.

Differential diagnosis

The main differential diagnosis is from accidental poisoning and so-called ‘recreational’ overdose in drug users. It must be remembered that SH is not a diagnosis but a presentation, and may be associated with any psychiatric diagnosis, the most common being adjustment disorder, substance and alcohol misuse, depressive disorder and personality disorder. In many cases, however, no psychiatric diagnosis can be made.

Initial management

A thorough psychiatric and social assessment should be attempted in all cases (Fig. 10.3), although some patients will discharge themselves before this can take place. The need for psychiatric assessment should not, however, delay urgent medical or surgical treatment, and may need to be deferred until the patient is well enough for interview. The purpose of the psychiatric assessment is to:

• establish the short-term risk of suicide
• identify potentially treatable problems, whether medical, psychiatric or social.

Topics to be covered when assessing a patient are listed in Box 10.14. The history should include events occurring immediately before and after the act, and especially any evidence of planning. The nature and severity of any current psychiatric symptoms must be assessed, along with the personal and social supports available to the patient outside hospital.

Most SH patients have depressive and anxiety symptoms on a background of chronic social and personal difficulties and alcohol misuse but no psychiatric disorder. They do not usually require psychotropic

---

**Box 10.14 Assessment of patients after self-harm**

<table>
<thead>
<tr>
<th>Current attempt</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Patient’s account</td>
</tr>
<tr>
<td>• Degree of intent at the time: preparations, plans, precautions against discovery, note</td>
</tr>
<tr>
<td>• Method used, particularly whether violent</td>
</tr>
<tr>
<td>• Degree of intent now</td>
</tr>
<tr>
<td>• Symptoms of psychiatric illness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Previous attempts and their outcome</td>
</tr>
<tr>
<td>• Family and personal history</td>
</tr>
<tr>
<td>• Social support</td>
</tr>
<tr>
<td>• Previous response to stress</td>
</tr>
<tr>
<td>• Extent of drug and alcohol misuse</td>
</tr>
</tbody>
</table>

**Fig. 10.3** Assessment of patients admitted following self-harm (SH).
medication or specialised psychiatric treatment but may benefit from personal support and practical advice from a GP, social worker or community psychiatric nurse. Admission to a psychiatric ward is necessary only for persons who:

- have an acute psychiatric disorder
- are at high risk of suicide
- need temporary respite from intolerable circumstances
- require further assessment of their mental state.

Approximately 20% of SH patients make a repeat attempt during the following year and 1–2% kill themselves. Factors associated with suicide after an episode of SH are listed in Box 10.9.

### Alcohol misuse

Misuse of alcohol is a major problem worldwide. It presents in a multitude of ways, which are discussed further on page 252 and in Box 10.35 (p. 253). In many cases, the link to alcohol will be all too obvious; in others, it may not be. Denial and concealment of alcohol intake are common. In the assessment of alcohol intake, the patient should be asked to describe a typical week’s drinking, quantified in terms of units of alcohol (1 unit contains approximately 8 g alcohol and is the equivalent of half a pint of beer, a single measure of spirits or a small glass of wine). Drinking becomes hazardous at levels above 21 units weekly for men and 14 units weekly for women. The history from the patient may need corroboration by the GP, earlier medical records and family members. The mean cell volume (MCV) and γ-glutamyl transferase (GGT) may be raised, but are abnormal in only half of problem drinkers; consequently, normal results on these tests do not exclude an alcohol problem. When abnormal, these measures may be helpful in challenging denial and monitoring treatment response. The prevention and management of alcohol-related problems are discussed on page 253.

### Substance misuse

The misuse of drugs of all kinds is also widespread. As well as the general headings listed for alcohol problems in Box 10.35 (p. 253), there are two additional sets of problems associated with drug misuse:

- problems linked with the route of administration, such as intravenous injection
- problems arising from pressure applied to doctors to prescribe the misused substances (Box 10.15).

Assessment and management are described on page 254.

### Psychological factors affecting medical conditions

Psychological factors may influence the presentation, management and outcome of medical conditions. Specific factors are shown in Box 10.16. The most common psychiatric diagnoses in the medically ill are anxiety and depressive disorders. Often these appear understandable as adjustments to illness and its treatment; however, if the anxiety and depression are severe and persistent, they may complicate the management of the medical condition and active management is required. Anxiety may present as an increase in somatic symptoms such as breathlessness, tremor or palpitations, or as the avoidance of medical treatment. It is most common in those facing difficult or painful treatments, deterioration of their illness or death. Depression may manifest as increased physical symptoms such as pain or fatigue and disability, as well as with depressed mood and loss of interest and pleasure. It is most common in patients who have suffered actual or anticipated losses, such as receiving a terminal diagnosis or undergoing disfiguring surgery.

Treatment is by psychological and/or pharmacological therapies, as described below. Care is required when prescribing psychotropic drugs to the medically ill in order to avoid exacerbation of the medical condition and harmful interactions with other prescribed drugs.

### TREATING PSYCHIATRIC DISORDERS

The multifactorial origin of most psychiatric disorders means that there are multiple potential targets for treatment.

#### Biological treatments

These aim to relieve psychiatric disorder by modifying brain function. The main biological treatments are psychotropic drugs. These are widely used for various
Treating psychiatric disorders

10.17 Classification of commonly used psychotropic drugs

<table>
<thead>
<tr>
<th>Action</th>
<th>Main groups</th>
<th>Clinical use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipsychotic</td>
<td>Phenothiazines</td>
<td>Schizophrenia</td>
</tr>
<tr>
<td></td>
<td>Butyrophenones</td>
<td>Bipolar mania</td>
</tr>
<tr>
<td></td>
<td>Second-generation antipsychotics</td>
<td>Acute confusion</td>
</tr>
<tr>
<td>Antidepressant</td>
<td>Tricyclics and related drugs</td>
<td>Depression/anxiety</td>
</tr>
<tr>
<td></td>
<td>Monoamine oxidase inhibitors</td>
<td>Obsessive-compulsive disorder</td>
</tr>
<tr>
<td></td>
<td>Serotonin and noradrenergic re-uptake inhibitors</td>
<td>Depression/anxiety</td>
</tr>
<tr>
<td>Mood-stabilising</td>
<td>Lithium (Semi-)Sodium valproate</td>
<td>Treatment and prophylaxis of bipolar disorder</td>
</tr>
<tr>
<td>Anti-anxiety</td>
<td>Benzodiazepines</td>
<td>Anxiety/insomnia (short term)</td>
</tr>
<tr>
<td></td>
<td>β-adrenoceptor antagonists</td>
<td>Alcohol withdrawal (short term)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anxiety (somatic symptoms)</td>
</tr>
</tbody>
</table>

purposes; a pragmatic classification is set out in Box 10.17. It should be noted that some drugs have applications to more than one condition; for example, antidepressants are also widely used in the treatment of anxiety and chronic pain. The specific subgroups of psychotropic drugs are discussed in the sections on the appropriate disorders.

Electroconvulsive therapy (ECT) entails producing a convulsion by the administration of high-voltage, brief, direct-current impulses to the head while the patient is anaesthetised and paralysed by muscle relaxant. If properly administered, it is remarkably safe, has few side-effects, and is of proven efficacy for severe depressive illness. There may be amnesia for events occurring a few hours before ECT (retrograde) and after it (anterograde). Pronounced amnesia can occur but is infrequent and difficult to distinguish from the effects of severe depression. Surgery to the brain (psychosurgery) has a very limited place and then only in the treatment of severe chronic psychiatric illness resistant to other measures.

Psychological treatments

These treatments are useful in many psychiatric disorders and also in non-psychiatric conditions. They are based on talking with patients, either individually or in groups. Sometimes discussion is supplemented by ‘homework’ or tasks to complete between treatment sessions. Psychological treatments take a number of forms based on the duration and frequency of contact, the specific techniques applied and their underlying theory.

General or supportive psychotherapy

This should be part of all medical treatment. It involves empathic listening to the patient’s account of their symptoms and associated fears and concerns, followed by the sympathetic provision of accurate information that addresses these.

Cognitive therapy

This therapy is based on the observation that some psychiatric disorders are associated with systematic errors in the patient’s conscious thinking; for example, a tendency to interpret events in a negative way or see them as unduly threatening. A triad of ‘cognitive errors’ has been described in depression (Box 10.18). Cognitive therapy aims to help patients to identify such cognitive errors and to learn how to challenge them. It is widely used for depression, anxiety and eating and somatoform disorders, and also increasingly in psychoses.

Behaviour therapy

This is a practically orientated form of treatment, in which patients are assisted in changing unhelpful behaviour: for example, helping patients to implement carefully graded exposure to the feared stimulus in phobias.

Cognitive behaviour therapy

Cognitive behaviour therapy (CBT) combines the methods of behaviour therapy and cognitive therapy. It is the most widely available and extensively researched psychological treatment.

Problem-solving therapy

This is a simplified brief form of CBT, which helps patients actively tackle problems in a structured way (Box 10.19). It is of benefit in mild to moderate depression, and can be delivered by non-psychiatric doctors and nurses after appropriate training.

Psychodynamic psychotherapy

This treatment, also known as ‘interpretive psychotherapy’, was pioneered by Freud, Jung and Klein, amongst others. It is based upon the theory that early life experience generates powerful but unconscious motivations. Psychotherapy aims to help the patient to
become aware of these unconscious factors on the assumption that, once identified, their negative effects are reduced. The relationship between therapist and patient is used as a therapeutic tool to identify issues in patients’ relationships with others, particularly parents, which may be replicated or transferred to their relationship with the therapist. Explicit discussion of this relationship (transference) is the basis for the treatment, which traditionally requires frequent sessions over a period of months or even years.

Interpersonal psychotherapy

Interpersonal psychotherapy (IPT) is a specific form of brief psychotherapy that focuses on patients’ current interpersonal relationships and is an effective treatment for mild to moderate depression.

Social interventions

Some adverse social factors, such as unemployment, may not be readily amenable to intervention, but others, such as access to benefits and poor housing, may be. Patients can be helped to address these problems themselves by being taught problem-solving. Befrienders and day centres can reduce social isolation, benefits advisers can ensure appropriate financial assistance, and medical recommendations can be made to local housing departments to help patients obtain more appropriate accommodation.

PSYCHIATRIC DISORDERS

Stress-related disorders

Acute stress reaction

Following a stressful event such as a serious medical diagnosis or a major accident, some people develop a characteristic pattern of symptoms. These include a sense of bewilderment, anxiety, anger, depression, altered activity and withdrawal. The symptoms are transient and usually resolve completely within a few days.

Adjustment disorder

A more common psychological response to a major stressor is a less severe but more prolonged emotional reaction. The predominant symptom is usually depression and/or anxiety, which is insufficiently persistent or intense to merit a diagnosis of depressive or anxiety disorder. There may also be anger, aggressive behaviour and associated excessive alcohol use. Symptoms develop within a month of the onset of the stress, and their duration and severity reflect the course of the underlying stressor.

Grief reactions following bereavement are a particular type of adjustment disorder. They manifest as a brief period of emotional numbing, followed by a period of distress lasting several weeks, during which sorrow, tearfulness, sleep disturbance, loss of interest and a sense of futility are common. Perceptual distortions may occur, including misinterpreting sounds as the dead person’s voice. ‘Pathological grief’ describes a grief reaction that is abnormally intense or persistent.

Management and prognosis

Ongoing contact with and support from a doctor or other who can listen, reassure, explain and advise are often all that is needed. Most patients do not require psychotropic medication, although benzodiazepines reduce arousal in acute stress reactions and can aid sleep in adjustment disorders. Psychotherapy may be helpful for patients with abnormal grief reactions. These conditions usually resolve with time but can develop into depressive or anxiety disorders and require treating as such.

Post-traumatic stress disorder

Post-traumatic stress disorder (PTSD) is a protracted response to a stressful event of an exceptionally threatening or catastrophic nature. Examples of such events include natural disasters, terrorist activity, serious accidents and witnessing violent deaths. PTSD may also sometimes occur after distressing medical treatments. There is usually a delay ranging from a few days to several months between the traumatic event and the onset of symptoms. Typical symptoms are recurrent intrusive memories (flashbacks) of the trauma, as well as sleep disturbance, especially nightmares (usually of the traumatic event) from which the patient awakes in a state of anxiety, symptoms of autonomic arousal, emotional blunting and avoidance of situations that evoke memories of the trauma. Anxiety and depression are often associated and excessive use of alcohol or drugs frequently complicates the clinical picture.

Management and prognosis

Immediate counselling for those who have survived a major trauma is only likely to benefit those who request it. The main aims are to provide support, direct advice and the opportunity for emotional catharsis. In established PTSD, structured psychological approaches (CBT, eye movement desensitisation and reprocessing (EMDR), and stress management) are effective. Anti-depressant drugs are moderately effective. The condition runs a fluctuating course, with most patients recovering within 2 years. In a small proportion, the symptoms become chronic.

Anxiety disorders

These are characterised by the emotion of anxiety, worrisome thoughts, avoidance behaviour and the somatic symptoms of autonomic arousal. Anxiety disorders are divided into three main subtypes: phobic, paroxysmal (panic) and generalised (Box 10.20). The nature and prominence of the somatic symptoms often lead the
patient to present initially to medical services. Anxiety may be stress-related and phobic anxiety may follow an unpleasant incident. Patients with anxiety often also have depression.

**Phobic anxiety disorder**
A phobia is an abnormal or excessive fear of an object or situation, which leads to avoidance of it (such as excessive fear of dying in an air crash leading to avoidance of flying). A generalised phobia of going out alone or being in crowded places is called agoraphobia. Phobic responses can develop to medical procedures such as vennupuncture.

**Panic disorder**
Panic disorder describes repeated attacks of severe anxiety, which are not restricted to any particular situation or circumstances. Somatic symptoms such as chest pain, palpitations and paraesthesiae in lips and fingers are common. The symptoms are in part due to involuntary over-breathing (hyperventilation). Patients with panic attacks often fear that they are suffering from a serious illness such as a heart attack or stroke, and seek emergency medical attention. Panic disorder is often associated with agoraphobia.

**Generalised anxiety disorder**
This is a chronic anxiety state associated with uncontrolled worry. The associated somatic symptoms of muscle tension and bowel disturbance often lead to a medical presentation.

**Management of anxiety disorders**

**Psychological treatment**
Explanation and reassurance are essential, especially when patients fear they have a serious medical condition. Specific treatment may be needed. Treatments include relaxation, graded exposure (desensitisation) to feared situations for phobic disorders, and CBT.

**Drug treatment**
Antidepressants are the drugs of choice. Benzodiazepines are useful in the short term but long-term use can lead to dependence. A β-blocker such as propranolol can help when somatic symptoms are prominent.

**Obsessive-compulsive disorder**
Obsessive-compulsive disorder (OCD) is characterised by obsessive thoughts, which are recurrent, unwanted and usually anxiety-provoking, but recognised as one’s own; and by compulsions, which are repeated acts performed to relieve the anxiety. An example is repeated hand-washing related to thoughts of contamination. The differential diagnosis is normal checking behaviour and delusional beliefs about thought possession. Unlike other anxiety disorders, which are more common in women, OCD is equally common in men and women.

**Management and prognosis**
OCD usually responds to some degree to antidepressant drugs (SSRIs; see Box 10.17) and to CBT, which helps patients expose themselves to the feared thought or situation without performing the anxiety-relieving compulsions. However, relapses are common and the condition often becomes chronic.

**Mood disorders**
Mood or affective disorders include:
- *unipolar depression*: one or more episodes of low mood and associated symptoms
- *bipolar disorder*: episodes of elevated mood interspersed with episodes of depression
- *dysthymia*: chronic low-grade depressed mood without sufficient other symptoms to count as ‘clinically significant’ or ‘major’ depression.

**Depression**
Major depressive disorder has a prevalence of 5% in the general population and approximately 10–20% in chronically ill medical outpatients. It is a major cause of disability and suicide. If comorbid with a medical condition, depression magnifies disability, diminishes adherence to medical treatment and rehabilitation, and may even shorten life expectancy.

**Aetiology**
There is a genetic predisposition to depression, especially when of early onset. The number and identity of the genes are largely unknown but the serotonin transporter gene is a candidate. Adversity and emotional deprivation early in life also predispose to depression. Depressive episodes are often, but not always, triggered by stressful life events (especially those that involve loss), including medical illnesses. Associated biological factors include hypofunction of monoamine neurotransmitter systems (5-HT and noradrenaline (norepinephrine)) and abnormal hypothalamo-pituitary-adrenal axis (HPA) regulation, which results in elevated cortisol levels that do not suppress with dexamethasone. Exclusion of Cushing’s syndrome is described on page 773.

**Diagnosis**
The symptoms are listed in Box 10.7 (p. 235). Depression may be mild, moderate or severe. It may also be episodic, recurrent or chronic. It can be both a complication of a medical condition and a cause of MUS (see below), so physical examination is essential; an associated medical condition should always be considered (Box 10.21).

**Management and prognosis**
There is evidence that both drug and psychological treatments work in depression. In practice, the choice is
determined by patient preference and local availability. Severe depression complicated by psychosis, dehydration or suicide risk may require ECT.

**Drug treatment**

Antidepressant drugs are effective in patients whose depression is secondary to medical illness, as well as those in whom it is the primary problem (Box 10.22). These agents are all effective in moderate and severe depression. Commonly used antidepressants are shown in Box 10.23.

- **Tricyclic antidepressants (TCAs).** These agents inhibit the re-uptake of the amines noradrenaline (norepinephrine) and 5-HT at synaptic clefts. The therapeutic effect is noticeable within a week or two. Side-effects, such as sedation, anticholinergic effects, postural hypotension, lowering of the seizure threshold and cardiotoxicity, can be troublesome during this period. TCAs may be dangerous in overdose and in people who have coexisting heart disease, glaucoma and prostatism.

  - **Selective serotonin re-uptake inhibitors (SSRIs).** These are less cardiotoxic and less sedative than TCAs, and have fewer anticholinergic effects. They are safer in overdose, but can still cause headache, nausea, anorexia and sexual dysfunction. They can also interact with other drugs increasing serotonin to produce ‘serotonin syndrome’. This is a rare syndrome of neuromuscular hyperactivity, autonomic hyperactivity and agitation, and potentially seizures, hyperthermia, confusion and even death.

  - **Neuer antidepressants.** A variety is available, including venlafaxine, mirtazapine and duloxetine. They have slightly different modes of action and adverse effects but are generally no more effective than the agents listed above.

- **Monoamine oxidase inhibitors (MAOIs).** These drugs increase the availability of neurotransmitters at synaptic clefts by inhibiting metabolism of noradrenaline (norepinephrine) and 5-HT. They are now rarely prescribed in the UK, since they can cause potentially dangerous interactions with drugs such as amphetamines, and foods rich in tyramine such as cheese and red wine. This is due to accumulation of amines in the systemic circulation, causing a potentially fatal hypertensive crisis.

These different classes of antidepressant have similar efficacy and about three-quarters of patients respond to treatment. Successful treatment requires the patient to take an appropriate dose of an effective drug for an adequate period. For patients who do not respond, a proportion will do so if changed to another antidepressant. The patient’s progress must be monitored and, after recovery, treatment should be continued for at least 6–12 months to reduce the high risk of relapse. The dose should then be tapered off over several weeks to avoid discontinuation symptoms. SIGN (www.sign.ac.uk) and NICE (www.nice.org.uk) have published treatment guidelines.

**Psychological treatments**

Both CBT and interpersonal therapy are as effective as antidepressants for mild to moderate depression. Antidepressant drugs are, however, preferred for severe depression. Drug and psychological treatments can be used in combination.

Over 50% of people who have had one depressive episode and over 90% of people who have had three or more episodes will have another. The risk of suicide in an individual who has had a depressive disorder is ten times greater than in the general population.

**Bipolar disorder**

Bipolar disorder is an episodic disturbance with interspersed periods of depressed and elevated mood; the latter is known as hypomania when mild or short-lived, or mania when severe or chronic. The lifetime risk of developing bipolar disorder is approximately 1–2%. Onset is usually in the twenties, and men and women

### Box 10.21 Pointers to an organic cause for psychiatric disorder

- Late age of onset of psychiatric illness
- No previous history of psychiatric illness
- No family history of psychiatric illness
- No apparent psychological precipitant

### Box 10.22 Antidepressants in the medically ill

‘Antidepressants are efficacious and safe in the treatment of depression occurring in the context of chronic physical health problems. The selective serotonin re-uptake inhibitors are probably the antidepressants of first choice, given their demonstrable effect on quality of life and their apparent safety in cardiovascular disease.’


### Box 10.23 Antidepressant drugs

<table>
<thead>
<tr>
<th>Group</th>
<th>Drug</th>
<th>Usual dose*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tricyclics</td>
<td>Amitriptyline</td>
<td>75–150 mg daily</td>
</tr>
<tr>
<td></td>
<td>Imipramine</td>
<td>75–150 mg daily</td>
</tr>
<tr>
<td></td>
<td>Dosulepin</td>
<td>75–150 mg daily</td>
</tr>
<tr>
<td></td>
<td>Clomipramine</td>
<td>75–150 mg daily</td>
</tr>
<tr>
<td>Selective serotonin re-uptake inhibitors (SSRIs)</td>
<td>Citalopram</td>
<td>20–40 mg daily</td>
</tr>
<tr>
<td></td>
<td>Escitalopram</td>
<td>10–20 mg daily</td>
</tr>
<tr>
<td></td>
<td>Fluoxetine</td>
<td>20–60 mg daily</td>
</tr>
<tr>
<td></td>
<td>Fluvoxamine</td>
<td>100–300 mg daily</td>
</tr>
<tr>
<td></td>
<td>Sertraline</td>
<td>50–100 mg daily</td>
</tr>
<tr>
<td></td>
<td>Paroxetine</td>
<td>20–50 mg daily</td>
</tr>
<tr>
<td>Monoamine oxidase inhibitors (MAOIs)</td>
<td>Phenelzine</td>
<td>45–90 mg daily</td>
</tr>
<tr>
<td></td>
<td>Tranylcypromine</td>
<td>20–40 mg daily</td>
</tr>
<tr>
<td></td>
<td>Moclobemide</td>
<td>300–600 mg daily</td>
</tr>
<tr>
<td>Noradrenergic re-uptake inhibitors and SSRIs</td>
<td>Venlafaxine</td>
<td>75–375 mg daily</td>
</tr>
<tr>
<td>Noradrenergic and specific serotonergic inhibitor</td>
<td>Mirtazapine</td>
<td>15–45 mg daily</td>
</tr>
</tbody>
</table>

*Higher doses may be required: see guidelines.
are equally affected. In DSM-IV, bipolar disorder has been divided into two types:

- **Bipolar I disorder** has a clinical course characterised by one or more manic episodes or mixed episodes. Often individuals have also had one or more major depressive episodes.
- **Bipolar II disorder** features depressive episodes that are more frequent and more intense than manic episodes, but there is a history of at least one hypomanic episode.

**Aetiology**

Bipolar disorder is strongly heritable (approximately 70%). Relatives of patients have an increased incidence of both bipolar and unipolar affective disorder. Life events, such as physical illness, sleep deprivation and medication, may play a role in triggering episodes.

**Diagnosis**

The diagnosis is based on clear evidence of episodes of depression and mania. Isolated episodes of hypomania or mania do occur but they are usually preceded or followed by an episode of depression. Psychosis may occur in both the depressive and the manic phases, with delusions and hallucinations that are usually in keeping with the mood disturbance. This is described as an affective psychosis. Patients who present with symptoms of both bipolar disorder and schizophrenia may be given a diagnosis of schizoaffective disorder. A clinical picture of recurrent depression with one or more episodes of hypomania may be referred to as type 2 bipolar disorder.

**Management and prognosis**

Depression should be treated as described above. However, if antidepressants are prescribed, they should be combined with a mood-stabilising drug (see below) to avoid ‘switching’ the patients into (hypo)mania. Manic episodes and psychotic symptoms usually respond well to antipsychotic drugs (see Box 10.30, p. 249).

Prophylaxis to prevent recurrent episodes of depression and mania with mood-stabilising agents is important. The main drugs used are lithium and sodium valproate. Olanzapine, quetiapine and risperidone are increasingly used. Caution must be exercised when stopping these drugs, as a relapse may follow.

- **Lithium carbonate** is the drug of choice. It is also used for acute mania, and in combination with a tricyclic as an adjuvant treatment for resistant depression. It has a narrow therapeutic range, so regular blood monitoring is required to maintain a serum level of 0.5–1.0 mmol/L. Toxic effects include nausea, vomiting, tremor and convulsions. With long-term treatment, weight gain, hypothyroidism, increased calcium and parathormone, nephrogenic diabetes insipidus (p. 794) and renal failure can occur. Thyroid and renal function should be checked before treatment is started and regularly thereafter. Lithium may be teratogenic, and should not be prescribed during the first trimester of pregnancy.
- **Sodium valproate (an anticonvulsant) and olanzapine (an antipsychotic)** are both used as prophylaxis in bipolar disorder, usually as a second-line alternative to lithium. Valproate conveys a high risk of birth defects and should also not be used in women of child-bearing age. Olanzapine can cause significant weight gain. (For a list of the adverse effects of antipsychotic drugs, see Box 10.31.)

The relapse rate of bipolar disorder is high, although patients may be perfectly well between episodes. After one episode the annual average risk of relapse is about 10–15%, which doubles after more than three episodes. There is a substantially increased lifetime risk of suicide of 5–10%.

**Somatoform disorders**

The essential feature of these disorders is that the somatic symptoms are not explained by a medical condition (medically unexplained symptoms), nor better diagnosed as part of a depressive or anxiety disorder. Several syndromes are described within this category; there is considerable overlap between them in both aetiology and clinical presentation.

**Aetiology**

The cause of somatoform disorders is incompletely understood but contributory factors include depression and anxiety, the erroneous interpretation of somatic symptoms as evidence of disease, excessive concern with physical illness and a tendency to seek medical care. A family history or previous history of a particular condition may have shaped the patient’s beliefs about illness. Doctors may exacerbate the problem, either by dismissing the complaints as non-existent or by over-emphasising the possibility of disease.

**Somatisation disorder**

Somatisation disorder (Briquet’s syndrome) is characterised by the occurrence of chronic multiple somatic symptoms for which there is no physical cause. Symptoms start in early adult life and may be referred to any part of the body. The disorder is much more common in women. Common complaints include pain, vomiting, nausea, headache, dizziness, menstrual irregularities and sexual dysfunction. Patients may undergo a multitude of negative investigations and unhelpful operations, particularly hysterectomy and cholecystectomy. There is no proven treatment but minimisation of iatrogenic harm from multiple investigations and attempts at medical treatment is important.

**Hypochondriacal disorder**

Patients with this condition, also known as health anxiety, have a strong fear or belief that they have a serious, often fatal, disease and that fear persists despite appropriate medical reassurance. They are typically highly anxious and seek many medical opinions and investigations in futile but repeated attempts to relieve their fears. Treatment with CBT may be helpful. The condition may become chronic.

In a small proportion of cases, the conviction that disease is present reaches delusional intensity. The best-known example is that of parasitic infestation (‘delusional parasitosis’), which leads patients to consult dermatologists. Antipsychotic medication may be effective in such cases.
Body dysmorphic disorder
This describes a preoccupation with bodily shape or appearance, with the belief that one is disfigured in some way (previously known as dysmorphophobia). People with this condition may make inappropriate requests for cosmetic surgery. CBT or antidepressants may be helpful. The belief in disfigurement may sometimes be delusional, in which case antipsychotic drugs may help.

Somatoform autonomic dysfunction
This describes somatic symptoms referable to bodily organs that are largely under the control of the autonomic nervous system. The most common examples involve the cardiovascular system (‘cardiac neurosis’), respiratory system (psychogenic hyperventilation) and gut (psychogenic vomiting and irritable bowel syndrome). Antidepressant drugs and CBT may be helpful.

Somatoform pain disorder
This describes severe, persistent pain that cannot be adequately explained by a medical condition. Antidepressant drugs (especially tricyclics and dual-action drugs such as duloxetine and mirtazapine) are helpful, as are some of the anticonvulsant drugs, particularly carbamazepine, gabapentin and pregabalin. CBT and multidisciplinary pain management teams are also useful.

Chronic fatigue syndrome
Chronic fatigue syndrome (CFS) is also referred to as neurasthenia. It is characterised by excessive fatigue after minimal physical or mental exertion, poor concentration, dizziness, muscular aches and sleep disturbance. This pattern of symptoms may follow a viral infection such as infectious mononucleosis, influenza or hepatitis. Symptoms overlap with those of depression and anxiety. There is good evidence that many patients improve with carefully graded exercise and with CBT, as long as the benefits of such treatment are carefully explained.

Dissociative (conversion) disorder
Dissociative disorder refers to a loss or distortion of neurological functioning that is not fully explained by organic disease. Psychological functions commonly affected include conscious awareness and memory. Physical functions affected (conversion) include changes in sensory or motor function that may mimic lesions in the motor or sensory nervous system (Box 10.24). The aetiology of dissociation is unknown. There is an association with adverse childhood experiences, including physical and sexual abuse. Organic disease may both facilitate dissociative mechanisms and provide a model for symptoms; thus, for example, non-epileptic seizures often occur in those with epilepsy. CBT may be of benefit. Coexisting depression should be treated with CBT or antidepressant drugs.

General management for medically unexplained symptoms
The management of the various syndromes of medically unexplained complaints described above is based on general principles (Box 10.25).

Reassurance
Patients should be asked what they are most worried about. Clearly, it may be unwise to state categorically that the patient does not have any disease, as that is difficult to establish with certainty. However, it can be emphasised that the probability of having a disease is low. If patients repeatedly ask for reassurance about the same health concern despite reassurance, they may have hypochondriasis.

Explanation
Patients need a positive explanation for their symptoms. It is unhelpful to say that symptoms are psychological or ‘all in the mind’. Rather, a term such as ‘functional’ (meaning that the symptoms represent a reversible disturbance of bodily function) may be more acceptable. When possible, it is useful to describe a plausible physiological mechanism that is linked to psychological factors such as stress and implies that the symptoms are reversible. For example, in irritable bowel syndrome, psychological stress results in increased activation of the autonomic nervous system, which leads to constriction of smooth muscle in the gut wall, which in turn causes pain and bowel disturbance.

Advice
This should focus on how to overcome factors perpetuating the symptoms; for example, by resolving stressful social problems or by practising relaxation. The doctor can offer to review progress, to prescribe (for example) an antidepressant drug and, if appropriate, to refer for physiotherapy or psychological treatments such as CBT. The attitudes of relatives may need to be addressed if they have adopted an over-protective role, unwittingly reinforcing the patient’s disability.

Drug treatment
Antidepressant drugs are often helpful, even if the patient is not depressed (Box 10.26).
Psychiatric disorders

EBM 10.26 Antidepressants for medically unexplained somatic symptoms

‘There is evidence for the efficacy of antidepressant drugs for patients with medically unexplained symptoms.’


EBM 10.27 CBT for medically unexplained somatic symptoms

‘CBT is consistently effective (11 of 13 trials) across a spectrum of somatoform disorders. Also, a psychiatric consultation letter to the primary care physician about strategies for managing the somatising patient seems to improve physical functioning and reduce costs.’


Psychological treatment

There is evidence for the effectiveness of CBT (Box 10.27). Other psychological treatments such as IPT may also have a role.

Rehabilitation

Where there is chronic disability, particularly in dissociative (conversion) disorder, conventional physical rehabilitation may be the best approach.

Shared care with the GP

Ongoing planned care is required for patients with chronic intractable symptoms, especially those of somatisation disorder. Review by the same specialist, interspersed with visits to the same GP, is probably the best way to avoid unnecessary multiple re-referral for investigation, to ensure that treatable aspects of the patient’s problems, such as depression, are actively managed, and to prevent the GP from becoming demoralised.

Factitious disorder and malingering

It is important to distinguish somatoform disorders from factitious disorder and malingering.

Factitious disorder

This describes the repeated and deliberate production of the signs or symptoms of disease to obtain medical care. It is uncommon. An example is the dipping of thermometers into hot drinks to fake a fever. The disorder feigned is usually medical but can be a psychiatric illness, with false reports of hallucinations or symptoms of depression.

Münchausen’s syndrome

This refers to a severe chronic form of factitious disorder. Patients are usually older and male, with a solitary, peripatetic lifestyle in which they travel widely, sometimes visiting several hospitals in one day. Although the condition is rare, such patients are memorable because they present so dramatically. The history can be convincing enough to persuade doctors to undertake investigations or initiate treatment, including exploratory surgery. It may be possible to trace the patient’s history and show that he has presented similarly elsewhere, often changing name several times. Some emergency departments hold lists of such patients.

Management is by gentle but firm confrontation with clear evidence of the fabrication of illness, together with an offer of psychological support. Treatment is usually declined but recognition of the condition may help to avoid further iatrogenic harm.

Malingering

Malingering is a description of behaviour, not a psychiatric diagnosis. It refers to the deliberate and conscious simulation of signs of disease and disability. Patients have motives that are clear to them but which they conceal from doctors. Examples include the avoidance of burdensome responsibilities (such as work or court appearances) or the pursuit of financial gain (fraudulent claims for benefits or compensation). Malingering can be hard to detect at clinical assessment, but is suggested by evasion or inconsistency in the history.

Schizophrenia

Schizophrenia is a psychosis characterised by delusions, hallucinations and lack of insight. Acute schizophrenia may present with disturbed behaviour, marked delusions, hallucinations and disordered thinking, or with insidious social withdrawal and other so-called negative symptoms and less obvious delusions and hallucinations. The prevalence is similar worldwide at about 1% and the disorder is more common in men. The children of one affected parent have approximately a 10% risk of developing the illness, but this rises to 50% if an identical twin is affected. The usual age of onset is the mid-twenties.

Aetiology

There is a strong genetic contribution, probably involving many susceptibility genes, each of small effect. The best candidates, such as disrupted in schizophrenia-1 (DISC1) and neuregulin-1 (NRG1), have supportive linkage, association, animal model and basic neurobiological evidence. Environmental risk factors include obstetric complications and urban birth. Brain imaging techniques have identified subtle structural abnormalities, including an enlargement of the lateral ventricles and an overall decrease in brain size (by about 3% on average), with relatively greater reduction in temporal lobe volume (5–10%). Episodes of acute schizophrenia may be precipitated by social stress and also by cannabis, which increase dopamine turnover and sensitivity. Consequently, schizophrenia is now viewed as a neurodevelopmental disorder, caused by abnormalities of brain development associated with genetic predisposition and early environmental influences, but precipitated by later triggers.

Diagnosis

Schizophrenia usually presents with an acute episode and progresses to a chronic state. Acute schizophrenia should be suspected in any individual with bizarre behaviour accompanied by delusions and hallucinations that are not due to organic brain disease or substance misuse. The diagnosis is made on clinical grounds, with
investigations used principally to rule out organic brain disease. The characteristic clinical features are listed in Box 10.28. Hallucinations are typically auditory, although they can occur in any sensory modality. They commonly involve voices from outside the head that talk to or about the person. Sometimes the voices repeat the person’s thoughts. Patients may also describe ‘passivity of thought’, experienced as disturbances in the normal privacy of thinking – for example, the delusional belief that their thoughts are being ‘withdrawn’ from them, perhaps ‘broadcast’ to others, and/or alien thoughts being ‘inserted’ into their mind. Other characteristic symptoms are delusions of control: believing that one’s emotions, impulses or acts are controlled by others. Another phenomenon is delusional perception, a delusion that arises suddenly alongside a normal perception (e.g. ‘I saw the moon and I immediately knew he was evil’). Many other, less specific symptoms may occur, including thought disorder, as manifest by incomprehensible speech and abnormalities of movement, such as those in which the patient can become immobile or adopt awkward postures for prolonged periods (catatonia).

The main differential diagnosis of schizophrenia (Box 10.29) is:

- **Other functional psychoses**, particularly psychotic depression and mania, in which delusions and hallucinations are congruent with a marked mood disturbance (negative in depression and grandiose in mania). If features of schizophrenia and affective disorder coexist in equal measure, a diagnosis of schizoaffective disorder is made. Schizophrenia must also be differentiated from specific delusional disorders that are not associated with the other typical features of schizophrenia.
- **Organic psychoses**, including delirium, in which there is impairment of consciousness and loss of orientation (not found in schizophrenia), typically with visual hallucinations, and drug misuse, the latter particularly in young people. Schizophrenia must also be differentiated from other organic psychoses such as temporal lobe epilepsy, in which olfactory and gustatory hallucinations may occur.

Many of those who experience acute schizophrenia go on to develop a chronic state in which the acute, so-called positive symptoms resolve, or at least do not dominate the clinical picture, leaving so-called negative symptoms that include blunt affect, apathy, social isolation, poverty of speech and poor self-care. Patients with chronic schizophrenia may also manifest positive symptoms, particularly when under stress, and it can be difficult for those who do not know the patient to judge whether or not these are signs of an acute relapse.

**Management**

First-episode schizophrenia usually requires admission to hospital because patients lack the insight that they are ill and are unwilling to accept treatment. In some cases, they may be at risk of harming themselves or others. Subsequent acute relapses and chronic schizophrenia are now usually managed in the community.

**Drug treatment**

Antipsychotic agents are effective against the positive symptoms of schizophrenia in the majority of cases. They take 2–4 weeks to be maximally effective but have some beneficial effects shortly after administration. Treatment is then ideally continued to prevent relapse. In a patient with a first episode of schizophrenia, this will usually be for 1–2 years, but in patients with multiple episodes, treatment may be required for many years. The benefits of prolonged treatment must be weighed against the adverse effects, which include extrapyramidal side-effects (EPS) like acute dystonic reactions (which may require treatment with parenteral anticholinergics), akathisia, Parkinsonism and tardive dyskinesia (abnormal movements, commonly of the face, over which the patient has no voluntary control). For long-term use, antipsychotic agents are often given in slow-release (depot) injected form to improve patient adherence.
A number of antipsychotic agents are available (Box 10.30). These may be divided into conventional (typical, first-generation) drugs such as chlorpromazine and haloperidol, and newer or atypical (also so-called novel or second-generation) drugs such as clozapine. All are believed to work by blocking D2 dopamine receptors in the brain. Patients who have not responded to conventional drugs may respond to newer agents, which are also less likely to produce unwanted EPS but do tend to cause greater weight gain and metabolic disturbances such as dyslipidaemia. Clozapine can also cause an agranulocytosis and consequently requires regular monitoring of the white blood cell count, initially on a weekly basis. Details of the side-effects of antipsychotic drugs are listed in Box 10.31.

Serious adverse effects of antipsychotic drugs include:

- **Neuroleptic malignant syndrome**, which is a rare but serious condition. It is characterised by fever, tremor and rigidity, autonomic instability and confusion. Characteristic laboratory findings are an elevated creatinine phosphokinase and leucocytosis. Antipsychotic medication must be stopped immediately and supportive therapy provided, often in an intensive care unit. Treatment includes ensuring hydration and reducing hyperthermia. Dantrolene sodium and bromocriptine may be helpful. Mortality is 20% untreated and 5% with treatment.

- **Prolongation of the QTc interval**, which may be associated with ventricular tachycardia, torsades de pointes and sudden death. Treatment is by stopping the drug, monitoring the electrocardiogram (ECG) and treating serious arrhythmias (p. 562).

**Psychological treatment**

Psychological treatment, including general support for the patient and his or her family, is now seen as an essential component of management. CBT may help patients to cope with symptoms. There is evidence that personal and/or family education, when given as part of an integrated treatment package, reduces the rate of relapse.

**Social treatment**

After an acute episode of schizophrenia has been controlled by drug therapy, social rehabilitation may be required. Recurrent illness is likely to cause disruption to patients’ relationships and their ability to manage their accommodation and occupation; consequently, they may need help to obtain housing and employment. A graded return to employment and sometimes a period of supported accommodation are required.

Patients with chronic schizophrenia have particular difficulties and may need long-term, supervised accommodation. This now tends to be in sheltered or hostel accommodation in the community. Patients may also benefit from sheltered employment if they are unable to participate effectively in the labour market. Ongoing contact with a health worker allows monitoring for signs of relapse, sometimes as part of a multidisciplinary team working to agreed plans (the ‘care programme approach’). Partly because of a tendency to inactivity, smoking and a poor diet, patients with chronic schizophrenia are at increased risk of cardiovascular disease, diabetes and stroke, and require proactive medical as well as psychiatric care.

**Prognosis**

About one-quarter of those who develop an acute schizophrenic episode have a good outcome. One-third develop chronic, incapacitating schizophrenia, and the remainder largely recover after each episode but suffer relapses. Most will not work or live independently. Prophylactic treatment with antipsychotic drugs reduces the rate of relapse in the first 2 years after an episode of schizophrenia from 50% to 10%. Schizophrenia is associated with suicide, with up to 1 in 10 patients taking their own lives.

### 10.30 Antipsychotic drugs

<table>
<thead>
<tr>
<th>Group</th>
<th>Drug</th>
<th>Dose range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenothiazines</td>
<td>Chlorpromazine</td>
<td>100–1000 mg daily</td>
</tr>
<tr>
<td></td>
<td>Haloperidol</td>
<td>5–30 mg daily</td>
</tr>
<tr>
<td>Thioxanthenes</td>
<td>Haloperidol</td>
<td>20–200 mg fortnightly (depot injection)</td>
</tr>
<tr>
<td>Butyrophenones</td>
<td>Pimozide</td>
<td>4–30 mg daily</td>
</tr>
<tr>
<td>Substituted benzamides</td>
<td>Sulpiride</td>
<td>200–1800 mg daily</td>
</tr>
<tr>
<td>Dibenzodiazepines*</td>
<td>Clozapine</td>
<td>200–900 mg daily</td>
</tr>
<tr>
<td>Benzisoxazole*</td>
<td>Risperidone</td>
<td>2–16 mg daily</td>
</tr>
<tr>
<td>Thienobenzodiazepines*</td>
<td>Olanzapine</td>
<td>5–20 mg daily</td>
</tr>
<tr>
<td>Dibenzothiazepines</td>
<td>Quetiapine</td>
<td>25–800 mg daily</td>
</tr>
</tbody>
</table>

*Second-generation antipsychotics.

### 10.31 Side-effects of antipsychotic drugs

**Weight gain due to increased appetite**

- Acute dystonia
- Akathisia (motor restlessness)
- Parkinsonism

**Effects due to dopamine blockade**

- Tardive dyskinesia
- Gynaecomastia
- Galactorrhoea

**Effects due to cholinergic blockade**

- Dry mouth
- Blurred vision
- Impotence

**Hypersensitivity reactions**

- Blood dyscrasias (neutropenia with clozapine)
- Cholestatic jaundice
- Photosensitive dermatitis

**Ocular complications (long-term use)**

- Corneal and lens opacities

*Less severe with clozapine, quetiapine and olanzapine, possibly because of strong 5-HT-blocking effect and relatively weak dopamine blockade.
Delirium, dementia and other organic disorders

Delirium, dementia and other organic disorders could be considered to be medical conditions rather than psychiatric disorders, as they are a result of reduced brain function; they are, however, included in psychiatric classifications and are sometimes misdiagnosed because they often manifest with disturbed behaviour.

Aetiology

Dementia may be divided into ‘cortical’ and ‘subcortical’ types, depending on the clinical features. Many of the primary degenerative diseases that cause dementia have characteristic features that may allow a specific diagnosis during life. Creutzfeldt–Jakob disease, for example, is usually relatively rapidly progressive (over months), is associated with myoclonus, and demonstrates characteristic abnormalities on electroencephalogram (EEG). The more slowly progressive dementias are more difficult to distinguish during life, but fronto-temporal dementia typically presents with focal (temporal or frontal lobe) dysfunction, and Lewy body dementia may present with visual hallucinations. The course may also help to distinguish types of dementia, as it may be gradual (as in Alzheimer’s disease) or step-wise (as in vascular dementia).

Clinical features

The usual presentation is with a disturbance of personality or memory dysfunction. A careful history is essential.

Delirium

Delirium is common in acute medical settings, especially in the elderly and patients in high-dependency and intensive care units. Aetiology, assessment and management are described in Chapters 7 and 26.

Dementia

Dementia is a clinical syndrome characterised by a loss of previously acquired intellectual function in the absence of impairment of arousal, and affects 5% of those over 65 and 20% of those over 85. It is defined as a global impairment of cognitive function, and is typically progressive and non-reversible. Although memory is most affected in the early stages, deficits in visuo-spatial function, language ability, concentration and attention gradually become apparent. There are many causes (Box 10.32) but Alzheimer’s disease and diffuse vascular disease are the most common. Rarer causes of dementia should be actively sought in younger patients and those with short histories.

<table>
<thead>
<tr>
<th>Type</th>
<th>Common</th>
<th>Unusual</th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular</td>
<td>Diffuse small-vessel disease</td>
<td>Amyloid angiopathy</td>
<td>Cerebral vasculitis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple emboli</td>
<td></td>
</tr>
<tr>
<td>Degenerative/inherited</td>
<td>Alzheimer’s disease</td>
<td>Fronto-temporal dementia (including Pick’s disease)</td>
<td>Mitochondrial encephalopathies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leucodystrophies</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huntington’s disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wilson’s disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dystrophia myotonica</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cortical Lewy body disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progressive supranuclear palsy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Others (e.g. cortico-basal degeneration)</td>
<td></td>
</tr>
<tr>
<td>Neoplastic (p. 1193)</td>
<td>Secondary deposits</td>
<td>Primary cerebral tumour</td>
<td>Paraneoplastic syndrome (limbic encephalitis)</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>–</td>
<td>Multiple sclerosis</td>
<td>Sarcoïdosis</td>
</tr>
<tr>
<td>Traumatic</td>
<td>Chronic subdural haematoma</td>
<td>Punch-drunk syndrome</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Post-head injury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrocephalus (p. 1216)</td>
<td></td>
<td>Communicating/non-communicating ‘normal pressure’ hydrocephalus</td>
<td>–</td>
</tr>
<tr>
<td>Toxic/nutritional</td>
<td>Alcohol</td>
<td>Thiamin deficiency</td>
<td>Anoxia/carbon monoxide poisoning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B12 deficiency</td>
<td>Heavy metal poisoning</td>
</tr>
<tr>
<td>Infective</td>
<td>–</td>
<td>Syphilis</td>
<td>Post-encephalitic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HIV</td>
<td>Whipple’s disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Subacute sclerosing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>panencephalitis</td>
</tr>
<tr>
<td>Prion diseases (p. 1211)</td>
<td>–</td>
<td>Sporadic Creutzfeldt–Jakob disease (CJD)</td>
<td>Variant CJD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kuru</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gerstmann–Sträussler–Scheinker disease</td>
</tr>
</tbody>
</table>
Alzheimer’s disease

Alzheimer’s disease is the most common cause of dementia, but is rare under the age of 45 years.

Aetiology

Genetic factors play an important role and about 15% of cases are familial. Familial cases fall into two main groups: early-onset disease with autosomal dominant inheritance and a later-onset group whose inheritance is polygenic. Mutations in several genes have been described. The inheritance of one of the alleles of apo lipoprotein ε (apo ε4) is associated with an increased risk of developing the disease (2–4 times higher in heterozygotes and 6–8 times in homozygotes). Its presence is, however, neither necessary nor sufficient for the development of the disease, so screening for its presence is not clinically useful. The brain in Alzheimer’s disease is macroscopically atrophic, particularly the cerebral cortex and hippocampus. Histologically, the disease is characterised by the presence of senile plaques and neurofibrillary tangles in the cerebral cortex. Histochemical staining demonstrates significant quantities of amyloid in the plaques (Fig. 10.4), which typically stain positive for the protein ubiquitin, involved in targeting unwanted or damaged proteins for degradation. This has led to the suggestion that the disease may be due to defects in the ability of neuronal cells to degrade unwanted proteins. Many different neurotransmitter abnormalities have also been described. In particular, there is impairment of cholinergic transmission, although

Investigations

The aim is to seek treatable causes and to estimate prognosis. This is done using a standard set of investigations (Box 10.33). Imaging of the brain can exclude potentially treatable structural lesions, such as hydrocephalus, cerebral tumour or chronic subdural haematoma, though the only abnormality usually seen is that of generalised atrophy. If the initial tests are negative, more invasive investigations, such as lumbar puncture or, rarely, brain biopsy, may be indicated.

Management

This is directed at addressing treatable causes, and providing support for patient and carers if no specific treatment exists. If the diagnosis is Alzheimer-type dementia, anticholinesterase inhibitors and memantine may arrest progression for a time. Treating vascular risk factors may slow deterioration in vascular dementia. Psychotropic drugs may help where there is associated disturbance of sleep, perception or mood, but should be used with care because of an increased mortality in patients who have been treated long-term with these agents. Sedation is not a substitute for good care for patients and carers or, in the later stages, attentive residential nursing care. In the UK, incapacity and mental health legislation may be required to manage patients’ financial and domestic affairs, as well as to determine their safe placement.

### Box 10.33 Initial investigation of dementia

<table>
<thead>
<tr>
<th>In most patients</th>
<th>In selected patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaging of head (computed tomography (CT) and/or magnetic resonance imaging (MRI))</td>
<td>Lumbar puncture</td>
</tr>
<tr>
<td>Blood tests</td>
<td>HIV serology</td>
</tr>
<tr>
<td>Full blood count, erythrocyte sedimentation rate</td>
<td>Brain biopsy</td>
</tr>
<tr>
<td>Urea and electrolytes, glucose</td>
<td></td>
</tr>
<tr>
<td>Calcium, liver function tests</td>
<td></td>
</tr>
<tr>
<td>Thyroid function tests</td>
<td></td>
</tr>
<tr>
<td>Vitamin B₁₂</td>
<td></td>
</tr>
<tr>
<td>Venereal Diseases Research Laboratory (VDRL) test</td>
<td></td>
</tr>
<tr>
<td>ANA, anti-dsDNA</td>
<td></td>
</tr>
<tr>
<td>Chest X-ray</td>
<td></td>
</tr>
<tr>
<td>EEG</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 10.4** Alzheimer’s disease. Section of neocortex stained with polyclonal antibody against βA4 peptide showing amyloid deposits in plaques in brain substance (arrow A) and in blood vessel walls (arrow B).
abnormalities of noradrenaline, 5-HT, glutamate and substance P have also been described.

**Clinical features**

The key clinical feature is impairment of the ability to remember new information. Hence, patients present with gradual impairment of memory, usually in association with disorders of other cortical functions. Short-term and long-term memory are both affected, but defects in the former are usually more obvious. Later in the course of the disease, typical features include apraxia, visuo-spatial impairment and aphasia. In the early stages of the disease, patients may notice these problems, but as the disease progresses it is common for patients to deny that there is anything wrong (anosognosia). In this situation, patients are often brought to medical attention by their carers. Depression is commonly present. Occasionally, patients become aggressive, and the clinical features can be made acutely worse by intercurrent physical disease.

**Investigations and management**

Investigation is aimed at excluding treatable causes of dementia (see Box 10.32), as histological confirmation of the diagnosis usually occurs only after death. There is no known treatment, though anticholinesterases such as donepezil, rivastigmine and galantamine, and the NMDA receptor antagonist, memantine, have been shown to be of some benefit. Management consists largely of providing a familiar environment for the patient and support for the carers. Many patients are depressed, and treatment with antidepressant medication may be helpful.

**Fronto-temporal dementia**

This term encompasses a number of different syndromes, including Pick’s diseases and primary progressive aphasia. Patients may present with personality change due to frontal lobe involvement or with language disturbance due to temporal lobe involvement. These diseases are much rarer than Alzheimer’s disease. Histological examination of the brain reveals argyrophilic cytoplasmic inclusion bodies of tau (τ) protein rather than the ubiquitin as in Alzheimer’s disease (Fig. 10.5). Memory is relatively preserved in the early stages. There is no specific treatment.

**Lewy body dementia**

This is a neurodegenerative disorder clinically characterised by dementia and signs of Parkinson’s disease. The cognitive state often fluctuates and there is a high incidence of visual hallucinations. Affected individuals are particularly sensitive to the side-effects of anti-Parkinsonian medication and also to antipsychotic drugs. The condition is associated with accumulation of abnormal protein aggregates in neurons that contain the protein α-synuclein in association with other proteins including ubiquitin (see Fig. 26.29, p. 1195). The condition is often inherited and mutations in the α-synuclein and β-synuclein genes have been identified in affected patients. There is no specific treatment but anticholinesterase agents may well be helpful.

**Alcohol misuse and dependence**

Alcohol consumption associated with social, psychological and physical problems constitutes misuse. The criteria for alcohol dependence, a more restricted term, are shown in Box 10.34. Approximately one-quarter of male patients in general hospital medical wards in the UK have a current or previous alcohol problem.

**Aetiology**

Availability of alcohol and social patterns of use appear to be the most important factors. Genetic factors predispose to dependence. The majority of alcoholics do not have an associated psychiatric disorder, but a few drink heavily in an attempt to relieve anxiety or depression.

### 10.34 Criteria for alcohol dependence

- Narrowing of the drinking repertoire
- Priority of drinking over other activities (salience)
- Tolerance of effects of alcohol
- Repeated withdrawal symptoms
- Relief of withdrawal symptoms by further drinking
- Subjective compulsion to drink
- Reinstatement of drinking behaviour after abstinence

---

**Fig. 10.5 Fronto-temporal dementia.**

A. Lateral view of formalin-fixed brain from a patient who died of Pick’s disease, showing gyral atrophy of frontal and parietal lobes and a more severe degree of atrophy affecting the anterior half of the temporal lobe (arrow).

B. High power (× 200) of hippocampal pyramidal layer, prepared with monoclonal anti-tau antibody. Many neuronal cell bodies contain sharply circumscribed, spherical cytoplasmic inclusion bodies (Pick bodies).
Diagnosis
Alcohol misuse may emerge during the patient’s history, although patients may minimise their intake. It may also present via its effects on one or more aspects of the patient’s life, listed below. Alcohol dependence commonly presents with withdrawal in those admitted to hospital, as they can no longer maintain their high alcohol intake in this setting.

Complications of chronic alcohol misuse
- **Social problems** include absenteeism from work, unemployment, marital tensions, child abuse, financial difficulties and problems with the law, such as violence and traffic offences.
- **Depression** is common. Alcohol has a direct depressant effect and heavy drinking creates numerous social problems. Attempted and completed suicide are associated with alcohol misuse.
- **Anxiety** is relieved by alcohol in the short term. People who are socially anxious may consequently use alcohol in this way and may develop dependence. Conversely, alcohol withdrawal increases anxiety.
- **Alcoholic hallucinosis** is a rare condition in which alcoholic individuals experience auditory hallucination in clear consciousness.
- **Alcohol withdrawal** is described in Box 10.35. Symptoms usually become maximal about 2–3 days after the last drink and can include seizures (‘rum fits’).
- **Delirium tremens** is a form of delirium associated with severe alcohol withdrawal. It has a significant mortality and morbidity (see Box 10.35).

Effects on the brain
The familiar features of drunkenness are ataxia, slurred speech, emotional incontinence and aggression. Very heavy drinkers may experience periods of amnesia for events that occurred during bouts of intoxication, termed ‘alcoholic blackouts’. Established alcoholism may lead to alcoholic dementia, a global cognitive impairment resembling Alzheimer’s disease, but which does not progress and may even improve if the patient becomes abstinent. Indirect effects on behaviour can result from head injury, hypoglycaemia and encephalopathy (p. 941).

A rare but important effect of chronic alcohol misuse is the Wernicke–Korsakoff syndrome. This organic brain disorder results from damage to the mamillary bodies, dorsomedial nuclei of the thalamus and adjacent areas of periventricular grey matter caused by a deficiency of thiamin (vitamin B1), which most commonly results from long-standing heavy drinking and an inadequate diet. It can also arise from malabsorption or even protracted vomiting. Without prompt treatment (see below), the acute presentation of Wernicke’s encephalopathy (nystagmus, ophthalmoplegia, ataxia and confusion) can progress to the irreversible deficits of Korsakoff’s syndrome (severe short-term memory deficits and confabulation, and also reduced red blood cell transketolase). In those who die in the acute stage, microscopic examination of the brain shows hyperaemia, petechial haemorrhages and astrocytic proliferation.

### Consequences of chronic alcohol misuse

#### Acute intoxication
- Emotional and behavioural disturbance
- Medical problems: hypoglycaemia, aspiration of vomit, respiratory depression
- Complication of other medical problems
- Accidents, injuries sustained in fights

#### Withdrawal phenomena
- Psychological symptoms: restlessness, anxiety, panic attacks
- Autonomic symptoms: tachycardia, sweating, pupil dilatation, nausea, vomiting
- Delirium tremens: agitation, hallucinations, illusions, delusions
- Seizures

#### Medical consequences
- Neurological: peripheral neuropathy, cerebellar degeneration, cerebral haemorrhage, dementia
- Hepatic: fatty change and cirrhosis, liver cancer
- Gastrointestinal: oesophagitis, gastritis, pancreatitis, oesophageal cancer, Mallory–Weiss syndrome, malabsorption, oesophageal varices
- Respiratory: pulmonary tuberculosis, pneumonia
- Skin: spider naevi, palmar erythema, Dupuytren’s contractures, telangiectasias
- Cardiac: cardiomyopathy, hypertension
- Musculoskeletal: myopathy, fractures
- Endocrine and metabolic: pseudo-Cushing’s syndrome, hypoglycaemia, gout
- Reproductive: hypogonadism, fetal alcohol syndrome, infertility

#### Psychiatric and cerebral consequences
- Depression
- Alcoholic hallucinosis
- Alcoholic ‘blackouts’
- Wernicke’s encephalopathy: nystagmus, ophthalmoplegia, ataxia, confusion
- Korsakoff’s syndrome: short-term memory deficits, confabulation

Effects on other organs
These are protean and virtually any organ can be involved (see Box 10.35). These effects are discussed in detail in the relevant chapters.

Management and prognosis
For the person misusing alcohol, provision of clear information from a doctor about the harmful effects of alcohol and the safe levels of consumption is often all that is needed. In more serious cases, patients may have to be advised to alter leisure activities or change jobs to help them to reduce their consumption. Psychological treatment is used for patients who have recurrent relapses and is usually available at specialised centres. Support to stop drinking is also provided by voluntary organisations, such as Alcoholics Anonymous (AA) in the UK.

Alcohol withdrawal syndromes can be prevented, or treated once established, with benzodiazepines. Large doses may be required (e.g. diazepam 20 mg 4 times daily), tailed off over a period of 5–7 days as...
symptoms subside. Prevention of the Wernicke-Korsakoff syndrome requires the immediate use of high doses of thiamine, which is initially given parenterally in the form of Pabrinex (two vials 3 times daily for 48 hours) and then orally (100 mg 3 times daily). There is no treatment for Korsakoff’s syndrome once it has arisen. The risk of side-effects, such as respiratory depression with benzodiazepines and anaphylaxis with Pabrinex, is small when weighed against the risks of no treatment.

Acamprosate (666 mg 3 times daily) may help to maintain abstinence by reducing the craving for alcohol. Disulfiram (200–400 mg daily) can be given as a deterrent to patients who have difficulty resisting the impulse to drink after becoming abstinent. It blocks the metabolism of alcohol, causing acetaldehyde to accumulate. When alcohol is consumed, an unpleasant reaction follows, with headache, flushing and nausea. Disulfiram is always an adjunct to other treatments, especially support psychotherapy. Treatment with antidepressants is always an adjunct to other treatments, especially supportive psychotherapy. Treatment with antidepressants may be required if depression is severe or does not resolve with abstinence. Antipsychotics (e.g. chlorpromazine 100 mg 3 times daily) are needed for alcoholic hallucinosis. Although such treatment may be successful, there is a high relapse rate.

Chronic alcohol misuse greatly increases the risk of death from accidents, disease and suicide (p. 100).

### Substance misuse disorder

Dependence on and misuse of both illegal and prescribed drugs is a major problem worldwide. Drugs of misuse are described in detail in Chapter 9. They can be grouped as follows.

#### Sedatives

These commonly give rise to physical dependence, the manifestations of which are tolerance and a withdrawal syndrome. Drugs include benzodiazepines, opiates (including morphine, heroin, methadone and dihydromorphine) and barbiturates (now rarely prescribed). Overdose with opiates and benzodiazepines can be fatal, primarily as a result of respiratory depression (Ch. 9). Withdrawal from opiates is notoriously unpleasant, and withdrawal from benzodiazepines (Box 10.36) and barbiturates may cause seizures.

Intravenous opiate users are prone to bacterial infections, hepatitis B (p. 950), hepatitis C (p. 954) and HIV infection (Ch. 14) through needle contamination. Accidental overdose is common, mainly because of the varied and uncertain potency of illicit supplies of the drug. The withdrawal syndrome, which can start within 12 hours of last use, presents with intense craving, rhinorrhea, lacrimation, yawning, perspiration, shivering, piloerection, vomiting, diarrhoea and abdominal cramps. Examination reveals tachycardia, hypertension, mydriasis and facial flushing.

#### Stimulants

Stimulant drugs include amphetamines and cocaine. They are less dangerous than the sedatives in overdose, although they can cause cardiac and cerebrovascular problems through their pressor effects. Physical dependence syndromes do not arise, but withdrawal causes a rebound lowering in mood and can give rise to an intense craving for further use, especially in any form of drug with a rapid onset and offset of effect, such as crack cocaine. Chronic ingestion can cause a paranoid psychosis similar to schizophrenia. A ‘toxic psychosis’ (delirium) can occur with high levels of consumption, and tactile hallucinations (formication) may be prominent.

#### Hallucinogens

The hallucinogens are a disparate group of drugs that cause prominent sensory disturbances. They include cannabis, ecstasy, lysergic acid diethylamide (LSD) and Psilocybin (magic mushrooms). A toxic confusional state can occur after heavy cannabis consumption. Acute psychotic episodes are well recognised, especially in those with a family or personal history of psychosis, and there is evidence that prolonged heavy use increases the risk of developing schizophrenia. Paranoid psychoses have been reported in association with ecstasy. A chronic psychosis has also been reported after regular LSD use.

#### Organic solvents

Solvent inhalation (glue sniffing) is popular in some adolescent groups. Solvents produce acute intoxication characterised by euphoria, excitement, dizziness and a floating sensation. Further inhalation leads to loss of consciousness; death can occur from the direct toxic effect of the solvent, or from asphyxiation if the substance is inhaled from a plastic bag.

#### Aetiology

Many of the aetiological factors for alcohol misuse also apply to drug dependence. The main factors are cultural pressures, particularly within a peer group, and availability of a drug. In the case of some drugs, medical over-prescribing (for example, of synthetic opiates) has increased their availability, but there has also been a relative decline in the price of illegal drugs. Most drug users take a range of drugs – so-called polydrug misuse.

#### Diagnosis

As with alcohol, the diagnosis either may be apparent from the history, or may only be made once the patient presents with a complication. Drug screening of samples of urine or blood can be valuable in confirming the diagnosis, especially if the patient persists in denial.

#### Management and prognosis

The first step is to determine whether patients wish to stop using the drug. If they do not, they can still benefit from advice about how to minimise harm from their habit: for example, how to obtain and use clean needles for those who inject. For those who are physically dependent on sedative drugs, substitute prescribing (using methadone, for example, in opiate dependence) may help stabilise their lives sufficiently to allow a
gradoal reduction in dosage until they reach abstinence. Some specialist units offer inpatient detoxification. For details of the medical management of overdose, see Chapter 9.

The drug lofexidine, a centrally acting α-agonist, can be useful in treating the autonomic symptoms of opiate withdrawal, as can clonidine, although this carries a risk of hypotension and is best used by specialists. Long-acting opiate antagonists, such as naltrexone, may also have a place, again in specialist hands, in blocking the euphoriant effects of the opiate, thereby reducing addiction.

In some cases, complete opiate withdrawal is not successful and the patient functions better if maintained on regular doses of oral methadone as an outpatient. This decision to prescribe long-term methadone should only be taken by a specialist, and carried out under long-term supervision at a specialist drug treatment centre.

Substitute prescribing is neither necessary nor possible for the hallucinogens and stimulants, but the principles of management are the same as those that should accompany prescribing for the sedatives. These include identifying problems associated with the drug misuse that may serve to maintain it, and intervening where possible. Intervention may be directed at physical illness, psychiatric comorbidity, social problems or family disharmony.

Relapsing patients and those with complications should be referred to specialist drug misuse services. Support can also be provided by self-help groups and voluntary bodies, such as Narcotics Anonymous in the UK.

**Personality disorders**

Personality refers to the set of characteristics and behavioural traits that best describes an individual’s patterns of interaction with the world. The intensity of particular traits varies from person to person, although many, such as shyness or irritability, are displayed to some degree by most people.

A personality disorder (PD) is diagnosed when an individual’s personality causes persistent and severe problems for the person or for others. For example, anxiety may be so pronounced that the individual rarely ventures into any situation where he or she fears scrutiny. Antisocial traits, such as disregard for the well-being of others and a lack of guilt concerning the adverse effects of one’s actions on others, if pronounced, may lead to damage to others and to criminal acts.

PD is classified into several subtypes (such as emotionally unstable, antisocial or schizotypal), depending on the particular behavioural traits in question. A patient who meets diagnostic criteria for one subtype commonly meets criteria for two or three others. As allocation to one particular subtype gives little guidance to management or prognosis, classification is of limited value. PD commonly accompanies other psychiatric conditions, making treatment of the latter more difficult and therefore affecting their prognosis.

**Aetiology**

Some personality disorders appear to have an inherited aspect (especially paranoid and schizotypal types) but most are more clearly related to an unsatisfactory upbringing and adverse childhood experiences.

**Management and prognosis**

Personality disorders usually persist throughout life and are not readily treated. However, they often become less extreme with age. Treatment options are limited but there is some evidence that emotionally unstable PD may also respond to dialectical behavioural therapy (an intensive type of CBT). Anxious (avoidant) and obsessive (anankastic) PD may benefit from prescription of anxiolytic drugs, while paranoid/schizotypal PD may benefit from treatment with low doses of antipsychotic agents.

**Eating disorders**

There are two well-defined eating disorders, anorexia nervosa (AN) and bulimia nervosa (BN); they share some overlapping features. Ninety per cent of cases are female. There is a much higher prevalence of abnormal eating behaviour in the population that does not meet diagnostic criteria for AN or BN. In developed societies, obesity is arguably a much greater problem but is usually considered to be more a disorder of lifestyle or physiology than a psychiatric disorder.

**Anorexia nervosa**

There is marked weight loss, arising from food avoidance, often in combination with bingeing, purging, excessive exercise, or the use of diuretics and laxatives. Body image is profoundly disturbed so that, despite emaciation, patients still feel overweight and are terrified of weight gain. These preoccupations are intense and pervasive, and the false beliefs may be held with a conviction approaching the delusional. Anxiety and depressive symptoms are common accompaniments. Downy hair (lanugo) may develop on the back, forearms and cheeks. Extreme starvation is associated with a wide range of physiological and pathological bodily changes. All organ systems may be affected, although the most serious problems are cardiac and skeletal (Box 10.37).

**Aetiology**

This is unknown but probably includes genetic and environmental factors, including, in many societies, the social pressure on women to be thin.

**Diagnosis**

The condition usually emerges in adolescence, with a marked female preponderance. Diagnostic criteria are shown in Box 10.38. Differential diagnosis is from other causes of weight loss, including psychiatric disorders such as depression, and medical conditions such as inflammatory bowel disease, malabsorption, hypopituitarism and cancer. The diagnosis is based on a pronounced fear of fatness despite being thin, and on the absence of alternative causes of weight loss.

**Management and prognosis**

The aims of management are to ensure patient’s physical well-being, whilst helping them to gain weight by addressing the beliefs and behaviours that maintain the
low weight. Treatment is usually given on an outpatient basis, inpatient treatment being indicated only if weight loss is intractable and severe (for example, less than 65% of normal), or if there is a risk of death from medical complications or from suicide. There is a limited evidence base for treatment, although CBT and family therapy are commonly used. Psychotropic drugs are of little benefit except in those with clear-cut comorbid depressive disorder.

Weight gain is best achieved in a collaborative fashion. Compulsory admission and refeeding (including tube feeding) are very occasionally resorted to when patients are at risk of death and other measures have failed. Whilst this may produce a short-term improvement in weight, it probably does not change long-term prognosis. About 20% of patients with AN have a good outcome, a further 20% develop a chronic intractable disorder and the rest have an intermediate outcome. There is a long-term mortality rate of 10–20%, either due to the complications of starvation or from suicide.

### Bulimia nervosa

In BN, patients are usually at or near normal weight (unlike in AN), but display a morbid fear of fatness associated with disordered eating behaviour. They recurrently embark on eating binges, often followed by corrective measures such as self-induced vomiting. The prevalence is similar to or slightly greater than that of AN, but only a small proportion of sufferers reach treatment services.

### Diagnosis

BN usually begins later in adolescence than AN, and is even more predominantly a female malady. Diagnostic criteria are shown in Box 10.38. Physical signs of repeated self-induced vomiting include pitted teeth (from gastric acid), calluses on knuckles (‘Russell’s sign’) and parotid gland enlargement. There are many associated physical complications, including the dental and oesophageal consequences of repeated vomiting, as well as electrolyte abnormalities, cardiac arrhythmias and renal problems (see Box 10.37).

### Management and prognosis

CBT achieves both short-term and long-term improvements. Guided self-help and interpersonal psychotherapy may also be of value. There is also evidence for benefit from the SSRI, fluoxetine, although high doses (60 mg daily) and long courses (1 year) may be required; this appears to be independent of the antidepressant effect.

Bulimia does not carry the mortality associated with AN, and few sufferers develop anorexia. At 10-year follow-up, approximately 10% are still unwell, 20% have a subclinical degree of BN, and the remainder have recovered.

### Puerperal disorders

There are three common psychiatric disorders that occur after childbirth. When managing these conditions, it is important always to consider both the mother and the baby, and their relationship (Box 10.39).

### Post-partum blues

These are characterised by irritability, labile mood and tearfulness. Most women are affected to some degree. Symptoms begin soon after childbirth, peak on about the fourth day and then resolve. They may be related to hormonal or psychological changes associated with childbirth. No treatment is required, other than to reassure the mother.
Further information

Post-partum depression
This occurs in 10–15% of women and within a month of delivery. Women with a previous history of depression are at risk. Explanation and reassurance are important. The usual psychological and drug treatments for depression should be considered, as well as practical help with childcare. If hospital admission is required, it should ideally be to a mother and baby unit. Further episodes of depression, both after childbirth and in response to other stressors, are likely.

Puerperal psychosis
This has its onset in the first 2 weeks after childbirth. It is a rare but serious complication affecting about 1 in 500 women and usually takes the form of a manic or depressive psychosis. There is an association with a personal or familial history of bipolar disorder. Delirium is rare with modern obstetric management but should still be considered in the differential diagnosis. Admission to a psychiatric mother and baby unit may be required. Management depends on the type of psychosis that presents. In addition, it is important to consider the welfare of the baby, especially if the mother has ideas of harming it. If so, the risk to the baby must be assessed and, if necessary, the baby temporarily removed. Most women recover but are at an increased (25%) risk of puerperal psychosis with the next pregnancy, and a 50% lifetime risk.

PSYCHIATRY AND THE LAW

Medicine takes place in a legal framework, made up of legislation (statute law) drafted by parliament or other governing bodies, and common law (case law) built up from court judgements over time. Psychiatry is similar to other branches of medicine in the applicability of common law but differs in that patients with psychiatric disorders can also be subject to legislative requirements to remain in hospital or to undergo treatments they refuse, such as the administration of antipsychotic drugs to a patient with acute schizophrenia who lacks insight, and whose symptoms and/or behaviour pose a risk to himself/herself or to others.

The UK has three different Mental Health Acts, covering England and Wales, Scotland, and Northern Ireland, and all of these have recently been revised. Other countries may have very different provisions. It is important for practitioners to be familiar with the relevant provisions that apply in their jurisdictions and are likely to arise in the clinical settings in which they work.

All the countries that make up the UK have also introduced Incapacity Acts in recent years, with detailed provisions covering medical treatments for patients incapable of consenting, whether this incapacity arises from physical or mental illness. In general, the guiding principle in British law is that people should be free to make their own decisions about medical treatment, except where their ability to decide is impaired by mental illness or physical incapacity, and where there are clear risks to the health and safety of themselves or others. Any restrictions or compulsions applied should be the minimum necessary, and they should only be applied for as long as is necessary; there should also be provisions for appeals and oversight.

Further information

Books

Websites
http://cebmh.warne.ox.ac.uk/cebmh/ Website of the Centre for Evidence-based Mental Health.
www.depressionalliance.org Information on depression.
www.rcpsych.ac.uk/info/index.htm Royal College of Psychiatrists: mental health information.
www.who.int/mental_health/ WHO website on mental health and brain disorders.
Clinical examination of the cancer patient 260
The ten hallmarks of cancer 262
Environmental and genetic determinants of cancer 266
Investigations 268
  Histology 268
  Imaging 269
  Biochemical markers 269
Presenting problems in oncology 269
  Palpable mass 270
  Weight loss and fever 271
  Finger clubbing 271
  Ectopic hormone production 271
  Neurological paraneoplastic syndromes 271
  Cutaneous manifestations of cancer 272
Emergency complications of cancer 272
Metastatic disease 274
Therapeutics in oncology 276
  Surgical treatment 276
  Systemic chemotherapy 276
  Radiation therapy 277
  Hormone therapy 278
  Immunotherapy 278
  Biological therapies 278
Specific cancers 279
  Breast cancer 279
  Ovarian cancer 280
  Endometrial cancer 281
  Cervical cancer 281
  Head and neck tumours 281
  Carcinoma of unknown origin 282
Observation
- Skin changes
- Ascites
- Cushingoid appearance
- Cachexia
- Dehydration

Cardiovascular
- Superior vena cava (SVC) obstruction
- Atrial fibrillation
- Pericardial effusion
- Hypo-/hypertension

Respiratory
- Stridor
- Consolidation
- Pleural effusion

Abdomen
- Surgical scars
- Umbilical nodule
- Mass in epigastrium
- Visible peristalsis
- Abdominal distension
- Ascites
- Hepatomegaly
- Splenomegaly
- Renal mass
- Pelvic or adnexal mass

Neurological
- Focal neurological signs
- Sensory deficit
- Spinal cord compression
- Memory deficit
- Personality change

Skeletal survey
- Focal bone tenderness (pelvis, spine, long bones)
- Wrist tenderness (hypertrophic pulmonary osteoarthropathy)

Periphery
- Calf tenderness, venous thrombosis
- Clubbing (if present in hands)
Clinical examination of the cancer patient

1 Examination of the skin

Important features of skin lesions that should alert suspicion include:
- Asymmetry: irregular shape
- Bleeding
- Border: not a smooth edge
- Colour: uneven, variegated or changing colour
- Diameter: > 6 mm in diameter or growing
- Itching or pain in a pre-existing mole

2 Abdominal examination

- Are there scars from previous surgery?
- Is the umbilicus everted, suggesting ascites?
- Is there a firm nodule at the umbilicus due to ovarian or gastric cancer metastasis, causing a Sister Mary Joseph’s nodule?
- Is there smooth hepatomegaly – possibly primary liver cancer or heart failure?
- Is the liver firm or knobbly, suggesting metastasis?
- Is the ascites too tense to demonstrate hepatomegaly?
- Are other masses palpable in the abdomen?
- Are there signs of obstruction or paralytic ileus with absence of bowel sounds?
- Palpate for inguinal nodes (occasionally involved in ovarian cancer)
- Percuss for flank dullness and shifting dullness
- Perform vaginal and rectal examinations to detect adnexal or rectal masses

3 Examination of the lymph nodes

4 Superior vena cava obstruction

- Venous distension of neck
- Elevated but non-pulsatile jugular venous pulse
- Venous distension of chest wall
- Facial oedema
- Cyanosis
- Plethora of face
- Oedema of arms

5 Pericardial effusion

- Tachycardia
- Falling blood pressure
- Rising jugular venous pressure
- Muffled heart sounds
- Kussmaul’s sign

6 Malignant pleural effusions

Large right pleural effusion

Inspection
Tachypnoea
Palpation
↓ Expansion on R
Trachea and apex may be moved to L
Percussion
Stony dull
R mid- and lower zones
Auscultation
Absent breath sounds and diminished or absent vocal resonance R base
Crackles above effusion

5 Examination of the lymph nodes

Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

3 Pre-auricular
Parotid
Submandibular
Submental
Anterior cervical
Posterior cervical
Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

6 Superior vena cava obstruction

- Venous distension of neck
- Elevated but non-pulsatile jugular venous pulse
- Venous distension of chest wall
- Facial oedema
- Cyanosis
- Plethora of face
- Oedema of arms

5 Pericardial effusion

- Tachycardia
- Falling blood pressure
- Rising jugular venous pressure
- Muffled heart sounds
- Kussmaul’s sign

6 Malignant pleural effusions

Large right pleural effusion

Inspection
Tachypnoea
Palpation
↓ Expansion on R
Trachea and apex may be moved to L
Percussion
Stony dull
R mid- and lower zones
Auscultation
Absent breath sounds and diminished or absent vocal resonance R base
Crackles above effusion

5 Examination of the lymph nodes

Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

3 Pre-auricular
Parotid
Submandibular
Submental
Anterior cervical
Posterior cervical
Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

6 Superior vena cava obstruction

- Venous distension of neck
- Elevated but non-pulsatile jugular venous pulse
- Venous distension of chest wall
- Facial oedema
- Cyanosis
- Plethora of face
- Oedema of arms

5 Pericardial effusion

- Tachycardia
- Falling blood pressure
- Rising jugular venous pressure
- Muffled heart sounds
- Kussmaul’s sign

6 Malignant pleural effusions

Large right pleural effusion

Inspection
Tachypnoea
Palpation
↓ Expansion on R
Trachea and apex may be moved to L
Percussion
Stony dull
R mid- and lower zones
Auscultation
Absent breath sounds and diminished or absent vocal resonance R base
Crackles above effusion

5 Examination of the lymph nodes

Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

3 Pre-auricular
Parotid
Submandibular
Submental
Anterior cervical
Posterior cervical
Supraclavicular
Axillary
Epitrochlear
Inguinal
Femoral
Popliteal fossa

6 Superior vena cava obstruction

- Venous distension of neck
- Elevated but non-pulsatile jugular venous pulse
- Venous distension of chest wall
- Facial oedema
- Cyanosis
- Plethora of face
- Oedema of arms

5 Pericardial effusion

- Tachycardia
- Falling blood pressure
- Rising jugular venous pressure
- Muffled heart sounds
- Kussmaul’s sign

6 Malignant pleural effusions

Large right pleural effusion

Inspection
Tachypnoea
Palpation
↓ Expansion on R
Trachea and apex may be moved to L
Percussion
Stony dull
R mid- and lower zones
Auscultation
Absent breath sounds and diminished or absent vocal resonance R base
Crackles above effusion
Cancer is a significant global health-care problem, with an estimated worldwide incidence of 10 million new cases per year, 46% of which are in developed countries. Mortality is high, with more than 7 million deaths per year. The global costs and socioeconomic impact are considerable. The most common solid organ malignancies arise in the lung, breast and gastrointestinal tract (Fig. 11.1), but the most common form worldwide is skin cancer. Tobacco is a major factor in the aetiology of 30% of cancers, including those of the lung, nasopharynx, bladder and kidney, and these could be prevented by smoking cessation. Diet and alcohol contribute to a further 30% of cancers, including those of the stomach, colon, oesophagus, breast and liver. Lifestyle modification could reduce these if steps were taken to avoid animal fat and red meat, reduce alcohol, increase fibre, fresh fruit and vegetable intake, and avoid obesity. Infections account for a further 15% of cancers, including those of the cervix, stomach, liver, nasopharynx and bladder, and some of these could be prevented by infection control and vaccination.

The formation and growth of cancer constitute a multi-step process, during which sequentially occurring gene mutations result in the formation of a cancerous cell. For cells to initiate carcinogenesis successfully, they require key characteristics, collectively referred to as the hallmarks of cancer.

1. **Genome instability and mutation**

Random genetic mutations occur continuously throughout all cells of the body and very rarely confer a selective advantage on single cells, allowing overgrowth and dominance in local tissue environments. Multistep carcinogenesis results from successive clonal expansions of pre-malignant cells, each expansion being triggered by acquisition of a random enabling genetic mutation. Under normal circumstances, cellular DNA repair mechanisms are so effective that almost all spontaneous mutations are corrected without producing phenotypic changes, keeping the overall mutation rates very low. In cancer cells, the accumulation of mutations can be accelerated by compromising the surveillance systems that normally monitor genomic integrity and force genetically damaged cells into either senescence or apoptosis. Therefore, they can become more sensitive to mutagenic actions or develop DNA repair mechanism failure.

2. **Resisting cell death**

There are three principal mechanisms through which cell death occurs in healthy tissues:

- **Apoptosis** is programmed cell death and is frequently found at markedly reduced rates in cancers, particularly those of high grade or those resistant to treatment. The cellular apoptotic system has regulatory elements which sense intrinsic and extrinsic pro-apoptotic signals and initiate a cascade of proteolysis and cell disassembly with nuclear fragmentation, chromosomal condensation, and shrinking of the cell with loss of intercellular contact, followed by cellular fragmentation and the formation of apoptotic bodies that are phagocytosed by neighbouring cells. The most important regulator of apoptosis is the TP53 tumour suppressor gene, often described as the ‘guardian of the genome’, as it is able to induce apoptosis in response to sufficient levels of genomic damage. The largest initiator of apoptosis via TP53 is cellular injury, particularly due to DNA damage from chemotherapy, oxidative damage and ultraviolet (UV) radiation.

- **Autophagy** is a catabolic process during which cellular constituents are degraded by lysosomal machinery within the cell. It is an important physiological mechanism, which usually occurs at low levels in cells but can be induced in response to environmental stresses, particularly radiotherapy and cytotoxic chemotherapy, which induce elevated levels of autophagy that are cytoprotective for malignant cells, thus impeding rather than perpetuating the killing actions of these stress situations. Severely stressed cancer cells have been shown to shrink via autophagy to a state of reversible dormancy.

- **Necrosis** is the premature death of cells and is characterised by the release of cellular contents into the local tissue microenvironment, in marked contrast to apoptosis, where cells are disassembled in a step-by-step fashion and the resulting cellular fragments phagocytosed. Necrotic cell death results in the recruitment of inflammatory immune cells, promotion of angiogenesis, cellular proliferation and tissue invasion. Necrotic cells also release stimulatory factors, which promote proliferation of neighbouring cells and can promote rather than inhibit carcinogenesis.
3. Sustaining proliferative signalling

Cancer cells can sustain proliferation beyond what would be expected for normal cells; this is typically due to growth factors, which are able to bind to cell surface-bound receptors that activate an intracellular tyrosine kinase-mediated signalling cascade, ultimately leading to changes in gene expression and promoting cellular proliferation and growth. Sustained proliferative capacity can result from over-production of growth factor ligands or receptors and production of structurally altered receptors, which can signal in the absence of ligand binding and activation of intracellular signalling pathway components so that signalling is no longer ligand-dependent.

The cell cycle

The cell cycle is comprised of four ordered, strictly regulated phases referred to as G₀ (gap 1), S (DNA synthesis), G₂ (gap 2) and M (mitosis) (Fig. 11.2). Normal cells grown in culture will stop proliferating and enter a quiescent state called G₀ once they become confluent or are deprived of serum or growth factors. The first gap phase (G₁) prior to the initiation of DNA synthesis represents the period of commitment that separates M and S phases as cells prepare for DNA duplication. Cells in G₀ and G₁ are receptive to growth signals, but once they have passed a restriction point, they are committed to enter DNA synthesis (S phase). Cells demonstrate arrest at different points in G₁ in response to different inhibitory growth signals. Mitogenic signals promote progression through G₁ to S phase, utilising phosphorylation of the retinoblastoma gene product (pRb). Following DNA synthesis, there is a second gap phase (G₂) prior to mitosis (M), allowing cells to repair errors that have occurred during DNA replication and thus preventing propagation of these errors to daughter cells. Although the duration of individual phases may vary, depending on cell and tissue type, most adult cells are in a G₀ state at any one time.

Cell cycle regulation

The cell cycle is orchestrated by a number of molecular mechanisms, most importantly by cyclins and cyclin-dependent kinases (CDKs). Cyclins bind to CDKs, and are regulated by both activating and inactivating phosphorylation, with two main checkpoints at G₁/S and G₂/M transition. The genes that inhibit progression play an important part in tumour prevention and are referred to as tumour suppressor genes (e.g. TP53, TP21, TP16 genes). The products of these genes deactivate the cyclin–CDK complexes and are thus able to halt the cell cycle. The complexity of cell cycle control is susceptible to dysregulation, which may produce a malignant phenotype.

Stimulation of the cell cycle

Many cancer cells produce growth factors, which drive their own proliferation by a positive feedback known as autocrine stimulation. Examples include transforming growth factor-alpha (TGF-α) and platelet-derived growth factor (PDGF). Other cancer cells express growth factor receptors at increased levels due to gene amplification or express abnormal receptors that are permanently activated. This results in abnormal cell growth in response to physiological growth factor stimulation or even in the absence of growth factor stimulation (ligand-independent signalling). The epidermal growth factor receptor (EGFR) is often over-expressed in lung and gastrointestinal tumours and the HER2/neu receptor is frequently over-expressed in breast cancer. Both receptors activate the Ras–Raf–MAP kinase pathway, causing cell proliferation.

Fig. 11.2 The cell cycle and sites of action of chemotherapeutic agents. (Rb = retinoblastoma gene; CDK = cyclin-dependent kinase)
4. Evading growth suppressors

In healthy tissues, cell-to-cell contact in dense cell populations acts as an inhibitory factor on proliferation. This contact inhibition is typically absent in many cancer cell populations. Growth-inhibitory factors can modulate the cell cycle regulators and produce activation of the CDK inhibitors, causing inhibition of the CDKs. Mutations within inhibitory proteins are common in cancer. Loss of restriction by disruption of pRb regulation can be found in human tumours, which produces a loss of restraint on transition from G1 to S phase of the cell cycle. Disruption of p53 function will have downstream effects on p21 that alter the coordination of DNA repair with cycle arrest and that result in the affected cell accumulating genomic defects. Down-regulation of p21 and p27, which can be found in tumours with normal p53 function, correlates notably with high tumour grade and poor prognosis.

5. Enabling replicative immortality

For cancer cells to evolve into macroscopic tumours, they need to acquire the ability for unlimited proliferation. Telomeric DNA sequences, which protect and stabilise chromosomal ends, play a central role in conferring this limitless replicative potential. During replication of normal cells, telomeres shorten progressively as small fragments of telomeric DNA are lost with successive cycles of replication. This shortening process is thought to represent a mitotic clock and eventually prevents the cell from dividing further. Telomerase, a specialised polymerase enzyme, adds nucleotides to telomeres, allowing continued cell division and thus preventing premature arrest of cellular replication. The telomerase enzyme is almost absent in normal cells but is expressed at significant levels in many human cancers.

6. Inducing angiogenesis

All cancers require a functional vascular network to ensure continued growth and will be unable to grow beyond 1 mm³ without stimulating the development of a vascular supply. Tumours require sustenance in the form of nutrients and oxygen, as well as an ability to evacuate metabolic waste products and carbon dioxide. This entails the development of new blood vessels, which is termed angiogenesis (Figs 11.3 and 11.4).

Angiogenesis is dependent on the production of angiogenic growth factors, of which vascular endothelial growth factor (VEGF) and platelet-derived endothelial growth factor (PDGF) are the best characterised. During tumour progression, an angiogenic switch is activated and remains on, causing normally quiescent vasculature to sprout new vessels continually that help sustain expanding tumour growth. Angiogenesis is governed by a balance of pro-angiogenic stimuli and angiogenesis inhibitors, such as thrombospondin (TSP)-1, which binds to transmembrane receptors on endothelial cells and evokes suppressive signals.

A number of cells can contribute to the maintenance of a functional tumour vasculature and therefore sustain angiogenesis. These include pericytes and a variety of bone marrow-derived cells such as macrophages, neutrophils, mast cells and myeloid progenitors.

---

**Fig. 11.3 Oncogenesis.** The multistep origin of cancer, showing events implicated in cancer initiation, progression, invasion and metastasis.
7. Activating invasion and metastasis

Invasion and metastasis are complex processes involving multiple discrete steps; it begins with local tissue invasion, followed by infiltration of nearby blood and lymphatic vessels by cancer cells. Malignant cells are eventually transported through haematogenous and lymphatic spread to distant sites within the body, where they form micrometastases that will eventually grow into macroscopic metastatic lesions (see Fig. 11.3).

Cadherin-1 (CDH1) is a calcium-dependent cell–cell adhesion glycoprotein that facilitates assembly of organised cell sheets in tissues, and increased expression is recognised as an antagonist of invasion and metastasis. In situ tumours usually retain cadherin-1 production, whereas loss of cadherin-1 production due to down-regulation or occasional mutational inactivation of CDH1 has been observed in human cancers, supporting the theory that CDH1 plays a key role in suppression of invasion and metastasis.

Cross-talk between cancer cells and cells of the surrounding stromal tissue is involved in the acquired capability for invasive growth and metastasis. Mesenchymal stem cells in tumour stroma have been found to secrete CCL5, a protein chemokine that helps recruit leucocytes into inflammatory sites. With the help of particular T-cell-derived cytokines (interleukin (IL)-2 and interferon (IFN)-γ), CCL5 induces proliferation and activation of natural killer cells and then acts reciprocally on cancer cells to stimulate invasive behaviour. Macrophages at the tumour periphery can foster local invasion by supplying matrix-degrading enzymes such as metalloproteinases and cysteine cathepsin proteases.

8. Reprogramming energy metabolism

Under aerobic conditions, oxidative phosphorylation functions as the main metabolic pathway for energy production; cells process glucose, first to pyruvate via glycolysis and thereafter to carbon dioxide in the mitochondria. Whilst under anaerobic conditions, glycolysis is favoured to produce adenosine triphosphate (ATP). Cancer cells can reprogramme their glucose metabolism to limit energy production to glycolysis, even in the presence of oxygen. This has been termed ‘aerobic glycolysis’. Up-regulation of glucose transporters, such as GLUT1, is the main mechanism through which aerobic glycolysis is achieved.

This reprogramming of energy metabolism appears paradoxical, as overall energy production from glycolysis is significantly lower (18-fold) than that from oxidative phosphorylation. One explanation may be that the increased production of glycolytic intermediates can be fed into various biosynthetic pathways, including those that generate the nucleosides and amino acids, necessary for the production of new cells.

9. Tumour-promoting inflammation

Almost all tumours show infiltration with immune cells on pathological investigation and historically this finding was thought to represent an attempt of the immune system to eradicate the cancer. It is now clear that tumour-associated inflammatory responses promote tumour formation and cancer progression.

Cytokines are able to alter blood vessels to permit migration of leucocytes (mainly neutrophils), in order to permeate from the blood vessels into the tissue, a process known as extravasation. Migration across the endothelium occurs via the process of diapedesis, where chemokine gradients stimulate adhered leucocytes to move between endothelial cells and pass through the basement membrane into the surrounding tissues. Once within the tissue interstitium, leucocytes bind to extracellular matrix proteins via integrins and CD44 to prevent their loss from the site.

As well as cell-derived mediators, several acellular biochemical cascade systems consisting of pre-formed plasma proteins act in parallel to initiate and propagate the inflammatory response. These include the
The immune system operates as a significant barrier to tumour formation and progression, and the ability to escape from immunity is a hallmark of cancer development. Cancer cells continuously shed surface antigens into the circulatory system, prompting an immune response that includes cytotoxic T cell, natural killer cell and macrophage production. The immune system is thought to provide continuous surveillance, with resultant elimination of cells that undergo malignant transformation.

However, deficiencies in the development or function of CD8+ cytotoxic T lymphocytes, CD4+ Th1 helper T cells, or natural killer cells can each lead to a demonstrable increase in cancer incidence. Also, highly immunogenic cancer cells may evade immune destruction by disabling components of the immune system. This is done through recruitment of inflammatory cells, including regulatory T cells and myeloid-derived suppressor cells, both actively immunosuppressive against the actions of cytotoxic lymphocytes (see Fig. 4.6, p. 80).

Cancers develop and progress when there is loss of recognition by the immune system, lack of susceptibility due to escape from immune cell action and induction of immune dysfunction, often via inflammatory mediators.

### Environmental and Genetic Determinants of Cancer

The majority of cancers do not have a single cause but rather are the result of a complex interaction between genetic factors and exposure to environmental carcinogens. These are often tumour type-specific but some general principles do apply.

#### Environmental factors

Environmental triggers for cancer have mainly been identified through epidemiological studies that examine patterns of distribution of cancers in patients in whom age, sex, presence of other illnesses, social class, geography and so on differ. Sometimes, these give strong
Environmental and genetic determinants of cancer

pointers to the molecular or cellular causes of the disease, such as the association between aflatoxin production within contaminated food supplies and hepatocellular carcinomas. However, for many solid cancers, such as breast and colorectal, there is evidence of a multifactorial pathogenesis, even when there is a principal environmental cause (Box 11.1).

Smoking is now established beyond all doubt as a major cause of lung cancer, but there are obviously additional predisposing factors since not all smokers develop cancer. Similarly, most carcinomas of the cervix are related to infection with human papilloma-virus (HPV subtypes 16 and 18). For carcinomas of the bowel and breast, there is strong evidence of an environmental component. For example, the risk of breast cancer in women of Far Eastern origin remains relatively low when they first migrate to a country with a Western lifestyle, but rises in subsequent generations to approach that of the resident population of the host country. The precise environmental factor that causes this change is unclear, but may include diet (higher intake of saturated fat and/or dairy products), reproductive patterns (later onset of first pregnancy) and lifestyle (increased use of artificial light and shift in diurnal rhythm).

**Genetic factors**

A number of inherited cancer syndromes are recognised that account for 5–10% of all cancers (Box 11.2). Their molecular basis is discussed in Chapter 3, but in general they result from inherited mutations in genes that regulate cell growth, cell death and apoptosis. Examples include the **BRCA1**, **BRCA2** and **AT** (ataxia telangiectasia) genes that cause breast and some other cancers, the **FAP** gene that causes bowel cancer, and the **Rb** gene that causes retinoblastoma. Although carriers of these gene mutations have a greatly elevated risk of cancer, none has 100% penetrance and additional modulating factors, both genetic and environmental, are likely to be operative. Exploration of a possible genetic contribution is a key part of cancer management, especially with regard to ascertaining the risk for an affected patient’s offspring.

### 11.2 Inherited cancer predisposition syndromes

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Malignancies</th>
<th>Inheritance</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxia telangiectasia</td>
<td>Leukaemia, lymphoma, ovarian, gastric, brain, colon</td>
<td>AR</td>
<td>AT</td>
</tr>
<tr>
<td>Breast/ovarian</td>
<td>Breast, ovarian, colonic, prostatic, pancreatic</td>
<td>AD</td>
<td>BRCA1, BRCA2</td>
</tr>
<tr>
<td>Bloom’s syndrome</td>
<td>Leukaemia, tongue, oesophageal, colonic, Wilms’ tumour</td>
<td>AR</td>
<td>BLM</td>
</tr>
<tr>
<td>Cowden’s syndrome</td>
<td>Breast, thyroid, gastrointestinal tract, pancreatic</td>
<td>AD</td>
<td>PTEN</td>
</tr>
<tr>
<td>Familial adenomatous polyposis</td>
<td>Colonic, upper gastrointestinal tract</td>
<td>AD</td>
<td>APC, MUTYH</td>
</tr>
<tr>
<td>Fanconi anaemia</td>
<td>Leukaemia, oesophageal, skin, hepatoma</td>
<td>AR</td>
<td>FACA, FACC, FACD</td>
</tr>
<tr>
<td>Gorlin’s syndrome</td>
<td>Basal cell skin, brain</td>
<td>AD</td>
<td>PTCH</td>
</tr>
<tr>
<td>Hereditary non-polyposis colon cancer (HNPPC)</td>
<td>Colonic, endometrial, ovarian, pancreatic, gastric</td>
<td>AD</td>
<td>MSH2, MLH1, MSH6, PMS1, PMS2</td>
</tr>
<tr>
<td>Li–Fraumeni syndrome</td>
<td>Sarcoma, breast, osteosarcoma, leukaemia, glioma, adrenocortical</td>
<td>AD</td>
<td>TP53</td>
</tr>
<tr>
<td>Melanoma</td>
<td>Melanoma</td>
<td>AD</td>
<td>CDK2 (TP16)</td>
</tr>
<tr>
<td>Multiple endocrine neoplasia (MEN)-1</td>
<td>Pancreatic islet cell, pituitary adenoma, parathyroid adenoma and hyperplasia</td>
<td>AD</td>
<td>MEN1</td>
</tr>
<tr>
<td>MEN-2</td>
<td>Medullary thyroid, phaeochromocytoma, parathyroid hyperplasia</td>
<td>AD</td>
<td>RET</td>
</tr>
<tr>
<td>Neurofibromatosis 1</td>
<td>Neurofibrosarcoma, phaeochromocytoma, optic glioma</td>
<td>AD</td>
<td>NF1</td>
</tr>
<tr>
<td>Neurofibromatosis 2</td>
<td>Vestibular schwannoma</td>
<td>AD</td>
<td>NF2</td>
</tr>
<tr>
<td>Papillary renal cell cancer syndrome</td>
<td>Renal cell cancer</td>
<td>AD</td>
<td>MET</td>
</tr>
<tr>
<td>Peutz–Jeghers syndrome</td>
<td>Colonic, ileal, breast, ovarian</td>
<td>AD</td>
<td>STK11</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Prostate</td>
<td>AD</td>
<td>HPC1</td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td>Retinoblastoma, osteosarcoma</td>
<td>AD</td>
<td>RB1</td>
</tr>
<tr>
<td>von Hippel–Lindau syndrome</td>
<td>Haemangioblastoma of retina and CNS, renal cell, phaeochromocytoma</td>
<td>AD</td>
<td>VHL</td>
</tr>
<tr>
<td>Wilms’ tumour</td>
<td>Nephroblastoma, neuroblastoma, hepatoblastoma, rhabdomyosarcoma</td>
<td>AD</td>
<td>WT1</td>
</tr>
<tr>
<td>Xeroderma pigmentosum</td>
<td>Skin, leukaemia, melanoma</td>
<td>AR</td>
<td>XPA, XPC, XPD (ERCC2), XPF</td>
</tr>
</tbody>
</table>

(AD = autosomal dominant; AR = autosomal recessive)
When a patient is suspected of having cancer, a full history should be taken; specific questions should be included as to potential risk factors such as smoking and occupational exposures. A thorough clinical examination is also essential to identify sites of metastases, and to discover any other conditions that may have a bearing on the management plan. In order to make a diagnosis and to plan the most appropriate management, information is needed on:

- the type of tumour
- the extent of disease, as assessed by staging investigations
- the patient’s general condition and any comorbidity.

The overall fitness of a patient is often assessed by the Eastern Cooperative Oncology Group (ECOG) performance scale (Box 11.3). The outcome for patients with a performance status of 3 or 4 is worse in almost all malignancies than for those with a status of 0–2, and this has a strong influence on the approach to treatment in the individual patient.

The process of staging determines the extent of the tumour; it entails clinical examination, imaging and in some cases surgery, to establish the extent of disease involvement. The outcome is recorded using a standard staging classification that allows comparisons to be made between different groups of patients. Therapeutic decisions and prognostic predictions can then be made using the evidence base for the disease. One of the most commonly used systems is the T (tumour), N (regional lymph nodes), M (metastatic sites) approach of the International Union against Cancer (UICC, Box 11.4). For some tumours, such as colon cancer, the Dukes system (p. 914) is used rather than the UICC classification.

### Histology

Histological analysis of a biopsy or resected specimen is pivotal in clinching the diagnosis and in deciding on the best form of management. The results of histological analysis are most informative when combined with knowledge of the clinical picture; therefore biopsy results should be reviewed and discussed within the context of a multidisciplinary team meeting.

### Light microscopy

Examination of tumour samples by light microscopy remains the core method of cancer diagnosis and, in cases where the primary site is unclear, may also give clues to the origin of the tumour:

- Signet-ring cells favour a gastric primary.
- Presence of melanin favours melanoma.
- Mucin is common in gut/lung/breast/endometrial cancers, but particularly common in ovarian cancer and rare in renal cell or thyroid cancers.
- Psammoma bodies are a feature of ovarian cancer (mucin +) and thyroid cancer (mucin −).

### Immunohistochemistry

Immunohistochemical (IHC) staining for tumour markers can provide useful diagnostic information and can help with treatment decisions. Commonly used examples of IHC in clinical practice include:

- Oestrogen (ER) and progesterone (PR) receptors. Positive results indicate that the tumour may be sensitive to hormonal manipulation.
- Alpha-fetoprotein (AFP) and human chorionic gonadotrophin (hCG) ± placental alkaline phosphatase (PLAP). These favour germ-cell tumours.
- Prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP). These favour prostate cancer.
- Carcinoembryonic antigen (CEA), cytokeratin and epithelial membrane antigen (EMA). These favour carcinomas.
- HER2 receptor. Breast cancers that have high levels of expression of HER2 indicate that the tumour may respond to trastuzumab (herceptin), an antibody directed against the HER2 receptor. The pattern of immunoglobulin, T-cell receptor and cluster designation (CD) antigen expression on the surface is also helpful in the diagnosis and classification of lymphomas. This can be achieved by IHC staining of biopsy samples or flow cytometry.
Electron microscopy

Electron microscopy (EM) can sometimes be of diagnostic value. Examples include the visualisation of melanosomes in amelanotic melanoma and dense core granules in neuro-endocrine tumours. EM may also help to distinguish adenocarcinoma from mesothelioma, as the ultrastructural properties of these two diseases are different (mesothelioma appears to have long, narrow, branching microvilli while adenocarcinomas appear to have short, stubby microvilli). EM is also useful for differentiating spindle-cell tumours (sarcomas, melanomas, squamous cell cancers) from small round-cell tumours, again due to their ultra-structural differences.

Cytogenetic analysis

Some tumours demonstrate typical chromosomal changes that help in diagnosis. The utilisation of fluorescent in situ hybridisation (FISH) techniques can be useful in Ewing’s sarcoma and peripheral neuroectodermal tumours where there is a translocation between chromosome 11 and 22-t(11; 22)(q24; q12). In some cases, gene amplification can also be detected via FISH (e.g. determining over-expression of HER2/neu).

Imaging

Imaging plays a critical role in oncology, not only in locating the primary tumour, but also in staging the disease. The imaging modality employed depends primarily on the site of the disease and likely patterns of spread, but usually more than one modality is required.

Radiography

Plain radiographs remain part of the initial workup, but have a limited role in defining disease extent and have been superseded by more sophisticated techniques.

Ultrasound

Ultrasound is useful in characterising lesions within the liver, kidney, pancreas and reproductive organs. It can be used for guiding biopsies of tumours in breast and liver. Endoscopic ultrasound is helpful in staging upper gastrointestinal and pancreatic cancers; it involves a special endoscope with an ultrasound probe attached.

Computerised tomography

Computerised tomography (CT) is a key investigation in cancer patients and is particularly useful in imaging the thorax and abdomen. With some modern scanners it is possible to visualise the bowel, and sometimes detection of colorectal adenomas and cancer is feasible.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) has a high resolution and because of this is the preferred technique for brain imaging. It is also used to image structures within the pelvis and is widely employed for staging of rectal, cervical and prostate cancers.

Positron emission tomography

Positron emission tomography (PET) visualises metabolic activity of tumour cells and is widely used, often in combination with CT (PET-CT), to evaluate patients with various cancers, including lung cancer and lymphoma (Fig. 11.5). It can accurately assess the severity and spread of cancer by detecting tumour metabolic activity following injection of small amounts of radioactive tracers such as fluorodeoxyglucose (FDG). In addition to having a role in diagnosis, PET can also be used in some patients to assess treatment response.

Biochemical markers

Many tumours produce substances called tumour markers, which can be used in diagnosis and surveillance. Some are useful in population screening, diagnosis, prognostication, treatment monitoring, detection of relapse and imaging of metastasis. Unfortunately, most tumour markers are not sufficiently sensitive or specific to be used in isolation and need to be interpreted in the context of the other clinical features. However, some can be used for antibody-directed therapy or imaging, where they have a greater role in diagnosis. Tumour markers in routine use are outlined in Box 11.5.
11.5 Commonly used serum tumour markers

<table>
<thead>
<tr>
<th>Name</th>
<th>Natural occurrence</th>
<th>Tumours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha-fetoprotein (AFP)</td>
<td>Glycoprotein found in yolk sac and fetal liver tissue. Transient elevation in liver diseases. Has a role in screening during pregnancy for the detection of neural tube defects and Down’s syndrome</td>
<td>Ovarian non-seminomatous germ cell tumours (80%), testicular teratoma (80%), hepatocellular cancer (50%)</td>
</tr>
<tr>
<td>Calcitonin</td>
<td>32 amino acid peptide from C cells of thyroid. Used to screen for MEN-2</td>
<td>Medullary cell carcinoma of thyroid</td>
</tr>
<tr>
<td>Cancer antigen 125 (CA-125)</td>
<td>Differentiation antigen of coelomic epithelium (Muller’s duct). Raised in any cause of ascites, pleural effusion or heart failure. Can be raised in inflammatory conditions</td>
<td>Ovarian epithelial cancer (75%), gastrointestinal cancer (10%), lung cancer (5%) and breast cancer (5%)</td>
</tr>
<tr>
<td>CA-19.9</td>
<td>A mucin found in epithelium of fetal stomach, intestine and pancreas. It is eliminated exclusively via bile and so any degree of cholostasis can cause levels to rise</td>
<td>Pancreatic cancer (80%), mucinous tumour of the ovary (65%), gastric cancer (30%), colon cancer (30%)</td>
</tr>
<tr>
<td>Carcinoembryonic antigen (CEA)</td>
<td>Glycoprotein found in intestinal mucosa during embryonic and fetal life. Elevated in smokers, cirrhosis, chronic hepatitis, ulcerative colitis, pneumonia</td>
<td>Colorectal cancer, particularly with liver metastasis, gastric cancer, breast cancer, lung cancer, mucinous cancer of the ovary</td>
</tr>
<tr>
<td>Human chorionic gonadotrophin (hCG)</td>
<td>Glycoprotein hormone, 14KD α subunit and 24KD β subunit from placental syncytiotrophoblasts. Used for disease monitoring in hydatidiform mole and as the basis of a pregnancy test</td>
<td>Choriocarcinoma (100%), hydatidiform moles (97%), ovarian non-seminomatous germ cell tumours (50–80%), seminoma (15%)</td>
</tr>
<tr>
<td>Placental alkaline phosphatase (PLAP)</td>
<td>Isoenzyme of alkaline phosphatase</td>
<td>Seminoma (40%), ovarian dygerminoma (50%)</td>
</tr>
<tr>
<td>Prostate-specific antigen (PSA)</td>
<td>Glycoprotein member of human kallikrein gene family. PSA is a serine protease that liquefies semen in excretory ducts of prostate. Can be elevated in benign prostatic hypertrophy and prostatitis</td>
<td>Prostate cancer (95%)</td>
</tr>
<tr>
<td>Thyroglobulin</td>
<td>Matrix protein for thyroid hormone synthesis in normal thyroid follicles</td>
<td>Papillary and follicular thyroid cancer</td>
</tr>
<tr>
<td>β-2-microglobulin</td>
<td>A human leucocyte antigen (HLA) common fragment present on surface of lymphocytes, macrophages and some epithelial cells. Can be elevated in autoimmune disease and renal glomerular disease</td>
<td>Non-Hodgkin’s lymphoma, myeloma</td>
</tr>
</tbody>
</table>

11.6 Local features of malignant disease

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Typical site or possible tumour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haemorrhage</td>
<td>Stomach, colon, bronchus, endometrium, bladder, kidney</td>
</tr>
<tr>
<td>Lump</td>
<td>Breast, lymph node (any site), testicle</td>
</tr>
<tr>
<td>Bone pain or fracture</td>
<td>Bone (primary sarcoma, secondary metastasis from breast, prostate, bronchus, thyroid, kidney)</td>
</tr>
<tr>
<td>Skin abnormality</td>
<td>Melanoma, basal cell carcinoma (rodent ulcer)</td>
</tr>
<tr>
<td>Ulcer</td>
<td>Oesophagus, stomach, anus, skin</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>Oesophagus, bronchus, gastric</td>
</tr>
<tr>
<td>Increasing constipation, abdominal discomfort or pain</td>
<td>Colon, rectum, ovary</td>
</tr>
<tr>
<td>Airway obstruction, stridor, cough, recurrent infection</td>
<td>Bronchus, thyroid</td>
</tr>
<tr>
<td>Odynophagia, early satiety, vomiting</td>
<td>Bronchus, stomach, oesophagus, colon, rectum</td>
</tr>
<tr>
<td>Abdominal swelling (ascites)</td>
<td>Ovary, stomach, pancreas</td>
</tr>
</tbody>
</table>

production of biologically active hormones by the tumour or as the result of an immune response to the tumour. The possible presentations are summarised in Boxes 11.6 and 11.7, and common presenting features discussed below. Although the incidence of cancer increases with patient age, the approach to investigation and management is similar at all ages (Box 11.8).

Palpable mass

A palpable mass detected by the patient or physician may be the first sign of cancer. Primary tumours of the thyroid, breast, testis and skin are often detected in this way, whereas palpable lymph nodes in the neck, groin
Presenting problems in oncology

11.7 Non-metastatic manifestations of malignant disease

<table>
<thead>
<tr>
<th>Feature</th>
<th>Common cancer site associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight loss and anorexia</td>
<td>Lung, gastrointestinal tract</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Any</td>
</tr>
<tr>
<td>Hypercalcaemia</td>
<td>Myeloma, breast, kidney</td>
</tr>
<tr>
<td>Prothrombotic tendency</td>
<td>Ovary, breast, gastrointestinal tract</td>
</tr>
<tr>
<td>SIADH Ectopic ACTH</td>
<td>Small cell lung cancer</td>
</tr>
<tr>
<td>Lambert–Eaton myasthenia-like syndrome</td>
<td>Small cell lung cancer</td>
</tr>
<tr>
<td>Subacute cerebellar degeneration</td>
<td>Small cell lung cancer, ovarian cancer</td>
</tr>
<tr>
<td>Acanthosis nigricans</td>
<td>Stomach, oesophagus</td>
</tr>
<tr>
<td>Dermatomyositis/polymyositis</td>
<td>Stomach, lung</td>
</tr>
</tbody>
</table>

(ACTH = adrenocorticotrophic hormone; SIADH = syndrome of inappropriate antidiuretic hormone secretion)

11.8 Cancer in old age

- **Incidence:** around 50% of cancers occur in the 15% of the population aged over 65 years.
- **Screening:** women aged over 65 in the UK are not invited to breast cancer screening but can request it. Uptake is low despite increasing incidence with age.
- **Presentation:** may be later for some cancers. When symptoms are non-specific, patients (and their doctors) may initially attribute them to age alone.
- **Life expectancy:** an 80-year-old woman can expect to live 8 years, so cancer may still shorten life and an active approach remains appropriate.
- **Prognosis:** histology, stage at presentation and observation for a brief period are better guides to outcome than age.
- **Rate of progress:** malignancy may have a more indolent course. This is poorly understood but may be due to reduced effectiveness of angiogenesis with age, inhibiting the development of metastases.
- **Response to treatment:** equivalent to that in younger people – well documented for a range of cancers and for surgery, radiotherapy, chemotherapy and hormonal therapy.
- **Treatment selection:** chronological age is of minor importance compared to comorbid illness and patient choice. Although older patients can be treated effectively and safely, aggressive intervention is not appropriate for all. Symptom control may be all that is possible or desired by the patient.

11.9 Ectopic hormone production by tumours

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Consequence</th>
<th>Tumours</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH</td>
<td>Hyponatraemia</td>
<td>SCLC</td>
</tr>
<tr>
<td>ACTH</td>
<td>Cushing’s syndrome</td>
<td>SCLC</td>
</tr>
<tr>
<td>FGF-23</td>
<td>Hypophosphataemic osteomalacia</td>
<td>Mesenchymal tumours</td>
</tr>
<tr>
<td>Insulin</td>
<td>Hypoglycaemia</td>
<td>Insulinoma</td>
</tr>
<tr>
<td>Erythropoietin</td>
<td>Polycythaemia</td>
<td>Kidney, hepatoma, cerebellar haemangioblastoma, uterine fibroids</td>
</tr>
<tr>
<td>PTHrP</td>
<td>Hypercalcaemia</td>
<td>NSCLC (squamous cell), breast, kidney</td>
</tr>
</tbody>
</table>

(ACTH = adrenocorticotrophic hormone; ADH = antidiuretic hormone; FGF = fibroblast growth factor; SCLC = non-small cell lung cancer; PTHrP = parathyroid hormone-related protein; NSCLC = small cell lung cancer)

Neurological paraneoplastic syndromes

These form a group of conditions associated with cancer thought to be due to an immunological response to the tumour that results in damage to the nervous system or muscle. The cancers most commonly implicated are as thyrotoxicosis, chronic inflammatory disease and chronic infective disorders. Fever can occur in any cancer secondary to infection, but may be a primary feature in Hodgkin’s disease, lymphoma, leukaemia, renal cancer and liver cancer. The presence of unexplained weight loss or fever warrants investigation to exclude the presence of occult malignancy.

Finger clubbing

Finger clubbing is a characteristic feature of lung cancer, and especially non-small cell lung cancer, although benign causes are recognised. It is often part of the wider process of hypertrophic osteoarthropathy in which there is periosteal new bone formation and arthritis due to increased levels of prostaglandin E. The diagnosis is primarily clinical, but X-rays show periosteal reaction and an isotope bone scan shows increased tracer uptake in the affected digits.

Ectopic hormone production

In some cases, the first presentation of cancer is with a metabolic abnormality due to ectopic production of hormones by tumour cells, including insulin, ACTH, ADH, fibroblast growth factor (FGF) 23, erythropoietin and parathyroid hormone-related protein (PTHrP). This can result in a wide variety of presentations, as summarised in Box 11.9. Further details on presentation and management of ACTH- and ADH-producing tumours are given on page 776, and of FGF23-producing tumours on page 1128. The management of hypercalcaemia associated with malignancy is discussed on page 273.

or axilla may indicate secondary spread of tumour. Hepatomegaly may be the first sign of primary liver cancer or tumour metastasis, whereas skin cancer may present as an enlarging or changing pigmented lesion.

**Weight loss and fever**

Unintentional weight loss is a characteristic feature of advanced cancer, but can be due to other causes such as thyrotoxicosis, chronic inflammatory disease and chronic infective disorders. Fever can occur in any cancer secondary to infection, but may be a primary feature in Hodgkin’s disease, lymphoma, leukaemia, renal cancer and liver cancer. The presence of unexplained weight loss or fever warrants investigation to exclude the presence of occult malignancy.
ONCOLOGY

those of the lung (small cell and non-small cell), pancreas, breast, prostate, ovary and lymphoma.

- *Peripheral neuropathy* results from axonal degeneration or demyelination.
- *Encephalomyelitis* can present with diverse symptoms, depending on which region of the brain is involved. Lumbar puncture shows raised protein in the cerebrospinal fluid (CSF) and a pleocytosis, predominantly that of lymphocytes. In some centres, flow cytometry of the CSF is also used to detect carcinomatous cells. MRI shows meningeal enhancement, particularly at the level of the brain stem, and anti-Hu antibodies may be detectable in serum. Encephalomyelitis is due to perivascular inflammation and selective neuronal degeneration. Most cases are caused by small cell lung cancer (75%).
- *Cerebellar degeneration* may be the presenting feature of an underlying malignancy and presents with rapid onset of cerebellar ataxia. Diagnosis is by MRI or CT, which may show cerebellar atrophy. Patients with these neurological paraneoplastic syndromes may be found to have circulating anti-Yo, Tr and Hu antibodies, but these are not completely specific and negative results do not exclude the diagnosis.
- *Retinopathy* is a rare complication of cancer and presents with blurred vision, episodic visual loss and impaired colour vision. If left untreated, it may lead to blindness. The diagnosis should be suspected if the electoretinogram is abnormal and anti-retinal antibodies are detected.
- *Lambert–Eaton syndrome* (LEMS) is due to underlying cancer in about 60% of cases. It presents with proximal muscle weakness that improves on exercise and is caused by the development of antibodies to pre-synaptic calcium channels (p. 1227). The diagnosis is made by electromyogram (EMG), which shows a low-amplitude compound muscle action potential that enhances to near normal following exercise.
- *Dermatomyositis* or *polymyositis* may be the first presentation of some cancers. Clinical features and management of these conditions are discussed in Chapter 28.

### Cutaneous manifestations of cancer

Many cancers can present with skin manifestations that are not due to metastases:

- *Pruritus* may be a presenting feature of lymphoma, leukaemia and CNS tumours.
- *Acanthosis nigricans* may precede cancers by many years and is particularly associated with gastric cancer.
- *Vitiligo* may be associated with malignant melanoma, and is possibly due to an immune response to melanocytes.
- *Pemphigus* may occur in lymphoma, Kaposi’s sarcoma and thymic tumours.
- *Dermatitis herpetiformis* associated with coeliac disease may precede tumour development by many years, and is associated with gastrointestinal lymphoma.

The clinical features and management of these skin conditions is discussed in Chapter 28.

### EMERGENCY COMPLICATIONS OF CANCER

#### Spinal cord compression

Spinal cord compression complicates 5% of cancers and is most common in myeloma, prostate, breast and lung cancers that involve bone. Cord compression often results from posterior extension of a vertebral body mass but intrathecal spinal cord metastases can cause similar signs and symptoms.

**Clinical features**

The earliest sign is back pain, particularly on coughing and lying flat. Subsequently, sensory changes develop in dermatomes below the level of compression and motor weakness distal to the block occurs. Finally, sphincter disturbance, causing urinary retention and bowel incontinence, is observed. Involvement of the lumbar spine may cause conus medullaris or cauda equina compression (Box 11.10). Physical examination reveals findings consistent with an upper motor neuron lesion, but lower motor neuron findings may predominate early on or in cases of nerve root compression.

**Management**

Spinal cord compression is a medical emergency and should be treated with analgesia and high-dose steroid therapy (Box 11.11). Neurosurgical treatment produces superior outcome and survival compared to radiotherapy alone, and should be considered first for all patients. Radiotherapy is used for the remaining patients and selected tumour types when the cancer is likely to be radiosensitive. The prognosis varies considerably, depending on tumour type, but the degree of neurological dysfunction at presentation is the strongest predictor of outcome irrespective of the underlying diagnosis.

#### 11.10 Comparison of features of neurological deficit

<table>
<thead>
<tr>
<th>Clinical feature</th>
<th>Spinal cord</th>
<th>Conus medullaris</th>
<th>Cauda equina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weakness</td>
<td>Symmetrical and profound</td>
<td>Symmetrical and variable</td>
<td>Asymmetrical, may be mild</td>
</tr>
<tr>
<td>Reflexes</td>
<td>Increased (or absent) knee and ankle reflexes with extensor plantar reflex</td>
<td>Increased knee reflex, decreased ankle reflex, extensor plantar reflex</td>
<td>Decreased knee and ankle reflexes with flexor plantar reflex</td>
</tr>
<tr>
<td>Sensory loss</td>
<td>Symmetrical, sensory level</td>
<td>Symmetrical, saddle distribution</td>
<td>Asymmetrical, radicular pattern</td>
</tr>
<tr>
<td>Sphincters</td>
<td>Late loss</td>
<td>Early loss</td>
<td>Often spared</td>
</tr>
<tr>
<td>Progression</td>
<td>Rapid</td>
<td>Variable</td>
<td>Variable</td>
</tr>
</tbody>
</table>
Ambulation can be preserved in more than 80% of patients who are ambulatory at presentation, but neurological function is seldom regained in patients with established deficits such as paraplegia.

Superior vena cava obstruction
Superior vena cava obstruction (SVCO) is a common complication of cancer that can occur through extrinsic compression or intravascular blockage. The most common causes of extrinsic compression are lung cancer, lymphoma and metastatic tumours. Patients with cancer can also develop SVCO due to intravascular blockage in association with a central catheter or thrombophilia secondary to the tumour.

Clinical features
The typical presentation is with oedema of the arms and face, distended neck and arm veins and dusky skin colouration over the chest, arms and face. Collaterals may develop over a period of weeks and the flow of blood in the collaterals helps to confirm the diagnosis. Headache secondary to cerebral oedema arising from the backflow pressure may also occur and tends to be aggravated by bending forward, stooping or lying down. The severity of symptoms is related to the rate of obstruction and the development of a venous collateral circulation. Accordingly, symptoms may develop rapidly or gradually. Clinical features are summarised in Box 11.12.

Investigations and management
The investigation of choice is CT of the thorax since it can clinch the diagnosis and distinguish between extra- and intravascular causes. A biopsy should be obtained when the tumour type is unknown because tumour type has a major influence on treatment. CT of the brain may be indicated if cerebral oedema is suspected. Tumours that are exquisitely sensitive to chemotherapy, such as germ cell tumours and lymphoma, can be treated with chemotherapy alone, but for most other tumours mediastinal radiotherapy is required. This relieves symptoms within 2 weeks in 50–90% of patients. In most centres, stenting is now increasingly favoured over radiotherapy, as it produces rapid results and can be repeated with reasonable effectiveness. This technique is particularly useful when dealing with tumours that are relatively chemoradio-resistant, such as non-small cell lung cancer or carcinoma of unknown primary. Where possible, these measures should be followed by treatment of the primary tumour, as long-term outcome is strongly dependent on the prognosis of the underlying cancer.

Hypercalcaemia
Hypercalcaemia is the most common metabolic disorder in patients with cancer and has a prevalence of 15–20 cases per 100000 persons. The incidence is highest in myeloma and breast cancer (approximately 40%), intermediate in non-small cell lung cancer, and uncommon in colon, prostate and small cell lung carcinomas. It is most commonly due to over-production of PTHrP, which binds to the PTH receptor and elevates serum calcium by stimulating osteoclastic bone resorption and increasing renal tubular reabsorption of calcium.

Clinical features
The symptoms of hypercalcaemia are often non-specific and may mimic those of the underlying malignancy. They include drowsiness, confusion, nausea and vomiting, constipation, polyuria, polydipsia and dehydration.

Investigations and management
The diagnosis is made by measuring serum total calcium and adjusting for albumin. It is especially important to correct for albumin in cancer because hypoalbuminaemia is common and total calcium values underestimate the level of ionised calcium. The principles of management are outlined in Box 11.13.

Patients should initially be treated with intravenous 0.9% saline to improve renal function and increase urinary calcium excretion. This alone often results in clinical improvement. Concurrently, intravenous bisphosphonates should be given to inhibit bone resorption. Calcitonin acts rapidly to increase calcium excretion and to reduce bone resorption and can be combined with fluid and bisphosphonate therapy for the first 24–48 hours in patients with life-threatening hypercalcaemia. Bisphosphonates will usually reduce the serum calcium levels to normal within 5 days, but if not, treatment can be repeated. The duration of action is up to 4 weeks and repeated therapy can be given at 3–4-weekly intervals as an outpatient. Hypercalcaemia is frequently a sign of tumour progression and the patient requires further investigation to establish disease status and review of the anti-cancer therapy.
Neutropenic fever

Neutropenia is a common complication of malignancy. It is usually secondary to chemotherapy but may occur with radiotherapy if large amounts of bone marrow are irradiated; it may also be a component of pancytopenia due to malignant infiltration of the bone marrow. Neutropenic fever is defined as a pyrexia of 38°C for over 1 hour in a patient with a neutrophil count < 1.0 x 10⁹/L. The risk of sepsis is related to the severity and duration of neutropenia and the presence of other risk factors such as intravenous or bladder catheters. Neutropenic fever is an emergency in cancer patients as, if left untreated, it can result in septicaemia with a high mortality rate.

Clinical features

The typical presentation is with high fever and affected patients are often non-specifically unwell. Examination is usually unhelpful in defining a primary source of the infection. Hypotension is an adverse prognostic feature and may progress to systemic circulatory shutdown and organ failure.

Investigations and management

An infection screen should be performed to include blood cultures (both peripheral and from central lines), urine culture, chest X-ray, and swabs for culture (throat, central line, wound). High-dose intravenous antibiotics should then be commenced, pending the results of cultures. Typical first-line empirical therapy consists of an anti-pseudomonal β-lactam (ceftazidime, cefotaxime or meropenem), or a combination of an aminoglycoside and a broad-spectrum penicillin with anti-pseudomonal activity (gentamicin and piperacillin), but this may need adjusting on the basis of local hospital policy and antibiotic resistance patterns. Metronidazole should be added if anaerobic infection is suspected, and flucloxacinilin or vancomycin or teicoplanin where Gram-positive infection is suspected (for example, in patients with central lines). If there is no response after 36–48 hours, treatment with amphotericin B or voriconazole should be considered to cover fungal infection. Antibiotics should be adjusted according to culture results, though these are often negative. Other supportive therapy, including intravenous fluids, inotrope therapy, ventilation or haemofiltration, may be required.

METASTATIC DISEASE

Metastatic disease is the major cause of death in cancer patients and the principal cause of morbidity. For the majority, the aim of treatment is palliative, but treatment of a solitary metastasis can occasionally be curative.

Brain metastases

Brain metastases occur in 10–30% of adults and 6–10% of children with cancer, and are an increasingly important cause of morbidity. Tumours that typically metastasise to the brain are shown in Box 11.14. Most involve the brain parenchyma but can also affect the cranial nerves, the blood vessels and other intracranial structures. In cases of solitary metastasis to the brain, the use of surgery and adjuvant radiotherapy has been shown to increase survival. However, practices vary for patients with more advanced brain metastases. In these cases, median survival without treatment is approximately 1 month. Steroids can increase survival to 2–3 months and whole-brain radiotherapy improves survival to 3–6 months, but the true efficacy of these interventions has not been proven adequately in a randomised trial setting. Patients with brain metastases as the only manifestation of an undetected primary tumour have a more favourable prognosis, with an overall median survival of 13.4 months. Tumour type also influences prognosis; breast cancer patients have a better prognosis than those with other types of tumour, and those with colorectal carcinoma tend to have a poorer prognosis.

Clinical features

Presentation is with headaches (40–50%), focal neurological dysfunction (20–40%), cognitive dysfunction (35%), seizures (10–20%) and papilloedema (< 10%).

Investigations and management

The diagnosis can be confirmed by CT or contrast-enhanced MRI. Treatment options include high-dose steroids (dexamethasone 4 mg 4 times daily) for tumour-associated oedema, anticonvulsants for seizures, whole-brain radiotherapy, and chemotherapy. Surgery may be considered for single sites of disease and can be curative; stereotactic radiotherapy may also be considered for solitary site involvement where surgery is not possible.

Liver metastases

These are common in breast cancer, colon cancer and tumours of the head and neck. The presentation is usually with a lesion on chest X-ray or CT. Solitary lesions require investigation, as single metastases can be difficult to distinguish from a primary lung tumour. Patients with two or more pulmonary nodules can be assumed to have metastases. The approach to treatment depends on the extent of disease in the lung and elsewhere. For solitary lesions, surgery should be considered with a generous wedge resection. Radiotherapy, chemotherapy or endocrine therapy can be used as systemic treatment and is dependent on the underlying primary cancer diagnosis.

Liver metastases

Metastatic cancer in the liver can represent the sole or life-limiting component of disease for many with colorectal cancer, ocular melanoma, neuro-endocrine tumours and, less commonly, other tumour types. The most common clinical presentations are with right upper quadrant pain due to stretching of the liver capsule,
jaundice, deranged liver function tests or an abnormality detected on imaging. In selected cases, resection of the metastasis can be contemplated. In colorectal cancer, successful resection of a metastasis improves 5-year survival from 3% to 30–40%. Other techniques, such as chemoembolisation or radiofrequency ablation, can also be used, provided the number and size of metastases remain small. If these are not feasible, symptoms may respond to systemic chemotherapy.

Bone metastases

Bone is the third most common organ involved by metastasis, after lung and liver. Bone metastases are a major clinical problem in patients with myeloma and breast or prostate cancers, but other tumours that commonly metastasise to bone include those of the kidney and thyroid. Bone metastases are an increasing management problem in other tumour types that do not classically target bone, due to the prolonged survival of patients generally. Accordingly, effective management of bony metastases has become a focus in the treatment of patients with many incurable cancers.

Clinical features

The main presentations are with pain, pathological fractures and spinal cord compression (p. 272). The pain tends to be progressive and worst at night, and may be partially relieved by activity, but subsequently becomes more constant in nature and is exacerbated by movement. Most pathological fractures occur in metastatic breast cancer (53%); other tumour types associated with fracture include the kidney (11%), lung (8%), thyroid (5%), lymphoma (5%) and prostate (3%).

Investigations and management

The most sensitive way of detecting bone metastases is by isotope bone scan. This can have false-positive results in multiple myeloma due to suppression of osteoblast activity. Therefore plain X-ray films are preferred for any sites of bone pain, as lytic lesions may not be detected by a bone scan. In patients with a single lesion, it is especially important to perform a biopsy to obtain a tissue diagnosis, since primary bone tumours may look very similar to metastases on X-ray.

The main goals of management are:

- pain relief
- preservation and restoration of function
- skeletal stabilisation
- local tumour control (e.g. relief of tumour impingement on normal structure)

Surgical intervention may be warranted where there is evidence of skeletal instability (e.g. anterior or posterior spinal column fracture) or an impending fracture (e.g. large lytic lesion on a weight-bearing bone with more than 50% cortical involvement). Intravenous bisphosphonates ( pamidronate, zoledronic acid or ibandronate) are widely used for bone metastases and are effective at improving pain and in reducing further skeletal related events, such as fractures and hypercalcaemia (Box 11.15). In certain types of cancer, such as breast and prostate, hormonal therapy may be effective. Radiotherapy, in the form of external beam therapy or systemic radionucleotides (strontium treatment), can also be useful for these patients. In some settings (e.g. breast carcinoma), chemotherapy may also be used in the management of bony metastases.

Malignant pleural effusion

This is a common complication of cancer and 40% of all pleural effusions are due to malignancy. The most common causes are lung and breast cancers, and the presence of an effusion indicates advanced and incurable disease. The presentation may be with dyspnoea, cough or chest discomfort, which can be dull or pleuritic in nature. Diagnosis and management of ascites is discussed on p. 938.

Investigations and management

Pleural aspirate is the key investigation and may show the presence of malignant cells. Malignant effusions are commonly blood-stained and are exudates with a raised fluid to serum lactate dehydrogenase (LDH) ratio (>0.6) and a raised fluid to serum protein ratio (>0.5). Treatment should focus on palliation of symptoms and be tailored to the patient’s physical condition and prognosis. Aspiration alone may be an appropriate treatment in frail patients with a limited life expectancy (Box 11.16). Those who present with malignant pleural effusion as the initial manifestation of breast cancer, small cell lung cancer, germ cell tumours or lymphoma should

---

**Box 11.15 Use of bisphosphonates in bony metastases**

‘The use of bisphosphonates in cancer patients with bony metastases results in decreased pain and a decrease in skeletal-related events.’

have the fluid aspirated and should be given systemic chemotherapy to try to treat disease in the pleural space. Treatment options for patients with recurrent pleural effusion include pleurodesis, pleurectomy and pleuroperitoneal shunt. Ideally, pleurodesis should be attempted once effusions recur after initial drainage.

**THERAPEUTICS IN ONCOLOGY**

Anti-cancer therapy may be either curative or palliative, and this distinction influences the approach to management of individual patients. The goal of treatment should be recorded in the medical notes.

- **Palliative chemotherapy** is the most common treatment and is primarily used to treat patients with metastasis. The goal is an improvement in symptoms with a focus on improving quality of life, and any survival increments are secondary. As a result, the treatment should be well tolerated and should aim to minimise adverse effects.
- **Adjuvant chemotherapy** is given after an initial intervention that is designed to cyto-reduce the tumour bulk and remove all macroscopic disease. Chemotherapy is then given with the intention of eradicating the micrometastatic disease that remains. The focus is on achieving an improvement in disease-free and overall survival.
- **Neoadjuvant chemotherapy** or primary medical therapy is where chemotherapy is administered first before a planned cytoreductive procedure. This can result in a reduced requirement for surgery, increase the likelihood of successful debulking, reduce the duration of hospitalisation and improve the fitness of the patient prior to interval debulking. This approach has the same goals as adjuvant treatment but creates an opportunity for translational research to measure responses to treatment and correlate with subsequent specimens removed at the time of surgery.
- **Chemoprevention** is the use of pharmacological agents to prevent cancer developing in patients identified as being at particular risk. Therefore the agents used aim to modify risk and, as such, should not have significant adverse effects.

**Surgical treatment**

Surgery has a pivotal role in the management of cancer. There are three main situations in which it is necessary.

**Biopsy**

In the vast majority of cases, a histological or cytological diagnosis of cancer is necessary, and tissue will also provide important information such as tumour type and differentiation, to assist subsequent management. Cytology can be obtained with fine needle aspiration but a biopsy is usually preferred. This can be a core biopsy, an image-guided biopsy or an excision biopsy.

**Excision**

The main curative management of most solid cancers is surgical excision. In early, localised cases of colorectal, breast and lung cancer, cure rates are high with surgery. There is increasing evidence that outcome is related to surgical expertise, and most multidisciplinary teams include surgeons experienced in the management of a particular cancer. There are some cancers for which surgery is one of two or more options for primary management, and the role of the multidisciplinary team is to recommend appropriate treatment for a specific patient. Examples include prostate and transitional cell carcinoma of the bladder, in which radiotherapy and surgery may be equally effective.

**Palliation**

Surgical procedures are often the quickest and most effective way of palliating symptoms. Examples include the treatment of faecal incontinence with a defunctioning colostomy; fixation of pathologica fractures and decompression of spinal cord compression; and the treatment of fungating skin lesions by ‘toilet’ surgery. A more specialist role for surgery is in resection of residual masses after chemotherapy and, in very selected cases, resection of metastases.

**Systemic chemotherapy**

Chemotherapeutic drugs are classified by their mode of action. They have the greatest activity in proliferating cells and this provides the rationale for their use in the treatment of cancer. Chemotherapeutic agents are not specific for cancer cells, however, and the side-effects of treatment are a result of their anti-proliferative actions in normal tissues such as the bone marrow, skin and gut.

**Combination therapy**

In order to overcome drug resistance and to limit the side-effects of different drugs, chemotherapy is most commonly given as a combination of agents. Combinations usually include drugs from different classes, with the aim of targeting several pathways and gaining maximum therapeutic effect. Drugs are conventionally given by intravenous injection every 3–4 weeks, allowing enough time for the patient to recover from short-term toxic effects before the next dose. Between four and eight such cycles of treatment are usually given in total. More recently, other strategies have been developed. For example, 5-fluorouracil (5-FU), which has a very short half-life, has increased efficacy when given by continuous intravenous infusion, using a semi-permanent in-dwelling intravenous catheter. However, the use of such catheters is not without risk and the potential of oral 5-FU is now being explored, using precursors such as capecitabine. Schedules of administration at weekly or 2-weekly intervals have also found their place in the management of both solid and haematological malignancies. Each tumour type has specific regimens that are used at various stages of the disease.

**Mode of administration**

Most drugs have to be given intravenously, and many are vesicant or locally irritant if there is an extravasation. Chemotherapy should be administered into a vein in which the infusion is free-flowing to minimise the risk of extravasation. A few patients require central venous
catheters due to the nature of their treatment or poor vascular access. Patients who receive chemotherapy through a peripheral line must be carefully observed, and the chemotherapy stopped at the first sign of any extravasation. Chemotherapy is potentially dangerous to the person giving the therapy, because cytotoxics are carcinogenic and teratogenic. In view of this, policies must be in place for the use of gloves and aprons and for the safe disposal of syringes containing cytotoxics. Other oral chemotherapeutic agents have been developed over the past 30 years, although not many have replaced their intravenous counterparts.

**Adverse effects**

Most cytotoxics have a narrow therapeutic window or index and can have significant adverse effects, as shown in Figure 11.6. Considerable supportive therapy is often required to enable patients to tolerate therapy and achieve benefit. Nausea and vomiting are common, but with modern antiemetics, regimens such as the combination of dexamethasone and highly selective 5-hydroxytryptamine (5-HT₃) receptor antagonists like ondansetron, most patients now receive chemotherapy without any significant problems. Myelosuppression is common to almost all cytotoxics. This not only limits the dose of drug, but also can cause life-threatening complications. The risk of neutropenia can be reduced with the use of specific growth factors that accelerate the re-population of myeloid precursor cells. The most commonly employed is granulocyte-colony-stimulating factor (G-CSF), which is widely used in conjunction with chemotherapy regimens that induce a high rate of neutropenia. More recently, it has also been used to ‘accelerate’ the administration of chemotherapy, enabling standard doses to be given at shorter intervals where the rate-limiting factor has been the time taken for the peripheral neutrophil count to recover. Accelerated chemotherapy regimens have now been demonstrated to offer therapeutic advantages in small cell lung cancer, lymphoma and possibly breast cancer.

**Radiation therapy**

Radiation therapy (radiotherapy) involves treating the cancer with ionising radiation; for certain localised cancers it may be curative. Ionising radiation can be delivered by radiation emitted from the decay of radioactive isotopes or by high-energy radiation beams, usually X-rays. Three methods are usually employed:

- **Teletherapy**: application from a distance by a linear accelerator.
- **Brachytherapy**: direct application of a radioactive source on to or into a tumour. This allows the delivery of a very high, localised dose of radiation and is integral to the management of localised cancer.

---

**Fig. 11.6** Adverse effects of chemotherapy and radiotherapy. Acute effects are shown in pink and late effects in blue.
ONCOLOGY

cancers of the head and neck and cancer of the cervix and endometrium.

• Intravenous injection of a radioisotope: such as ¹³¹iodine for cancer of the thyroid and ⁹⁰strontium for the treatment of bone metastases from prostate cancer.

The majority of treatments are now delivered by linear accelerators; these produce electron or X-ray beams of high energy that are used to target tumour tissue. Whatever the method of delivery, the biological effect of ionising radiation is to cause lethal and sublethal damage to DNA. Since normal tissues are also radiosensitive, treatment has to be designed to maximise exposure of the tumour and minimise exposure of normal tissues. This is now possible with modern imaging techniques such as CT and MRI, which allow better visualisation of normal and tumour tissue. In addition, techniques such as conformal radiotherapy, where shaped rather than conventional square or rectangular beams are used, allow much more precise targeting of therapy to the tumour, and reduce the volume of normal tissue irradiated by up to 40% compared to non-conformal techniques.

Biological differences between normal and tumour tissues are also used to obtain therapeutic gain. Fundamental to this is fractionation, which entails delivering the radiation as a number of small doses on a daily basis. This allows normal cells to recover from irradiation damage but recovery occurs to a lesser degree in malignant cells. Fractionation regimens vary from centre to centre but radical treatments given with curative intent are often delivered in 20–30 fractions given daily, 5 days a week over 4–6 weeks. Radiotherapy can also be extremely useful for the alleviation of symptoms, and for palliative treatments such as this a smaller number of fractions (1–5) is usually adequate.

Both normal and malignant tissues vary widely in their sensitivity to radiotherapy. Germ cell tumours and lymphomas are extremely radiosensitive and relatively low doses are adequate for cure, but most cancers require doses close to or beyond that which can be tolerated by adjacent normal structures. Normal tissue also varies in its radiosensitivity, the central nervous system, small bowel and lung being amongst the most sensitive. The side-effects of radiotherapy (see Fig. 11.6) depend on the normal tissues treated, their radiosensitivity and the dose delivered.

Adverse effects

An acute inflammatory reaction commonly occurs towards the end of most radical treatments and is localised to the area treated. For example, skin reactions are common with breast or chest wall radiotherapy, and proctitis and cystitis with treatment to the bladder or prostate. These acute reactions settle over a period of a few weeks after treatment, assuming normal tissue tolerance has not been exceeded. Late effects of radiotherapy develop 6 weeks or more after treatment and occur in 5–10% of patients. Examples include brachial nerve damage and subcutaneous fibrosis after breast cancer treatment, and shrinkage and fibrosis of the bladder after treatment for bladder cancer. There is a risk of inducing cancer after radiotherapy, which varies depending on the site treated and whether the patient has had other treatment such as chemotherapy.

Hormone therapy

Hormone therapy is most commonly used in the treatment of breast cancer and prostate cancer. Breast tumours that are positive for expression of the oestrogen receptor (ER) respond well to anti-oestrogen therapy, and assessment of ER status is now standard in the diagnosis of breast cancer. Several drugs are now available that reduce oestrogen levels or block the effects of oestrogen on the receptor. When targeted appropriately, adjuvant hormone therapy reduces the risk of relapse and death at least as much as chemotherapy, and in advanced cases can induce stable disease and remissions that may last months to years, with acceptable toxicity. Hormonal manipulation may be effective in other cancers. In prostate cancer, hormonal therapy (e.g. luteinising hormone releasing hormone (LHRH) analogues such as goserelin and/or anti-androgens such as bicalutamide) aimed at reducing androgen levels can provide good long-term control of advanced disease, but there is no convincing evidence that it is an effective therapy following potentially curative surgery. Progestogens are active in the treatment of endometrial and breast cancer. In the metastatic setting, progestogen use (e.g. megestrol acetate) is associated with response rates of 20–40% in endometrial cancer. In breast cancer, progestogens are used in patients whose disease has progressed with conventional anti-oestrogen therapy. Their exact mechanism in this setting is not fully understood.

Immunotherapy

A profound stimulus to the patient’s immune system can sometimes alter the natural history of a malignancy, and the discovery of interferons was the impetus for much research. Although solid tumours show little benefit, interferons are active in melanoma and lymphoma, and there is evidence that they are beneficial as adjuvants (after surgery and chemotherapy respectively) to delay recurrence. Whether interferon-induced stimulation of the immune system is capable of eradicating microscopic disease remains unproven. More powerful immune responses can be achieved with potent agents like interleukin-2 (IL-2), but the accompanying systemic toxicity is a problem still to be overcome. The most striking example of successful immunotherapy is that with rituximab, an antibody against the common B-cell antigen CD20. It increases complete response rates and improves survival in diffuse large cell non-Hodgkin’s lymphoma when combined with chemotherapy, and is also effective in palliating advanced follicular non-Hodgkin’s lymphoma (p. 1043).

Biological therapies

Advances in knowledge about the molecular basis of cancer have resulted in the development of a new generation of treatments to block the signalling pathways responsible for the growth of specific tumours. This has created the potential to target cancer cells more selectively, with reduced toxicity to normal tissues. Some
examples are discussed below, but in the years to come many more such agents will come into clinical use, with the potential to revolutionise our approach to some cancers.

**Gefitinib/erlotinib**

These agents inhibit the activity of the epidermal growth factor receptor, which is over-expressed in many solid tumours. However, the drugs’ activity does not depend on the amount of receptor over-expression but on factors such as gene copy number and mutation status.

**Imatinib**

Imatinib was developed to inhibit the **BCR-ABL** gene product, tyrosine kinase, that is responsible for chronic myeloid leukaemia (p. 1039), and it does this extremely effectively. It is also active in gastrointestinal stromal tumour (GIST), a type of sarcoma that has over-expression of another cell surface tyrosine kinase, c-kit. This agent has good tolerability and is particularly useful in GIST, where conventional chemotherapy is less effective.

**Bevacizumab**

This is a humanised monoclonal antibody that inhibits vascular endothelial growth factor A (VEGF-A), a key stimulant of angiogenesis in tumours. This has activity in colorectal, lung, breast, renal and ovarian cancers, although the licence was subsequently revoked for breast cancer; while bevacizumab slows the rate of progression of metastatic breast cancer, it had little impact on survival or improved quality of life.

**Trastuzumab**

Trastuzumab (herceptin) targets the HER2 receptor, an oncogene that is over-expressed in around one-third of breast cancers and in a number of other solid tumours. It is effective as a single-agent therapy, but also improves survival in patients with advanced breast cancer when used in conjunction with chemotherapy. Unfortunately, trastuzumab can induce cardiac failure by an unknown biological mechanism, especially in combination with doxorubicin.

**SPECIFIC CANCERS**

The diagnosis and management of cancers are discussed in more detail elsewhere in the book (Box 11.17). Here we discuss the clinical features, pathogenesis and management of common tumours that are not covered elsewhere.

**Breast cancer**

Globally, the incidence of breast cancer is only second to that of lung cancer and the disease represents the leading cause of cancer-related deaths among women. Invasive ductal carcinoma with or without ductal carcinoma in situ (DCIS) is the most common histology, accounting for 70%, whilst invasive lobular carcinoma accounts for most of the remaining cases. DCIS constitutes 20% of breast cancers detected by mammography screening. It is multifocal in one-third of women and has a high risk of becoming invasive (10% at 5 years following excision only). Pure DCIS does not cause lymph node metastases, although these are found in 2% of cases where nodes are examined, owing to undetected invasive cancer. Lobular carcinoma in situ (LCIS) is a predisposing risk factor for developing cancer in either breast (7% at 10 years). The survival for breast cancer by stage is outlined in Box 11.18.

**Pathogenesis**

Both genetic and hormonal factors play a role; about 5–10% of breast cancers are hereditary and occur in patients with mutations of **BRCA1**, **BRCA2**, **AT** or **TP53** genes. Prolonged oestrogen exposure associated with early menarche, late menopause and use of hormone replacement therapy (HRT) has been associated with an increased risk. Other risk factors include obesity, alcohol intake, nulliparity and late first pregnancy. There is no definite evidence linking use of the contraceptive pill to breast cancer.

**SPECIFIC CANCERS covered in other chapters**

<table>
<thead>
<tr>
<th>Tumour</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder cancer</td>
<td>516</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>910</td>
</tr>
<tr>
<td>Familial cancer syndromes</td>
<td>66</td>
</tr>
<tr>
<td>Gastric cancer</td>
<td>877</td>
</tr>
<tr>
<td>Hepatocellular carcinoma</td>
<td>966</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>1035</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>699</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1041</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>719</td>
</tr>
<tr>
<td>Myeloma</td>
<td>1046</td>
</tr>
<tr>
<td>Oesophageal cancer</td>
<td>870</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>895</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>517</td>
</tr>
<tr>
<td>Renal cancer</td>
<td>515</td>
</tr>
<tr>
<td>Seminoma</td>
<td>518</td>
</tr>
<tr>
<td>Skin cancer</td>
<td>1269</td>
</tr>
<tr>
<td>Teratoma</td>
<td>518</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>754</td>
</tr>
</tbody>
</table>

**Five-year survival rates for breast cancer by stage**

<table>
<thead>
<tr>
<th>Tumour stage</th>
<th>Stage definition</th>
<th>5-year survival (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage I</td>
<td>Tumour &lt; 2 cm, no lymph nodes</td>
<td>96</td>
</tr>
<tr>
<td>Stage II</td>
<td>Tumour 2–5 cm, and/or mobile</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>axillary lymph nodes</td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>Chest wall or skin fixation, and/or</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>fixed axillary lymph nodes</td>
<td></td>
</tr>
<tr>
<td>Stage IV</td>
<td>Metastasis</td>
<td>18</td>
</tr>
</tbody>
</table>
Clinical features

Breast cancer usually presents as a result of mammographic screening or as a palpable mass with nipple discharge in 10% and pain in 7% of patients. Less common presentations include inflammatory carcinoma with diffuse induration of the skin of the breast, and this confers an adverse prognosis. Around 40% of patients will have axillary nodal disease, with likelihood correlating with increasing size of the primary tumour. Distant metastases are infrequently present at diagnosis and the most common sites of spread are: bone (70%), lung (60%), liver (55%), pleura (40%), adrenals (35%), skin (30%) and brain (10–20%).

Investigations

Following clinical examination, patients should have imaging with mammography or ultrasound evaluation and a biopsy using fine needle aspiration for cytology or core biopsy for histology. Histological assessment should be carried out to assess tumour type and to determine oestrogen and progesterone receptor (ER/PR) status and HER2 status. If distant spread is suspected, a tumour biopsy for histology. Histological assessment should be carried out to assess tumour type and to determine oestrogen and progesterone receptor (ER/PR) status and HER2 status. If distant spread is suspected, CT of the thorax and abdomen and an isotope bone scan are required.

Management

Surgery is the mainstay of treatment for most patients, and this can range from a lumpectomy, where only the tumour is removed, to mastectomy, where the whole breast is removed. Lymph node sampling is performed at the time of surgery. Adjuvant radiotherapy is given to reduce the risk of local recurrence to 4–6%. Adjuvant hormonal therapy improves disease-free and overall survival in pre- and post-menopausal patients who have tumours that express ER. Patients at low risk, with tumours that are small and ER-positive, only require adjuvant hormonal therapy with tamoxifen. Patients with tumours that are ER-positive and who are pre-menopausal should receive an LHRH analogue. Aromatase inhibitors also have benefit in this setting and are still under investigation.

Adjuvant chemotherapy is considered for patients at higher risk of recurrence. Factors that increase the risk of recurrence include a tumour of greater than 1 cm, tumour that is ER-negative or the presence of involved axillary lymph nodes. Such patients should be offered adjuvant chemotherapy, which improves disease-free and overall survival (Box 11.19). The role of adjuvant treatment has been studied by meta-analyses and data support the use of adjuvant trastuzumab, a humanised monoclonal antibody to HER2, in addition to standard chemotherapy for women with early HER2-positive breast cancer.

Metastatic disease management includes radiotherapy to palliate painful bone metastases and second-line endocrine therapy with aromatase inhibitors, which inhibit peripheral oestrogen production in adrenal and adipose tissues. Advanced ER-negative disease may be treated with combination chemotherapy.

Ovarian cancer

Ovarian cancer is the most common gynaecological tumour in Western countries. Most ovarian cancers are epithelial in origin (90%), and up to 7% of women with ovarian cancer have a positive family history. Patients often present late in ovarian cancer with vague abdominal discomfort, low back pain, bloating, altered bowel habit and weight loss. Occasionally, peritoneal deposits are palpable as an omental ‘cake’ and nodules in the umbilicus (Sister Mary Joseph nodules).

Pathogenesis

Genetic and environmental factors play a role. The risk of ovarian cancer is increased in patients with BRCA1 or BRCA2 mutations, and Lynch type II families (a subtype of hereditary non-polyposis colon cancer (HNPPC)) have ovarian, endometrial, colorectal and gastric tumours due to mutations of mismatch repair enzymes. Advanced age, nulliparity, ovarian stimulation and Caucasian descent all increase the risk of ovarian cancer, whilst suppressed ovulation appears to protect, so pregnancy, prolonged breastfeeding and the contraceptive pill have all been shown to reduce the risk of ovarian cancer.

Investigations

Initial workup for patients with suspected ovarian malignancy includes imaging in the form of ultrasound and CT. Serum levels of the tumour marker CA-125 are often measured. Surgery plays a key role in the diagnosis, staging and treatment of ovarian cancer, and in early cases, palpation of viscera, intraoperative washing and biopsies are generally performed to define disease extent.

Management

In early disease, surgery followed by adjuvant chemotherapy with carboplatin, or carboplatin plus paclitaxel, is the treatment of choice. Surgery should include removal of the tumour along with total hysterectomy, bilateral salpingo-oophorectomy, and omentectomy. Even in advanced disease, surgery is undertaken to debulk the tumour and is followed by adjuvant chemotherapy, typically using carboplatin and paclitaxel. Bevacizumab is indicated for patients with high-grade tumours that are suboptimally debulked or those with a more aggressive biological pattern. Monitoring for relapse is achieved through a combination of serum CA-125 and clinical examination with CT imaging for those with suspected relapse. Second-line chemotherapy is aimed at improving symptoms and should not be used for CA-125 elevation only in the absence of symptoms. Treatments can include further platinum/
paclitaxel combination, liposomal doxorubicin or topotecan. These regimens are associated with a response rate of 10–40%. The best responses are observed in patients with a treatment-free interval of more than 12 months.

**Endometrial cancer**

Endometrial cancer accounts for 4% of all female malignancies, producing a 1 in 73 lifetime risk. The majority of patients are post-menopausal, with a peak incidence at 50–60 years of age. Mortality from endometrial cancer is currently falling. The most common presentation is with post-menopausal bleeding, which often results in detection of the disease before distant spread has occurred.

**Pathogenesis**

Oestrogen plays an important role in the pathogenesis of endometrial cancer, and factors that increase the duration of oestrogen exposure, such as nulliparity, early menarche, late menopause and unopposed HRT, increase the risk. Endometrial cancer is 10 times more common in obese women, and this is thought to be due to elevated levels of oestrogens.

**Investigations**

The diagnosis is confirmed by endometrial biopsy.

**Management**

Surgery is the treatment of choice and is also used for staging. A hysterectomy and bilateral salpingo-oophorectomy are performed with peritoneal cytology and, in some cases, lymph node dissection. Where the tumour extends beyond the inner 50% of the myometrium, involves the cervix and local lymph nodes, or there is lymphovascular space invasion, adjuvant pelvic radiotherapy is recommended. Chemotherapy and hormonal therapies have not demonstrated a sufficient survival advantage to be recommended for routine use in the adjuvant setting but have a role in recurrent disease.

**Cervical cancer**

This is the second most common gynaecological tumour worldwide. The incidence is decreasing in developed countries but continues to rise in developing nations. Cervical cancer is the leading cause of death from gynaecological cancer. The most common presentation is with an abnormal smear test, but with locally advanced disease the presentation is with vaginal bleeding, discomfort, discharge or symptoms attributable to involvement of adjacent structures, such as bladder, or rectal or pelvic wall. Occasionally, patients present with distant metastases to bone and lung.

**Pathogenesis**

There is a strong association between cervical cancer and sexual activity that includes sex at a young age and multiple sexual partners. Infection with HPV has an important causal role, and this has underpinned the introduction of programmes to immunise teenagers against HPV in an effort to prevent the later development of cervical cancer (p. 425).

**Investigations**

Diagnosis is made by smear or cone biopsy. Dilatation and curettage is also used diagnostically, with cystoscopy and flexible sigmoidoscopy if there are symptoms referable to the bladder, colon or rectum. In contrast to other gynaecological malignancies, cervical cancer is a clinically staged disease. MRI is often used to characterise the primary tumour. A routine chest X-ray should be obtained to help rule out pulmonary metastasis. CT of the abdomen and pelvis is performed to look for metastasis in the liver and lymph nodes and to exclude hydronephrosis and hydroureter.

**Management**

This depends on the stage of disease. Pre-malignant disease can be treated with laser ablation or diathermy, whereas in microinvasive disease a large loop excision of the transformation zone (LLETZ) or a simple hysterectomy is employed. Invasive but localised disease requires radical surgery, while chemotherapy and radiotherapy, including brachytherapy, may be given as primary treatment, especially in patients with adverse prognostic features such as bulky or locally advanced disease, or lymph node or parametrium invasion. In metastatic disease, cisplatin-based chemotherapy may be beneficial in improving symptoms but does not improve survival significantly.

**Head and neck tumours**

Head and neck cancers are typically squamous tumours that arise in the nasopharynx, hypopharynx and larynx. They are most common in elderly males, but now occur with increasing frequency in a younger cohort, as well as in women, especially where oropharyngeal cancers are concerned. The rising incidence of oropharyngeal cancers, especially in the developed world, is thought to be secondary to HPV infection. Presentation depends on the location of the primary tumour and the extent of disease. For example, early laryngeal cancers may present with hoarseness, while more extensive local disease may present with pain due to invasion of local structures or with a lump in the neck. Patients who present late often have pulmonary symptoms, as this is the most common site of distant metastases (Box 11.20).

**Pathogenesis**

The tumours are strongly associated with a history of smoking and excess alcohol intake, but other recognised risk factors include Epstein–Barr virus for nasopharyngeal cancer and HPV infection for oropharyngeal tumours.

**Investigations**

Careful inspection of the primary site is required as part of the staging process, and most patients will require endoscopic evaluation and examination under anaesthesia. Tissue biopsies should be taken from the most accessible site. CT of the primary site and the thorax is the
Investigations

In this situation, there is a temptation to investigate the patient endlessly in order to determine the original primary site. However, there is a compromise between exhaustive investigation and obtaining sufficient information to plan appropriate management. For all patients, histological examination of an accessible site of metastasis is required. The architecture of the tissue can assist the pathologist in determining the likely primary site, and therefore it is better to perform a biopsy rather than fine needle aspiration. The greater volume of tissue also permits the use of immunohistochemistry. Extensive imaging to search for the primary is rarely indicated; a careful history to identify symptoms and risk factors (including familial) will often permit a judicious choice of imaging.

Management

Management of the patient will depend on that person’s circumstances, as well as on the site(s) involved and the likely primary sites. The overriding principle is to ensure that a curable diagnosis has not been overlooked. For example, lung metastases from a testicular teratoma do not preclude cure; nor do one or two liver metastases from a colorectal cancer. Early discussion with an oncologist within a multidisciplinary team is essential and avoids unnecessary investigation; for example, a single hCG-based pregnancy test in a young man with lung metastases might confirm the presence of a teratoma and allow rapid administration of potentially curative chemotherapy. Treatment should not necessarily wait for a definitive diagnosis; appropriate analgesia, radiotherapy and surgical palliation can all be helpful. Some patients remain free of cancer for some years after resection of a single metastasis of an adenocarcinoma of unknown primary, justifying this approach in selected patients.

In those with no obvious primary, systemic chemotherapy may achieve some reduction in tumour burden and alleviation of symptoms, but long-term survival is rare.

Further information

Books and journal articles


Websites

http://info.cancerresearchuk.org/cancerstats/ A wide range of cancer statistics that can be sorted by type or geographical location.

Carcinoma of unknown origin

Some patients are found to have evidence of metastatic disease at their initial presentation prior to diagnosis of a primary site. In many cases, a subsequent biopsy reveals adenocarcinoma but the primary site is not always clear.

<table>
<thead>
<tr>
<th>Location</th>
<th>Common presenting features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypopharynx</td>
<td>Dysphagia, Odynophagia, Referred otalgia, Enlarged lymph nodes</td>
</tr>
<tr>
<td>Mouth</td>
<td>Non-healing ulcers, Ipsilateral otalgia</td>
</tr>
<tr>
<td>Nasal cavity and sinuses</td>
<td>Discharge (bloody) or obstruction</td>
</tr>
<tr>
<td>Nasopharynx</td>
<td>Nasal discharge or obstruction, Conduction deafness, Atypical facial pain, Diplopia, Hoarse voice, Horner’s syndrome</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>Dysphagia, Pain, Otalgia</td>
</tr>
<tr>
<td>Salivary gland</td>
<td>Painless swelling, Facial nerve palsy</td>
</tr>
</tbody>
</table>
Principles of palliative care 284
Presenting problems in palliative care 284
- Pain 284
- Breathlessness 289
- Cough 289
- Nausea and vomiting 289
- Gastrointestinal obstruction 290
- Weight loss and general weakness 290
- Anxiety and depression 290
- Delirium and terminal agitation 290

Death and dying 290
- Talking about and planning towards dying 290
- Diagnosing dying 291
- Management 291
- Ethical issues at the end of life 291
Palliative care is the active total care of patients with far advanced, rapidly progressive and ultimately fatal disease. Its focus is quality of life rather than cure, and it encompasses a distinct body of knowledge and skills that all good physicians must possess to allow them to care effectively for patients at the end of life. In palliative care, there is a fundamental change of emphasis in decision-making away from a focus on prolonging life towards decisions that balance comfort and the individual’s wishes with treatments that might prolong life. There is a growing recognition that the principles of, and some specific interventions developed in the palliative care of patients with cancer are equally applicable to other conditions. The principles of palliative care may therefore be applied not only to cancer but also to any chronic disease state.

Palliative care is often seen as a means of managing distress and symptoms in patients with cancer, where metastatic disease has been diagnosed and death is seen as inevitable. In other illnesses, the challenge is recognising when patients have entered this phase of their illness, as there are fewer clear markers and the course of the illness is much more variable.

Different chronic disease states progress at different rates, allowing some general trajectories of illness or dying to be defined (Fig. 12.1). These trajectories are useful in helping decision-making in individual patients and also in planning services.

Traditionally, palliative care has been associated with cancer because the latter is typified by a progressive decline in function which was more predictable than in many other diseases. This ‘rapid decline’ trajectory is the best-recognised pattern need for palliative care and many traditional hospice services are designed to meet the needs of people on this trajectory: for example, motor neuron disease, or AIDS where antiretroviral therapy (ART) is not available. With the improvements in management of malignant disease, this is no longer so true for all patients with cancer, whose illness may follow an erratic or intermittent decline trajectory.

Many chronic diseases, such as advanced chronic obstructive pulmonary disease (COPD) and intractable congestive heart failure, carry as high a burden of symptoms as cancer, as well as psychological and family distress. The ‘palliative phase’ of these illnesses may be more difficult to identify because of periods of relative stability interspersed with acute episodes of severe illness. However, it is still possible to recognise those patients whose care may benefit from a palliative approach. The challenge is that symptom management needs to be delivered at the same time as treatment for acute exacerbations. This leads to difficult decisions as to the balance between symptom relief and aggressive management of the underlying disease. The starting point of need for palliative care in these conditions is the point at which consideration of comfort and individual values becomes important in decision-making, often alongside management of the underlying disease.

The third major trajectory is categorised by years of poor function and frailty before a relatively short terminal period; it is exemplified by dementia, but is also increasingly true for patients with many different chronic illnesses. As medical advances extend survival, this mode of dying is being experienced by increasing numbers of people. The main challenge lies in providing nursing care and ensuring that plans are agreed for the time when medical intervention is no longer beneficial.

In a situation where death is inevitable and foreseeable, palliative care balances the ‘standard textbook’ approach with the wishes and values of the patient and a realistic assessment of the benefits of medical interventions. This often results in a greater focus on comfort, symptom control and support for patient and family, and may enable withdrawal of interventions that are ineffective or burdensome. Commonly, the outcome is less certain. In many cases, there is a substantial risk that the patient will die but there may be a small chance of improvement with further treatment. In these circumstances, it is often (but not always) correct and helpful to share this information with the patient so that better decisions can be made about further care.

The principles of palliative care are being used increasingly in many different diseases so that death can be managed effectively and compassionately. Palliative management of the most common symptoms is discussed in the next section.

### PRINCIPLES OF PALLIATIVE CARE

#### Pain

The International Association for the Study of Pain (IASP) has defined pain as ‘an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage’. It follows that each patient’s experience and expression of pain are different, and that severity of pain does not correlate with the degree of tissue damage. Effective pain treatment facilitates recovery from injury or surgery, aids rapid recovery of function, and may minimise chronic pain and disability. Unfortunately, the delivery of effective pain relief is often impeded by factors such as poor assessment and concerns about the use of opioid analgesia.

Fig. 12.1 Archetypal trajectories of dying. From Murray, et al. 2005 – see p. 292.
Pain classification and mechanisms

Pain can be classified into two types:

- **Nociceptive**: due to direct stimulation of peripheral nerve endings by a noxious stimulus such as trauma, burns or ischaemia.
- **Neuropathic**: due to dysfunction of the pain perception system within the peripheral or central nervous system as a result of injury, disease or surgical damage, such as continuing pain experienced from a limb which has been amputated (phantom limb pain). This should be identified early (Box 12.1) because it is more difficult to treat once established.

The pain perception system (p. 1147) is not a simple hard-wired circuit of nerves connecting tissue pain receptors to the brain, but a dynamic system in which a continuing pain stimulus can cause central changes that lead to an increase in pain perception. This plasticity (changeability) applies to all the peripheral and central components of the pain pathway. Early and appropriate treatment of pain reduces the potential for chronic undesirable changes to develop.

Assessment and measurement of pain

Accurate assessment of the patient is the first step in providing good analgesia.

**History and measurement of pain**

A full pain history should be taken, to establish its causes and the underlying diagnoses. Patients may have more than one pain; for example, bone and neuropathic pain may both arise from skeletal metastases (Box 12.2).

A diagram of the body on which the patient can mark the pain site can be helpful. When asked to score pain, patients consistently rate it higher than health professionals and should, if able, always be asked to rate pain themselves. Methods include:

- **Verbal rating scale**: Different verbal descriptions are used to rate pain - ‘no pain’, ‘mild pain’, ‘moderate pain’ and ‘severe pain’.
- **Visual analogue scale**: A question is used, such as ‘Over the past 24 hours, how would you rate your pain, if 0 is no pain and 10 is the worst pain you could imagine?’
- **Behavioural rating scale**: It can be particularly difficult to decide whether a patient with cognitive impairment is suffering pain. A variety of measures are available which use observed behaviours, such as agitation and withdrawn posture, to assess levels of pain. Commonly used scales include Abbey and Dolorplus. Changes in behavioural rating pain scores can indicate whether drug measures have been successful.

Regular recording of formal pain assessment and patient-rated pain scores improves pain management and reduces the time taken to achieve pain control.

**Psychological aspects of chronic pain**

Perception of pain is influenced by many factors other than the painful stimulus, and pain cannot therefore be easily classified as wholly physical or psychogenic in any individual (Fig. 12.2). Patients who suffer chronic pain will be affected emotionally and, conversely, emotional distress can exacerbate physical pain (p. 240). Full assessment for symptoms of anxiety and depression is essential to effective pain management.

**Examination**

This should include careful assessment of the painful area, looking for signs of neuropathic pain (see Box 12.1) or bony tenderness suggestive of bone metastases. In
Nearly all types of pain respond to morphine to some degree. Some are completely opioid-responsive but others, such as neuropathic and ischaemic pain, are relatively unresponsive. Opioid-unresponsive or poorly responsive pain will only be relieved by opioids at a dose which causes significant side-effects. In these situations, effective pain relief may only be achieved with the use of adjuvant analgesics (see below).

**Pharmacological treatments**

**Non-opioids**
- **Paracetamol.** This is often effective for mild to moderate pain. For severe pain, it is inadequate alone, but is a useful and well-tolerated adjunct.

**Weak opioids** Codeine and dihydrocodeine are weak opioids. They have lower analgesic efficacy than strong opioids and a ceiling dose. They are effective for mild to moderate pain.

**Strong opioids**
Immediate-release (IR) oral morphine takes about 20 minutes to have an effect and usually provides pain relief for 4 hours. Most patients with continuous pain should be prescribed IR oral morphine every 4 hours initially, as this will provide continuous pain relief over the whole 24-hour period. Controlled-release (CR) morphine lasts for 12 or 24 hours but takes longer to provide analgesia.

**The WHO analgesic ladder**
The basic principle of the WHO ladder (Fig. 12.3) is that analgesia which is appropriate for the degree of pain should be prescribed. If pain is severe or remains poorly controlled, strong opioids should be prescribed and increased as indicated by the patient’s need for additional analgesia (opioid titration).

A patient with mild pain is started on a non-opioid analgesic drug, such as paracetamol 1 g 4 times daily (step 1). If the maximum recommended dose is not
sufficient or the patient has moderate pain, a weak opioid, such as codeine 60 mg 4 times daily, should be added (step 2). If adequate pain relief is still not achieved with the maximum recommended dosages or if the patient has severe pain, a strong opioid is substituted for the weak opioid (step 3). It is important not to move ‘sideways’ (change from one drug to another of equal potency) on a particular step of the ladder. All patients with severe pain should receive a full trial of strong opioids with appropriate adjuvant analgesia, as described below.

In addition to the regular dose, an extra dose of IR morphine should be prescribed ‘as required’ for when the patient has pain that is not controlled by the regular prescription (breakthrough pain). This should be one-sixth of the total 24-hour dose of opiate. The frequency of breakthrough doses should be dictated by their efficacy and any side-effects, rather than by a fixed time interval. A patient may require breakthrough analgesia as frequently as hourly if pain is severe, but this should lead to early review of the regular prescription. The patient and/or carer should note the timing of any breakthrough doses and the reason for them. These should be reviewed daily and the regular 4-hourly dose increased for the next 24 hours on the basis of:

- frequency of and reasons for breakthrough analgesia
- degree and acceptability of side-effects.

The regular dose should be increased by adding the total of the breakthrough doses over the previous 24 hours, unless there are significant problems with unacceptable side-effects. When the correct dose has been established, a CR preparation can be prescribed, usually twice daily.

Worldwide, the most effective and appropriate route of administration is oral, though transdermal preparations of strong opioids (usually fentanyl) are useful in certain situations, such as in patients with dysphagia or those who are reluctant to take tablets on a regular basis. Diamorphine is a highly soluble strong opioid used for subcutaneous infusions, particularly in the last few days of life, but is only available in certain countries.

Common side-effects of opioids are shown in Box 12.4. Nausea and vomiting occur initially but usually settle after a few days. Confusion and drowsiness are dose-related and reversible. In acute dosing, respiratory depression can occur but this is rare in those on regular opioids.

**Opioid toxicity**

All patients will develop dose-related side-effects, such as nausea, drowsiness, confusion or myoclonus; this is termed opiate toxicity. The dose at which this occurs varies from 10 to 5000 mg of morphine, depending on the patient and the type of pain. When opiates are being titrated, side-effects should be assessed regularly. The earliest side-effects are often visual hallucinations (often a sense of movement at the periphery of vision) and a distinct myoclonic movement. Pain should be re-assessed to ensure that appropriate adjuvants are being used. Parenteral rehydration is often helpful to speed up excretion of active metabolites of morphine. The dose of opioid may need to be reduced or changed to an alternative strong opiate.

Different opioids have different side-effect profiles in different people. If a patient develops side-effects, switching to an alternative strong opioid may be helpful. Options include oxycodone, transdermal fentanyl, alfentanil, hydromorphone and occasionally methadone, any of which may produce a better balance of benefit against side-effects. Fentanyl and alfentanil have no renally excreted active metabolites and may be particularly useful in patients with renal failure. Pethidine is used in acute pain management but not for chronic or cancer pain because of its short half-life and ceiling dose.

It is important to be very careful when switching opiates, as it is easy to make calculation mistakes and prescribe too much or too little.
PALLIATIVE CARE AND PAIN

Adjuvant analgesics
An adjuvant analgesic is a drug with a primary indication other than pain but which provides analgesia in some painful conditions and may enhance the effect of the primary analgesic. At each step of the WHO analgesic ladder, an adjuvant analgesic should be considered, the choice depending on the type of pain (Boxes 12.5 and 12.6).

Non-pharmacological and complementary treatments
Radiotherapy
Radiotherapy can improve pain from bone metastases and may be considered for cancer in other sites (see Box 12.2).

Physiotherapy
This helps to alleviate pain and restore function, through active mobilisation and specific physiotherapy techniques, such as spinal manipulation, massage, application of heat or cold, and exercise. Immediate application of cold with ice packs can reduce subsequent swelling and inflammation after a direct injury.

Psychological techniques
These include simple relaxation, hypnosis, cognitive behavioural therapies and biofeedback (pp. 240–241), which train the patient to use coping strategies and behavioural techniques. This is often more relevant in chronic non-malignant pain than in cancer pain.

EBM 12.5 Treatment of neuropathic pain
‘Tricyclic antidepressants, a variety of anticonvulsants, and gabapentin are effective treatments for neuropathic pain.’

For further information: www.sign.ac.uk

Drug Example Indications Side-effects*

<table>
<thead>
<tr>
<th>NSAIDs</th>
<th>Diclofenac</th>
<th>Bone metastases, soft tissue infiltration, liver pain, inflammatory pain</th>
<th>Gastric irritation and bleeding, fluid retention, headache; caution in renal impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corticosteroids</td>
<td>Dexamethasone 8–16 mg per day, titrated to lowest dose that controls pain</td>
<td>Raised intracranial pressure, nerve compression, soft tissue infiltration, liver pain</td>
<td>Gastric irritation if used together with NSAID, fluid retention, confusion, Cushingoid appearance, candidiasis, hyperglycaemia</td>
</tr>
<tr>
<td>Anticonvulsants</td>
<td>Evidence strongest for: Gabapentin Pregabalin Duloxetine</td>
<td>Neuropathic pain of any aetiology</td>
<td>Mild sedation, tremor, confusion</td>
</tr>
<tr>
<td>Tricyclic antidepressants</td>
<td>Amitriptyline</td>
<td>Neuropathic pain of any aetiology</td>
<td>Sedation, dizziness, confusion, dry mouth, constipation, urinary retention; avoid in cardiac disease</td>
</tr>
<tr>
<td>NMDA blockers</td>
<td>Ketamine</td>
<td>Severe neuropathic pain (only under specialist supervision)</td>
<td>Confusion, anxiety, agitation, hypertension</td>
</tr>
</tbody>
</table>

*In old age, all drugs can cause confusion.
(NMDA = N-methyl-D-aspartate)

Stimulation therapies
Acupuncture (Fig. 12.4) has been used successfully in Eastern medicine for centuries. It causes release of endogenous analgesics (endorphins) within the spinal cord. It can be particularly effective in pain related to muscle spasm. Transcutaneous electrical nerve stimulation (TENS) may have a similar mechanism of action to acupuncture and can be used in both acute and chronic pain.

Herbal medicine and homeopathy
These are widely used for pain, but often with little evidence for efficacy (p. 15). Safety regulations for these treatments are limited, compared with conventional drugs, and the doctor should be wary of unrecognised side-effects which may result.

Fig. 12.4 Acupuncture.
**Presenting problems in palliative care**

**Cough**

Cough can be a troubling symptom in cancer and other illnesses such as motor neuron disease, cardiac failure and COPD. There are many possible causes (p. 654). Management should focus on treating the underlying condition if possible. If this fails to bring about the desired response, antitussives, such as codeine linctus, are sometimes effective, particularly for cough at night.

**Nausea and vomiting**

The presentation of nausea and vomiting differs, depending on the underlying cause, of which there are many (p. 853). Large-volume vomiting with little nausea is common in intestinal obstruction, whereas constant nausea with little or no vomiting is often due to metabolic abnormalities or drugs. Vomiting related to raised intracranial pressure is worse in the morning.

Different receptors are activated, depending on the cause or causes of the nausea (Fig. 12.5). For example, dopamine receptors in the chemotactic trigger zone in the fourth ventricle are stimulated by metabolic and drug causes of nausea, whereas gastric irritation stimulates histamine receptors in the vomiting centre via the vagus nerve.

---

**Breathlessness**

The sensation of breathlessness is the result of a complex interaction between different factors at the levels of production (the pathophysiological cause), perception (the severity of breathlessness perceived by the patient) and expression (the symptoms expressed by an individual patient). A patient’s perception and expression of breathlessness can be significantly improved, even if there is no reversible ‘cause’ (Box 12.7). Assessment and treatment should therefore be targeted at modifying these factors, particularly when there is no reversible pathophysiology.

Clearly reversible causes of breathlessness (p. 655) should be identified and managed, but investigation and treatment should be appropriate to the prognosis and stage of disease. A therapeutic trial of corticosteroids (dexamethasone 8 mg for 5 days) and/or nebulised salbutamol may be helpful.

Breathlessness may be worsened by specific anxieties and beliefs; these should be explored. Many people with heart failure are concerned that exertional breathlessness will lead to worsening of their heart condition. Patients with advanced disease have specific panic-breathlessness cycles in which breathlessness leads to panic, which leads to worsening breathlessness and worsening panic. These should be identified and explained to the patient. Many fear that they will die during one of these episodes, and explanation of the panic cycle can be very reassuring. Another frequently expressed fear is that breathlessness will continue to worsen until it is continuous and unbearable, leading to a distressing and undignified death. Reassurance should again be given that this is uncommon and can be effectively managed with opioids and benzodiazepines.

A rapidly acting benzodiazepine, such as sublingual lorazepam, or non-drug measures, such as relaxation techniques, may help panic-breathlessness cycles. Attention to energy conservation (thinking clearly about using limited energy reserves sensibly) and pacing of activity is also extremely helpful. Physiotherapists are good at this and should be involved in developing an individual plan for each patient.

Perception of breathlessness may also be improved by night-time or regular morphine, or by regular benzodiazepines.

Oxygen does not help breathlessness unless the patient is hypoxic. An electric fan, piped air or an open window can be as effective as oxygen in patients who are breathless but not hypoxic. The patient’s, family’s or even professional beliefs about the benefits and need for oxygen may be the main reason for its apparent efficacy in non-hypoxic patients who feel less breathless when using oxygen.

**EBM 12.7 Palliative treatment of breathlessness**

‘Interventions based on psychosocial support, breathing control and coping strategies can help patients to cope with the symptom of breathlessness and reduce physical and emotional distress.’


---

**Fig. 12.5 Mechanisms of nausea.** (ACh = acetylcholine; D₂ = dopamine; 5-HT = 5-hydroxytryptamine, serotonin; H₁ = histamine)
Reversible causes, such as hypercalcaemia and constipation, should be treated appropriately. Drug-induced causes should be considered and the offending drugs stopped if possible. As different classes of antiemetic drug act at different receptors, antiemetic therapy should be based on a careful assessment of the probable causes and a rational decision to use a particular class of drug (Box 12.8). The subcutaneous route is often required initially to overcome gastric stasis and poor absorption of oral medicines.

**Gastrointestinal obstruction**

Gastrointestinal obstruction is a frequent complication of intra-abdominal cancer. Patients may have multiple levels of obstruction and symptoms may vary greatly in nature and severity. Surgical mortality is high in patients with advanced disease and obstruction should normally be managed without surgery.

The key to effective management is to address the presenting symptoms – colic, abdominal pain, nausea, vomiting, intestinal secretions – individually or in combination, using drugs which do not cause or worsen other symptoms. This can be problematic when a specific treatment worsens another symptom. Cyclizine improves nausea and colic responds well to anticholinergic agents, such as hyoscine butylbromide, but both slow gut motility. Nausea will improve with metoclopramide, although this is contraindicated in the presence of colic because of its prokinetic effect. There is some evidence that corticosteroids (dexamethasone 8 mg) can shorten the length of obstructive episodes. Somatostatin analogues, such as octreotide, will reduce intestinal secretions and therefore large-volume vomits. Occasionally, a nasogastric tube is required to reduce gaseous or fluid distension.

**Weight loss and general weakness**

Patients with cancer lose weight due to an alteration of metabolism by the tumour known as the cancer cachexia syndrome. NSAIDs and megestrol may be helpful in early-stage disease but are unlikely to be effective in advanced cancer. Corticosteroids can temporarily boost appetite and general well-being, but may cause false weight gain by promoting fluid retention. Their benefits need to be weighed against the risk of side-effects.

**Anxiety and depression**

Depression is common in palliative care but diagnosis is more difficult, as the physical symptoms of depression are similar to those of advanced disease. Anxiety and depression may still respond to treatment with a combination of drugs and psychotherapeutic approaches (p. 243). Citalopram and mirtazapine are better tolerated in patients with advanced disease. It should not be assumed that depression is an ‘understandable’ consequence of the patient’s situation.

**Delirium and terminal agitation**

Many patients become confused or agitated in the last days of life. It is important to identify and treat potentially reversible causes (pp. 173 and 280), unless the patient is too close to death for this to be feasible. Early diagnosis and effective management of delirium are extremely important. As in other palliative situations, it may not be possible to identify and treat the underlying cause, and the focus of management may be to ensure comfort. It is important to distinguish between behavioural change due to pain and that due to delirium, as opioids will improve one and worsen the other.

The management of delirium is detailed on page 174. It is important, even in palliative care to treat delirium with antipsychotic medicines such as haloperidol rather than regard it as distress or anxiety and use benzodiazepines only.

**DEATH AND DYING**

**Talking about and planning towards dying**

There have been dramatic improvements in medical treatment and care of patients with cancer and other illnesses over recent years, but the inescapable fact remains that everyone will die at some time. Planning for death is not required for people who die suddenly but should be actively considered in patients with chronic diseases when the death is considered to be foreseeable or inevitable. Doctors rarely know exactly when a patient will die but we are often aware that the risk of dying is increasing and that medical interventions are unlikely to prolong life or improve function. Many people wish their doctors to be honest about this situation to allow them time to think ahead, make plans and address practical issues. A smaller number do not wish to discuss future deterioration or death; this avoidance of discussion should be respected.

For doctors, it is helpful to understand an individual’s wishes and values about medical interventions at this time, as this can help guide decisions about ceilings of intervention. Some interventions will not work in patients with far advanced disease. It is useful to distinguish between those that will not work (a medical decision) and those that do not confer sufficient benefit to be worthwhile (a decision that can only be reached with a patient’s involvement and consent). A common example...
of this would be decisions about not attempting cardio-pulmonary resuscitation.

In general, people wish for a dignified and peaceful death and most prefer to die at home. Families also are grateful for the chance to prepare themselves for the death of a relative, by timely and gentle discussion with their doctor or other health professionals. Early discussion and effective planning improve the chances that an individual’s wishes will be achieved.

Diagnosing dying

When patients with cancer become bed-bound, semi-comatose, unable to take tablets and only able to take sips of water, with no reversible cause, they are likely to be dying and many will have died within 2 days. Patients with other conditions also reach a stage where death is predictable and imminent. Doctors are sometimes poor at recognising this, and should be alert to the views of other members of the multidisciplinary team. A clear decision that the patient is dying should be agreed and recorded.

Management

Once the conclusion has been reached that a patient is going to die in the next few days, there is a significant shift in management (Box 12.9). Symptom control, relief of distress and care for the family become the most important elements of care. Medication and investigation are only justifiable if they contribute to these ends. When patients can no longer drink because they are dying, intravenous fluids are usually not necessary and may cause worsening of bronchial secretions. Medicines should always be prescribed for the relief of symptoms. For example, morphine or diamorphine may be used to control pain, levomepromazine to control nausea, haloperidol to treat confusion, diazepam or midazolam to treat distress, and hyoscine hydrobromide to reduce respiratory secretions. Side-effects, such as drowsiness, may be acceptable if the principal aim of relieving distress is achieved. It is important to discuss and agree the aims of care with the patient’s family.

Ethical issues at the end of life

In Europe, between 25 and 50% of all deaths are associated with some form of decision which may affect the length of a patient’s life. The most common form of decision involves withdrawing or withholding further treatment: for example, not treating a chest infection in a patient who is clearly dying from advanced cancer. It is important to have a framework for considering such decisions (such as the four ethical principles: autonomy, beneficence, non-maleficence and justice, p. 10), which balances degrees of importance when there is conflict: for example, when a patient wishes to receive treatment which a doctor believes will be ineffective or which may cause harm. A decision has to be taken as to which principle is most important: whether it is better to respect a patient’s wishes, even if it causes harm, or to reduce the risk of harm but not adhere to those wishes.

A futile treatment is one which has no chance of achieving worthwhile benefit: that is, the treatment cannot achieve a result that the patient would consider, now or in the future, to be worthwhile. Doctors are not required to institute futile treatments, such as resuscitation, in the event of cardiac arrest in a patient with terminal cancer.

Incapacity and advance directives

Patients’ wishes are very important in Western medical ethics, although other cultures emphasise the views of the family. If a patient is unable to express his or her view because of communication or cognitive impairment, that person lacks ‘capacity’. In order to decide what the patient would have wished, as much information as possible should be gained about any previously expressed wishes, along with the views of relatives and other health professionals.

An advance directive is a previously recorded, written document of a patient’s wishes (p. 171). It should carry the same weight in decision-making as a patient’s contemporaneously expressed wishes, but may not be sufficiently specific to be used in a particular clinical situation. The legal framework for decision-making varies in different countries.

Hydration

Deciding whether to give intravenous fluids can be difficult when a patient is very unwell and the prognosis
PALLIATIVE CARE AND PAIN

is uncertain. If a patient is clearly dying and has a prognosis of a few days, rehydration may cause harm by increasing bronchial secretions, and will not benefit the patient by prolonging life. A patient with a major stroke, who is unable to swallow but expected to survive the event, will develop renal impairment and thirst if not given fluids and should be hydrated. Each decision should be individual and discussed with the patient’s family.

Euthanasia

In the UK and Europe, between 3 and 6% of dying patients ask a doctor to end their life. Many of these requests are transient; some are associated with poor control of physical symptoms or a depressive illness. All expressions of a wish to die are an opportunity to help the patient discuss and address unresolved issues and problems.

Reversible causes, such as pain or depression, should be treated. Sometimes, patients may choose to discontinue life-prolonging treatments, such as diuretics or anticoagulation, following discussion and the provision of adequate alternative symptom control. However, there remain a small number of patients who have a sustained, competent wish to end their lives, despite good control of physical symptoms. Euthanasia is now permitted or legal in some countries but remains illegal in many others; public ethical and legal debate over this issue is likely to continue.

Further information and acknowledgements

Websites

www.anaesthetist.com/ Information on pain physiology and acute management of pain.
www.helpthehospices.org.uk/clip/index.htm Useful online tutorials on all aspects of palliative symptom control.
www.palliativecareguidelines.scot.nhs.uk Regularly reviewed, evidence-based clinical guidelines.
www.palliativedrugs.co.uk Information for health professionals about the use of drugs in palliative care. It highlights drugs given for unlicensed indications or by unlicensed routes, and the administration of multiple drugs by continuous subcutaneous infusion.

Figure acknowledgements

Fig. 12.1 Reproduced from Murray SA, Kendall M, Boyd K, et al. Illness trajectories and palliative care. BMJ 2005; 330:7498; reproduced with permission from the BMJ Publishing Group.

Fig. 12.3 WHO. Cancer pain relief. 2nd edn. Geneva: WHO; 1996.
Clinical examination of patients with infectious disease 294

Presenting problems in infectious diseases 296
- Fever 296
- Positive blood culture 303
- Sepsis 304
- Acute diarrhoea and vomiting 306
- Infections acquired in the tropics 308
- Infections in adolescence 313
- Infections in pregnancy 313

Viral infections 314
- Systemic viral infections with exanthem 314
- Systemic viral infections without exanthem 319
- Viral infections of the skin 325
- Gastrointestinal viral infections 327
- Respiratory viral infections 328
- Viral infections with neurological involvement 328
- Viral infections with rheumatological involvement 329

Prior diseases 329

Bacterial infections 329
- Bacterial infections of the skin, soft tissues and bones 329
- Systemic bacterial infections 333
- Gastrointestinal bacterial infections 341
- Respiratory bacterial infections 345
- Bacterial infections with neurological involvement 347
- Mycobacterial infections 347
- Rickettsial and related intracellular bacterial infections 350
- Chlamydial infections 353

Protozoal infections 353
- Systemic protozoal infections 353
- Leishmaniasis 362
- Gastrointestinal protozoal infections 367

Infections caused by helminths 369
- Intestinal human nematodes 369
- Tissue-dwelling human nematodes 372
- Zoonotic nematodes 375
- Trematodes (flukes) 376
- Cestodes (tapeworms) 378

Ectoparasites 381

Fungal infections 381
- Superficial mycoses 381
- Subcutaneous mycoses 381
- Systemic mycoses 383
Observation

- Temperature
- Sweating
- Weight loss
- Respiratory distress
- Altered consciousness
- Pallor
- Jaundice

1. Skin
   - Generalised erythema
   - Rash (see opposite)
   - IV injection track marks
   - Surgical scars
   - Prosthetic devices, e.g. central venous catheters
   - Tattoos

2. Hands and nails
   - Finger clubbing
   - Splinter haemorrhages
   - Janeway lesions
   - Signs of chronic liver disease
   - Vasculitis lesions

3. Oropharynx
   - Dental caries
   - Tonsillar enlargement or exudate
   - Candidiasis

4. Head and neck
   - Lymphadenopathy
   - Parotidomegaly
   - Abnormal tympanic membranes

5. Eyes
   - Conjunctival petechiae
   - Painful red eye in uveitis
   - Loss of red reflex in endophthalmitis
   - Roth’s spots in infective endocarditis
   - Haemorrhages and exudates of cytomegalovirus retinitis
   - Choroidal lesions of tuberculosis

6. Neurological
   - Neck stiffness
   - Photophobia
   - Delirium
   - Focal neurological signs

7. Heart and lungs
   - Tachycardia, hypotension
   - Murmurs or prosthetic heart sounds
   - Pericardial rub
   - Signs of consolidation
   - Pleural or pericardial effusion

8. Abdomen
   - Hepatosplenomegaly
   - Ascites
   - Renal angle tenderness
   - Mass lesions
   - Surgical drains

9. Musculoskeletal
   - Joint swelling, erythema or tenderness
   - Localised tender spine suggestive of epidural abscesses or discitis
   - Draining sinus of chronic osteomyelitis

10. Genitalia and rectum
    - Ulceration or discharge
    - Testicular swelling or nodules
    - Inguinal lymphadenopathy
    - Prostatic tenderness
    - Rectal fluctuance

- Chest X-ray consolidation in pneumonia
- Testicular swelling in adult mumps
Fever

- Documentation of fever. ‘Feeling hot’ or sweaty does not necessarily signify fever. Fever is diagnosed only when a body temperature of over 38.0°C has been recorded. Axillary and aural measurement is less accurate than oral or rectal. Outpatients may be trained to keep a temperature chart.

- Rigs. Shivering (followed by excessive sweating) occurs with a rapid rise in body temperature from any cause.

- Night sweats. These are associated with particular infections (e.g. tuberculosis, infective endocarditis), but sweating from any cause is worse at night.

- Excessive sweating. Alcohol, anxiety, thyrotoxicosis, diabetes mellitus, acromegaly, lymphoma and excessive environmental heat all cause sweating without temperature elevation.

- Recurrent fever. There are various causes, e.g. Borrelia recurrentis, bacterial abscess.

- Accompanying features.

  - HEADACHE. Severe headache and photophobia, although characteristic of meningitis, may accompany other infections.
  - DELIRIUM. Mental confusion during fever is more common in young children or the elderly.
  - MUSCLE PAIN. Myalgia may occur with viral infections, such as influenza, and with septicaemia, including meningococcal sepsis.
  - SHOCK. Shock may accompany severe infections and sepsis (Ch. 8).

**History-taking in suspected infectious disease**

**Presenting complaint**
- Diverse manifestations of infectious disease make accurate assessment of features and duration critical; e.g. fever and cough lasting 2 days imply an acute respiratory tract infection but suggest TB if they last 2 months

**Review of systems**
- Must be comprehensive

**Past medical history**
- Define the ‘host’ and likelihood of infection(s)
- Include surgical and dental procedures involving prosthetic materials
- Document previous infections

**Medication history**
- Include non-prescription drugs, use of antimicrobials and immunosuppressive drugs
- Identify medicines which interact with antimicrobials

**Allergy history**
- Esp. to antimicrobials, noting allergic manifestation (e.g. rash versus anaphylaxis)

**Family and contact history**
- Note infections and their time course
- Sensitively explore exposure to key infections, e.g. TB and HIV-1

**Travel history**
- Include countries visited and where previously resident (relevant to exposure and likely vaccination history, e.g. likelihood of BCG vaccination in childhood)

**Occupation**
- E.g. Anthrax in leather tannery workers

**Recreational pursuits**
- E.g. Leptospirosis in canoeists and windsurfers

**Animal exposures**
- Include pets, e.g. dog exposure and hydatid disease

**Dietary history**
- Consider undercooked meats, shellfish, unpasteurised dairy products or well water
- Establish who else was exposed, e.g. to food-borne pathogens

**History of intravenous drug injection or receipt of blood products**
- Risks for blood-borne viruses, e.g. HIV-1, and HBV and HCV

**Sexual history**
- Explore in a confidential and non-threatening way (Ch. 19), remember that the most common mechanism of HIV-1 transmission is heterosexual contact (Ch. 14)

**Vaccination history and use of prophylactic medicines**
- Consider occupation- or age-related vaccines
- In a traveller or infection-predisposed patient, establish compliance with prophylactic treatments

*Always consider non-infectious aetologies in the differential diagnosis.

(Ch. 13, 14, 15; Box 14, 18, and 28; Ch. 19; Table 82; Ch. 130)

**Skin lesions in infectious diseases**

- Diffuse erythema, e.g. A
- Migrating erythema, e.g. B
- Enlarging rash of erythema migrans in Lyme disease (see Fig. 13.20, p. 335)
- Purpuric or petechial rashes, e.g. C
- Macular or papular rashes, e.g. primary infection with HIV (Box 14.8, p. 395)
- Vesicular or blistering rash, e.g. D
- Erythema multiforme (Fig. 28.48 and Box 28.41, p. 1302)
- Nodules or plaques, e.g. Kaposi’s sarcoma (p. 384)
- Erythema nodosum (D and Box 28.42, p. 1303)
### INFECTION

The principles of infection and its investigation and therapy are described in Chapter 6. This chapter and the following ones on human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) and sexually transmitted infection (STI) describe the approach to patients with potential infectious disease, the individual infections and the resulting syndromes.

### PRESENTING PROBLEMS IN INFECTION

Infectious diseases present with myriad clinical manifestations. Many of these are described either in other chapters of this book or below.

#### Fever

‘Fever’ implies an elevated core body temperature of more than 38.0°C (p. 138). Fever is a response to cytokines and acute phase proteins (pp. 74 and 82) and occurs in infections and in non-infectious conditions.

#### Clinical assessment

The differential diagnosis is very broad so clinical features are used to guide the most appropriate investigations. The systematic approach described on pages 294–295 should be followed. Box 13.1 describes the assessment of elderly patients.

#### Investigations

If the clinical features do not suggest a specific infection, then initial investigations should include:

- a full blood count (FBC) with differential, including eosinophil count
- urea and electrolytes, liver function tests (LFTs), blood glucose and muscle enzymes
- inflammatory markers, erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP)
- a test for antibodies to HIV-1 (p. 392)
- autoantibodies, including antinuclear antibodies (ANA)

#### 13.1 Fever in old age

<table>
<thead>
<tr>
<th><strong>Temperature measurement:</strong> fever may be missed because oral temperatures are unreliable. Rectal measurement may be needed but core temperature is increasingly measured using eardrum reflectance.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute confusion:</strong> common with fever, especially in those with underlying cerebrovascular disease or dementia.</td>
</tr>
<tr>
<td><strong>Prominent causes of pyrexia of unknown origin:</strong> include tuberculosis and intra-abdominal abscesses, complicated urinary tract infection and infective endocarditis. Non-infective causes include polymyalgia rheumatica/temporal arteritis and tumours. A smaller fraction of cases remain undiagnosed than in young people.</td>
</tr>
<tr>
<td><strong>Pitfalls in the elderly:</strong> conditions such as cerebrovascular accident or thromboembolic disease can cause fever but every effort must be made to exclude concomitant infection.</td>
</tr>
<tr>
<td><strong>Common infectious diseases in the very frail</strong> (e.g. nursing home residents): pneumonia, urinary infection, soft tissue infection and gastroenteritis.</td>
</tr>
</tbody>
</table>

- chest X-ray and electrocardiogram (ECG)
- urinalysis and urine culture
- blood culture (p. 140)
- throat swab for culture
- other specimens, as indicated by history and examination, e.g. wound swab; sputum culture; stool culture, microscopy for ova and parasites, and Clostridium difficile toxin assay; if relevant, malaria films on 3 consecutive days or a malaria rapid diagnostic test (antigen detection, p. 355).

Subsequent investigations in patients with HIV-related (p. 396), immune-deficient (p. 301), nosocomial or travel-related (p. 309) pyrexia and in individuals with associated symptoms or signs of involvement of the respiratory, gastrointestinal or neurological systems are described elsewhere.

#### Management

Fever and its associated systemic symptoms can be treated with paracetamol, and by tepid sponging to cool the skin. Replacement of salt and water is important in patients with drenching sweats. Further management is focused on the underlying cause.

#### Fever with localising symptoms or signs

In most patients, the site of infection is apparent after clinical evaluation (p. 294), and the likelihood of infection is reinforced by investigation results (e.g. neutrophilia with raised ESR and CRP in bacterial infections). Not all apparently localising symptoms are reliable; however; headache, breathlessness and diarrhoea can occur in sepsis without localised infection in the central nervous system (CNS), respiratory tract or gastrointestinal tract. Careful interpretation of the clinical features is vital (e.g. severe headache associated with photophobia, rash and neck stiffness suggests meningitis, whereas moderate headache with cough and rhinorrhea is consistent with a viral upper respiratory tract infection).

Common infections that present with fever are shown in Figure 13.1. Further investigation and management are specific to the cause, but may include empirical antimicrobial therapy (p. 149) pending confirmation of the microbiological diagnosis.

#### Pyrexia of unknown origin

Pyrexia of unknown origin (PUO) is defined as a temperature persistently above 38.0°C for more than 3 weeks, without diagnosis, despite initial investigation during 3 days of inpatient care or after more than two outpatient visits. Subsets of PUO are described by medical setting: HIV-1 related, immune-deficient or nosocomial. Up to one-third of cases of PUO remain undiagnosed.

#### Clinical assessment

Major causes of PUO are outlined in Box 13.2. Rare causes, such as periodic fever syndromes (p. 85), should be considered in those with a positive family history. Children and younger adults are more likely to have infectious causes – in particular, viral infections. Older adults are more likely to have certain infectious and non-infectious causes (see Box 13.1). Detailed history and examination should be repeated at regular intervals.
Fig. 13.1 Common infectious syndromes presenting with fever and localised features. Major causes are grouped by approximate anatomical location and include central nervous system infection; respiratory tract infections; abdominal, pelvic or urinary tract infections; and skin and soft tissue infections (SSTI) or osteomyelitis. For each site of infection, particular syndromes and their common causes are described elsewhere in the book. The causative organisms vary, depending on host factors, which include whether the patient has lived or resided in a tropical country or particular geographical location, has acquired the infection in a health-care environment or is immunocompromised.
INFECTIONOUS DISEASE

13.2 Aetiology of pyrexia of unknown origin (PUO)

Infections (~30%)

Specific locations
- Abscesses: hepatobiliary*, diverticular*, urinary tract* (including prostate), pulmonary, CNS
- Infections of oral cavity (including dental), head and neck (including sinuses)
- Bone and joint infections
- Infective endocarditis*

Specific organisms
- Tuberculosis (particularly extrapulmonary)*
- HIV-1 infection
- Other viral infections (cytomegalovirus (CMV), Epstein–Barr virus (EBV))
- Fungal infections (e.g. Aspergillus spp., Candida spp. or dimorphic fungi)
- Infections with fastidious organisms (e.g. Bartonella spp., Tropheryma whippelii)

Specific patient groups
- Imported infections
  - Malaria, dengue, rickettsial infections, Brucella spp., amoebic liver abscess, enteric fevers, Leishmania spp. (southern Europe, India, Africa and Latin America), Burkholderia pseudomallei (South-east Asia), HIV and respiratory tract infections
- Nosocomial infections
- Infections related to prosthetic materials and surgical procedures
- HIV-positive individuals
  - Acute retroviral syndrome
- AIDS-defining infections (disseminated Mycobacterium avium complex (DMAC), Pneumocystis jirovecii (carinii) pneumonia, CMV and others)

Malignancy (~20%)

Haematological malignancy
- Lymphoma*, leukaemia and myeloma

Solid tumours
- Renal, liver, colon, stomach, pancreas, kidney

Connective tissue disorders (~15%)

Older adults
- Temporal arteritis/polymyalgia rheumatica*

Younger adults
- Still’s disease (juvenile rheumatoid arthritis)*
- Systemic lupus erythematosus (SLE)
- Vasculitic disorders (including polyarteritis nodosa, rheumatoid disease with vasculitis and granulomatosis with polyangiitis (also known as Wegener’s granulomatosis))
- Polymyositis
- Behçet’s disease

Geographically restricted
- Rheumatic fever

Miscellaneous (~20%)

Cardiovascular
- Atrial myxoma, aortitis, aortic dissection

Respiratory
- Sarcoïdosis, pulmonary embolism and other thromboembolic disease, extrinsic allergic alveolitis

Gastrointestinal
- Inflammatory bowel disease, granulomatous hepatitis, alcoholic liver disease, pancreatitis

Endocrine/metabolic
- Thyrotoxicosis, thyroiditis, hypothalamic lesions, phaeochromocytoma, adrenal insufficiency, hypertriglyceridaemia

Haematological
- Haemolytic anaemia, paroxysmal nocturnal haemoglobinuria, thrombotic thrombocytopenic purpura, myeloproliferative disorders, Castleman’s disease, graft-versus-host disease (after allogeneic haematopoietic stem cell transplantation)

Inherited
- Familial Mediterranean fever and periodic fever syndromes

Drug reactions*
- e.g. Antibiotic fever, drug hypersensitivity reactions etc.

Factitious fever
- Idiopathic (~15%)

*Most common causes within each group.

to detect emerging features (e.g. rashes, signs of infective endocarditis (p. 625) or features of vasculitis. In men, the prostate should be considered as a potential source of infection.

Clinicians should be alert to the possibility of factitious fever, in which high temperature recordings are engineered by the patient (Box 13.3).

Investigations

If initial investigation of fever (see above) is negative, a series of further microbiological and non-microbiological investigations should be considered (Boxes 13.4 and 13.5). These will usually include:
- induced sputum or other specimens for mycobacterial stains and culture
- serological tests
- imaging of the abdomen by ultrasonography or computed tomography (CT)
- echocardiography.

Lesions identified on imaging should usually be biopsied in order to seek evidence of relevant pathogens by culture, histopathology or nucleic acid detection. The chance of a successful diagnosis is greatest if procedures for obtaining and transporting the correct samples in the appropriate media are carefully planned in advance;
13.4 Microbiological investigation of PUO

**Microscopy**
- Blood for atypical lymphocytes (EBV, CMV, HIV-1, hepatitis viruses or *Toxoplasma gondii*, trypanosomiasis, malaria, *Borrelia* spp.)
- Respiratory samples for mycobacteria, fungi
- Stool for ova, cysts and parasites
- Biopsy for light microscopy (bacteria, mycobacteria, fungi, *Leishmania* and other parasites) and/or electron microscopy (viruses, protozoa (e.g. *microsporidia*) and other fastidious organisms (e.g. *T. whipplei*)
- Urine for white or red blood cells, schistosome ova, mycobacteria (*early morning urine × 3*)

**Culture**
- Aspirates and biopsies (e.g. joint, deep abscess, debrided tissues)
- Blood, including prolonged culture and special media conditions
- Sputum for mycobacteria
- Cerebrospinal fluid (CSF)
- Gastric aspirate for mycobacteria
- Stool
- Swabs
- Urine + prostatic massage in older men

**Antigen detection**
- Blood, e.g. HIV p24 antigen, cryptococcal antigen, *Histoplasma* antigen (restricted availability) and *Aspergillus* galactomannan enzyme-linked immunosorbent assay (ELISA)
- CSF for cryptococcal antigen
- Bronchoalveolar lavage fluid for *Aspergillus* galactomannan
- Nasopharyngeal aspirate/throat swab for respiratory viruses
- Urine, e.g. for *Legionella* antigen

**Nucleic acid detection**
- Blood for *Bartonella* spp. and viruses
- CSF for viruses and key bacteria (meningococcus, pneumococcus)
- Nasopharyngeal aspirate/throat swab for respiratory viruses
- Bronchoalveolar lavage fluid, e.g. for respiratory viruses
- Tissue specimens, e.g. for *Tropheryma whipplei*
- Urine, e.g. for *Chlamydia trachomatis*, *Neisseria gonorrhoeae*
- Stool, e.g. for norovirus, rotavirus

**Immunological tests**
- Serology (antibody detection) for viruses, dimorphic fungi and some bacteria and protozoa
- Interferon-γ release assay for diagnosis of tuberculosis

**Note** This list does not apply to every patient with a PUO. Appropriate tests should be selected in a stepwise manner, according to specific predisposing factors, epidemiological exposures and local availability, and should be discussed with a microbiologist.

---

13.5 Additional investigations in PUO

**Serological tests for connective tissue disorders**
- Autoantibody screen
- Complement levels
- Immunoglobulins
- Cryoglobulins

**Echocardiography**

**Ultrasound of abdomen**
- CT/MRI of thorax, abdomen and/or brain

**Imaging of the skeletal system**
- Plain X-rays
- CT/MRI spine
- Isotope bone scan

**Labelled white cell scan**
- Positron emission tomography (PET)/single photon emission computed tomography (SPECT)

**Biopsy**
- Bronchoscopy and lavage ± transbronchial biopsy
- Lymph node aspirate or biopsy
- Biopsy of radiological lesion
- Biopsy of liver
- Bone marrow aspirate and biopsy
- Lumbar puncture
- Laparoscopy and biopsy
- Temporal artery biopsy

---

this requires discussion between the clinical team, the radiologist or surgeon performing the procedure, and the local microbiologist and histopathologist. Liver biopsy may be justified, e.g. to identify idiopathic granulomatous hepatitis, if there are biochemical or radiological abnormalities. Bone marrow biopsies have a diagnostic yield of up to 15%, most often revealing haemato logical malignancy, myelodysplasia or tuberculosis, and also identifying brucellosis, typhoid fever or visceral leishmaniasis. Bone marrow should be sent for culture, as well as microscopy. Laparoscopy is occasionally undertaken with biopsy of abnormal tissues. Splenic aspiration in specialist centres is the diagnostic test of choice for suspected visceral leishmaniasis. Temporal artery biopsy should be considered in patients over the age of 50 years, even in the absence of physical signs or a raised ESR. ‘Blind’ biopsy of other structures in the absence of localising signs, or laboratory or radiology results is unhelpful.

**Prognosis**

The overall mortality of PUO is 30–40%, mainly attributable to malignancy in older patients. If no cause is found, the long-term mortality is low and fever often settles spontaneously.

**Fever in the injection drug-user**

Intravenous injection of recreational drugs is widespread in many parts of the world (p. 240). Infective organisms are introduced by non-sterile (often shared) injection equipment (Fig. 13.2), and infection is facilitated by immunodeficiency due to malnutrition or the toxic effects of drugs. The risks increase with prolonged drug use and injection into large veins of the groin and neck because of progressive thrombosis of superficial peripheral veins. The most common causes of fever are soft tissue or respiratory infections.

**Clinical assessment**

The history should address the following risk factors:
- Site of injection. Femoral vein injection is associated with vascular complications such as deep venous thrombosis (50% of which are septic) and accidental arterial injection with false aneurysm formation or a compartment syndrome due to swelling within the fascial sheath. Local complications include iliopsoas abscess, and septic arthritis of the hip joint or sacroiliac joint. Injection of the jugular vein can be
associated with cerebrovascular complications. Subcutaneous and intramuscular injection has been related to infection by clostridial species, the spores of which contain heroin. *Clostridium novyi* causes a local lesion with significant toxin production, leading to shock and multi-organ failure. Tetanus, wound botulism and gas gangrene also occur.

- **Technical details of injection.** Sharing of needles and other injecting paraphernalia (including spoons and filters) greatly increases the risk of blood-borne virus infection (e.g. HIV-1, hepatitis B or C virus). Some users lubricate their needles by licking them prior to injection, thus introducing mouth organisms (e.g. anaerobic streptococci, *Fusobacterium* spp. and *Prevotella* spp.). Contamination of commercially available lemon juice, used to dissolve heroin before injection, has been associated with blood-stream infection with *Candida* spp.

- **Substances injected.** Injection of cocaine is associated with a variety of vascular complications. Certain formulations of heroin have been linked with particular infections, e.g. wound botulism with black tar heroin. Drugs are often mixed with other substances, e.g. talc.

- **Blood-borne virus status.** Results of previous HIV-1 and hepatitis virus tests or vaccinations for hepatitis viruses should be recorded.
- **Surreptitious use of antimicrobials.** Addicts may use antimicrobials to self-treat infections, masking initial blood culture results.

Key findings on clinical examination are shown in Figure 13.2. It can be difficult to distinguish the effects of infection from the effects of drugs or drug withdrawal (excitement, tachycardia, sweating, marked myalgia, confusion). Stupor and delirium may result from drug administration but may also indicate meningitis or encephalitis. Non-infected venous thromboembolism is also common in this group.

**Investigations**

The initial investigations are as for any fever (see above), including a chest X-ray and blood cultures. Since blood sampling may be difficult, contamination is often a problem. Echocardiography to detect infective endocarditis should be performed in all injection drug-users with: bacteraemia due to *Staphylococcus aureus* or other organisms associated with endocarditis (Fig. 13.3A); thromboembolic phenomena; or a new or previously undocumented murmur. Endovascular infection should also be suspected if lung abscesses or pneumatoceles are detected radiologically. Additional imaging should be focused on sites of injection or of localising symptoms and signs (Fig. 13.3B). Any pathological fluid collections should be sampled.

Urinary toxicology tests may suggest a non-infectious cause of the presenting complaint. While being investigated, all injection drug-users should be offered testing for infection with hepatitis B and C virus and HIV-1.

Microbiological results are crucial in guiding therapy. Injection drug-users may have more than one infection. Skin and soft tissue infections are most often due to *Staph. aureus* or streptococci, and sometimes to *Clostridium* spp. or anaerobes. Pulmonary infections are most often due to the common pathogens causing community-acquired pneumonia, tuberculosis or septic emboli (Fig. 13.3C). Endocarditis with septic emboli commonly involves *Staph. aureus* and viridans streptococci, but *Pseudomonas aeruginosa* and *Candida* spp. are also encountered.

**Management**

Empirical therapy of fever in the injection drug-user includes an antistaphylococcal penicillin (e.g. flucloxacinil) or, if meticillin-resistant *Staph. aureus* (MRSA) is prevalent in the community, a glycopeptide (e.g. vancomycin). Once a particular pathogen is identified, specific therapy is commenced, with modification when antimicrobial susceptibility is available. In injection drug-users, right-sided endocarditis due to *Staph. aureus* is customarily treated with high-dose intravenous flucloxacinil. In left-sided *Staph. aureus* endocarditis, aminoglycoside therapy may be added. Right-sided endocarditis caused by MRSA is usually treated with 4 weeks of vancomycin plus gentamicin for the first week. Specialist advice should be sought.

For localised infections of the skin and soft tissues, oral therapy with agents active against staphylococci, streptococci and anaerobes is appropriate (e.g. flucloxacinil plus co-amoxiclav or clindamycin). Non-adherence with prescribed antimicrobial regimens leads to a high rate of relapse for all infections in this patient group.

**Fever in the immunocompromised host**

Immunocompromised hosts include those with congenital immunodeficiency (p. 78), HIV infection (Ch. 14) and iatrogenic immunosuppression induced by chemotherapy (p. 276), transplantation (p. 95) or immunosuppressant medicines, including high-dose corticosteroids. Metabolic abnormalities, such as under-nutrition or hyperglycaemia, may also contribute. Multiple elements of the immune system are potentially compromised. A patient may have impaired neutrophil function from chemotherapy, impaired T-cell and/or B-cell responses due to underlying malignancy, T-cell and phagocytosis defects due to corticosteroids, mucositis from chemotherapy and an impaired skin barrier due to insertion of a central venous catheter.

Fever may result from infectious or from non-infectious causes, including drugs, vasculitis, neoplasm, organising pneumonitis, lymphoproliferative disease,
**INFECTION DISEASE**

graft-versus-host disease (in recipients of haematopoietic stem cell transplants; p. 1017) or Sweet’s syndrome (reddish nodules or plaques with fever and leucocytosis, in association with haematological malignancy).

**Clinical assessment**

The following should be addressed in the history:

- Identification of the immunosuppressant factors, and nature of the immune defect.
- Any past infections and their treatment. Infections may recur and antimicrobial resistance may have been acquired in response to prior therapy.
- Exposure to infections, including opportunistic infections that would not cause disease in an immunocompetent host.
- Prophylactic medicines and vaccinations administered.

Examination should include inspection of the normal physical barriers provided by skin and mucosal surfaces and, in particular, central venous catheters, the mouth, sinuses, ears and perianal area, (digital rectal examination should be avoided). Disseminated infections can manifest as cutaneous lesions. The areas around fingernails and toenails should also be inspected closely.

**Investigations**

Initial screening tests are as described above (p. 296). Immunocompromised hosts often have decreased inflammatory responses leading to attenuation of physical signs, such as neck stiffness with meningitis, radiological features and laboratory findings, such as leucocytosis. Chest CT scan should be considered in addition to chest X-ray when respiratory symptoms occur. Abdominal imaging may also be warranted, particularly if there is right lower quadrant pain, which may indicate typhilitis (inflammation of the caecum) in neutropenic patients. Blood cultures from a central venous catheter, urine cultures, and stool cultures if diarrhoea is present are also recommended.

Nasopharyngeal aspirates are sometimes diagnostic, as immunocompromised hosts may shed respiratory viruses for prolonged periods. Skin lesions should be biopsied if nodules are present, and investigation should include fungal stains. Useful molecular techniques include polymerase chain reaction (PCR) for CMV and *Aspergillus* spp. DNA, and antigen assays (e.g. cryptococcal antigen (CrAg) for *Cryptococcus neoforms*, and galactomannan for *Aspergillus* spp. in blood or *Legionella pneumophila* type 1 in urine). Antibody detection is rarely useful in immunocompromised patients. Patients with respiratory signs or symptoms should be considered for bronchoscopy in order to obtain bronchoalveolar lavage fluid to detect pathogens, including *Pneumocystis jiroveci* (carinii), as well as bacteria, fungi and viruses.

**Neutropenic fever**

Neutropenic fever is strictly defined as a neutrophil count of less than 0.5 x 10⁹/L (p. 1004) and a single axillary temperature above 38.5°C or three recordings above 38.0°C over a 12-hour period, although the infection risk increases progressively as the neutrophil count drops below 1.0 x 10⁹/L. Patients with neutropenia are particularly prone to bacterial or fungal infection. Gram-positive organisms are the most common pathogens, particularly in association with in-dwelling catheters.

Empirical broad-spectrum antimicrobial therapy is commenced as soon as neutropenic fever occurs and cultures have been obtained. The most common regimens for neutropenic sepsis are broad-spectrum penicillins, such as piperacillin–tazobactam IV. Although aminoglycosides are commonly used in combination, routine use is not supported by trial data (Box 13.6). If fever has not resolved after 3–5 days, empirical antifungal therapy (e.g. caspofungin) is added (p. 159). An alternative antifungal strategy is to use azole prophylaxis in high-risk patients and markers of early fungal infection, such as galactomannan antigen, to guide initiation of antifungal treatment (a ‘pre-emptive approach’).

**Post-transplantation fever**

Fever in transplant recipients may be due to infection, episodes of graft rejection in solid organ transplant recipients, or graft-versus-host disease following haematopoietic stem cell transplantation (HSCT; p. 1017).

Infections in solid transplant recipients are grouped according to the time of onset (Box 13.7). Those in the first month are related to the underlying condition or surgical complications. Those occurring 1–6 months after transplantation are characteristic of impaired T-cell function. Risk factors for CMV infection have

---

**EBM 13.6 Treatment of neutropenic fever**

‘Broad spectrum β-lactam monotherapy is as effective as β-lactam-aminoglycoside combination therapy for neutropenic fever in many settings.’

been identified and patients commonly receive either prophylaxis or intensive monitoring involving regular testing for CMV DNA by PCR and early initiation of anti-CMV therapy using intravenous ganciclovir or oral valganciclovir if tests become positive.

Following HSCT, infections in the first 4 weeks are more common in patients receiving a myeloablative conditioning regimen (see Box 13.7). Later infections are more common if an allogeneic procedure is performed.

Post-transplant lymphoproliferative disorder (PTLD) is an Epstein–Barr virus (EBV)-associated lymphoma that can complicate transplantation, particularly when primary EBV infection occurs after transplantation.

### 13.8 Common causes of blood-stream infection

<table>
<thead>
<tr>
<th>Community-acquired</th>
<th>Nosocomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>Staph. aureus</strong>, including MRSA</td>
<td>- <strong>Other streptococci</strong></td>
</tr>
<tr>
<td>- <strong>Streptococcus pneumoniae</strong></td>
<td>- <strong>Escherichia coli</strong></td>
</tr>
<tr>
<td>- <strong>Coagulase-negative staphylococci</strong></td>
<td>- <strong>Enterococci, including VRE</strong></td>
</tr>
<tr>
<td>- <strong>Candida spp.</strong></td>
<td>- <strong>Gram-negative bacteria</strong></td>
</tr>
</tbody>
</table>

(MRSA = meticillin-resistant Staph. aureus; VRE = vancomycin-resistant enterococci)

---

### Positive blood culture

Blood-stream infection (BSI) or bacteraemia is a frequent presentation of infection. This can be community-acquired or may arise in hospital (‘nosocomial’). The most common causes are shown in Box 13.8. In immunocompromised hosts, a wider range of microorganisms may be isolated, e.g. fungi in neutropenic hosts.

Primary bacteraemia refers to cases in which the site of infection is unknown; this applies in approximately 10% of community-acquired cases and approximately 30% of nosocomial cases, and is more common in *Staph. aureus* bacteraemias. In community-acquired *Staph. aureus* bacteraemia, 20–30% of cases are associated with infective endocarditis and up to 10% are due to osteomyelitis. Peripheral and central venous catheters are an important source of nosocomial BSI.

BSI has an associated mortality of 15–40%, depending on the setting, host and microbial factors.

### Clinical assessment

The history should determine the setting in which BSI has occurred. Host factors predisposing to infection include skin disease, diabetes mellitus, injection drug use, the presence of a central venous, urinary or haemodialysis catheter, and surgical procedures, especially those involving the implantation of prosthetic materials (in particular, endovascular prostheses).

Physical examination should focus on signs of endocarditis (p. 623), evidence of bone or joint infection (tenderness or restriction of movement), and abdominal or flank tenderness. Central venous catheters should be examined for erythema or purulence at the exit site. Particularly in cases with *Candida* spp. infection or suspected infective endocarditis, fundoscopy after pupil dilatation should be performed.

### Investigations

Positive blood cultures may be caused by contaminants. When isolated from only one bottle, or from all bottles from one venesection, coagulase-negative staphylococci often represent contamination. Repeated isolation of this organism, however, should raise suspicion of infective endocarditis or, in a patient with any form of prosthetic material, prosthetic infection. Viridans streptococci occasionally cause transient non-significant bacteraemia or blood culture contamination but, in view of their association with infective endocarditis, significant infection must always be sought clinically. *Bacillus* spp. (‘aerobic spore bearers’) and *Clostridium* spp. often represent incidental transient bacteraemia or contamination, but certain species (e.g. *C. septicum*) may be genuine pathogens.

Further investigations are influenced by the causative organism and setting. Initial screening tests are similar to those for fever (p. 296) and should include chest X-ray, urine culture and, in many cases, ultrasound or other imaging of the abdomen. Imaging should also include any areas of bone or joint pain and any prosthetic material, e.g. a prosthetic joint or an aortic graft.

Echocardiography should be considered for those patients with BSI who have valvular heart disease or clinical features of endocarditis (p. 623), whose cultures reveal an organism that is a common cause of endocarditis (e.g. *Staph. aureus*, viridans streptococci or enterococci), those in whom multiple blood cultures are positive for the same organism, and those with rapid positive result on culture. The sensitivities of trans-thoracic echocardiography (TTE) and transoesophageal echocardiography (TOE) for the detection of vegetations are 50–90% and over 95%, respectively. Therefore, if TTE is negative, TOE should be performed (Box 18.115, p. 627).

Certain rare causes of BSI have specific associations that warrant further investigation. *Strep. bovis* (biotype I, *Strep. gallolyticus*) endocarditis and *C. septicum* BSI are both associated with colonic carcinoma and their isolation is an indication for colonoscopy.

### Management

BSI requires antimicrobial therapy and attention to the source of infection, including surgical drainage if appropriate. Two weeks of therapy may be sufficient for *Staph. aureus* BSI from central and peripheral venous catheter infections when the source is identified and removed, for uncomplicated skin and soft tissue infections, and for uncomplicated right-sided infective endocarditis. Other *Staph. aureus* BSIs are usually treated for 4–6 weeks.

### Central venous catheter infections

Infections of central venous catheters typically involve the catheter lumen and are associated with fever, positive blood cultures and, in some cases, signs of purulence or exudate at the site of insertion. Infection is more common in temporary catheters inserted into the groin or jugular vein than those in the subclavian vein. Tunneled catheters, e.g. Hickman catheters, may also develop tunnel site infections.

Staphylococci account for 70–90% of catheter infections, with coagulase-negative staphylococci more
common than *Staph. aureus*. Other causes include enterococci and Gram-negative bacilli. Unusual Gram-negative organisms, such as *Citrobacter freundii* and *Pseudomonas fluorescens*, cause pseudo-outbreaks and raise the possibility of non-sterile infusion equipment or infusion. *Candida* spp. are a common cause of line infections, particularly in association with total parenteral nutrition. Non-tuberculous mycobacteria may cause tunnel infections.

**Investigations and management**

In bacteraemic patients with fever and no other obvious source of infection, a catheter infection is likely. Local evidence of erythema, purulence or thrombophlebitis supports the diagnosis. However, microbiological confirmation is essential (p. 140). Catheter-related infection is suggested by higher colony counts or shorter time to positivity in blood cultures obtained through the catheter than in peripheral blood cultures. If the line is removed, a semi-quantitative culture of the tip should confirm the presence of 15 or more colony-forming units, but this is retrospective and does not detect luminal infection.

For coagulase-negative staphylococcal line infections, the options are to remove the line or, particularly in the case of tunnelled catheters, to treat empirically with a glycopeptide antibiotic, e.g. vancomycin, with or without the use of antibiotic-containing lock therapy to the catheter for approximately 14 days. For *Staph. aureus* infection, the chance of curing an infection with the catheter in situ is low and the risks from infection are high. Therefore, unless the risks of catheter removal outweigh the benefits, treatment involves catheter removal, followed by 14 days of antimicrobial therapy; the same applies to infections with *Candida* spp. or *Bacillus* spp.

Infection prevention is a key component of the management of vascular catheters. Measures include strict attention to hand hygiene, optimal sitting, full aseptic technique on insertion and subsequent interventions, skin antisepsis with chlorhexidine and isopropyl alcohol, daily assessment of catheter sites (e.g. with visual infusion phlebitis (VIP) score (p. 330)), and daily consideration of the continuing requirement for catheterisation. The use of catheters impregnated with antimicrobials such as chlorhexidine or silver is advocated in some settings.

### Sepsis

Sepsis is discussed on page 200. It describes patients with evidence of infection and signs of the systemic inflammatory response syndrome (SIRS), which entails two of: temperature over 38°C or under 36°C; pulse rate more than 90 beats per minute; respiratory rate over 20 breaths per minute or PCO2 below 4.3 kPa (32.5 mmHg); and white blood cell count over 12 or below 4 × 109/L (Box 8.2, p. 184). Septic shock describes sepsis plus hypotension (systolic blood pressure below 90 mmHg systolic or a fall of more than 40 mmHg from baseline that is not responsive to fluid challenge or due to another cause). It may be complicated by multi-organ failure and requires intensive care unit admission.

Sepsis largely results from host responses to microbial lipopolysaccharide, peptidoglycans, lipoproteins or
superantigens, and there are many infectious causes (Box 13.9). The results of blood cultures and known host factors guide initial investigations. Patients who are immunocompromised may have a broader range of causal pathogens which may be harder to culture, including mycobacteria and fungi. In any individual who has recently visited the tropics, malaria must also be considered.

**Severe skin and soft tissue infections**

Skin and soft tissue infections (SSTIs) are an important cause of sepsis. Cases can be classified as in Box 13.10, according to the clinical features and microbiological findings. In some cases, severe systemic features may be out of keeping with mild local features.

**Necrotising fasciitis**

In necrotising fasciitis, cutaneous erythema and oedema progress to bullae or areas of necrosis. Unlike in cellulitis, pain may be disproportionately intense in relation to the visible cutaneous features. The infection spreads quickly along the fascial plane. Type 1 necrotising fasciitis is a mixed infection with Gram-negative bacteria and anaerobes, often seen post-operatively in diabetic or immunocompromised hosts. Subcutaneous gas may be present. Type 2 necrotising fasciitis is caused by group A or other streptococci. Approximately 60% of cases are associated with streptococcal toxic shock syndrome (p. 331).

Necrotising fasciitis is a medical emergency, requiring immediate surgical débridement with inspection of the involved muscle groups, in addition to antimicrobial therapy (Fig. 13.4). Empiric treatment is with broad-spectrum agents (e.g. piperacillin–tazobactam plus clindamycin and ciprofloxacin; meropenem monotherapy; or third-generation cephalosporin plus metronidazole). Hyperbaric oxygen therapy may be considered for polymicrobial infection. Group A streptococcal infection is treated with benzylpenicillin plus clindamycin, and often immunoglobulin.

**Gas gangrene**

Although *Clostridium* spp. may colonise or contaminate wounds, no action is required unless there is evidence of spreading infection. Infection may be limited to tissue that is already damaged (anaerobic cellulitis) or involve healthy muscle (gas gangrene).

In anaerobic cellulitis, usually that due to *C. perfringens* or to other strains infecting devitalised tissue following a wound, gas forms locally and extends along tissue planes but bacteraemia does not occur. Prompt surgical débridement of devitalised tissue and therapy with penicillin or clindamycin is usually effective.

Gas gangrene (clostridial myonecrosis) is defined as acute invasion of healthy living muscle undamaged by previous trauma, and is most commonly caused by *C. perfringens*. In at least 70% of cases, it follows deep penetrating injury sufficient to create an anaerobic (ischaemic) environment and allow clostridial introduction and proliferation. Severe pain at the site of the injury progresses rapidly over 18–24 hours. Skin colour changes from pallor to bronze/purple discoloration and the skin is tense, swollen, oedematous and exquisitely tender. Gas in tissues may be obvious, with crepitus on clinical examination, or visible on X-ray, CT or ultrasound. Signs of systemic toxicity develop rapidly, with high leukocytosis, multi-organ dysfunction, raised creatine kinase and evidence of disseminated intravascular coagulation and haemolysis. Antibiotic therapy with high-dose intravenous penicillin and clindamycin is recommended, coupled with aggressive surgical débridement of the affected tissues. Alternative agents include cephalosporins and metronidazole. Hyperbaric oxygen has a putative but controversial role.

**Other SSTIs**

‘Synergistic gangrene’ is a combined infection with anaerobes and other bacteria (*Staph. aureus* or Gram-negatives). When this affects the genital/perineal area, it is known as ‘Fournier’s gangrene’. Severe gangrenous cellulitis in immunocompromised hosts may involve Gram-negative bacteria or fungi. *Entamoeba histolytica* can cause soft tissue necrosis following abdominal surgery in areas of the world where infection is common. Contact with shellfish in tropical areas and regions such as the Gulf of Mexico can lead to infection with *Vibrio vulnificus*, which causes soft tissue necrosis and bullae. Patients with chronic liver disease are particularly susceptible.

---

**Box 13.10 Severe necrotising soft tissue infections**

- Necrotising fasciitis (primarily confined to subcutaneous fascia and fat)
- Clostridial anaerobic cellulitis (confined to skin and subcutaneous tissue)
- Non-clostridial anaerobic cellulitis
- Progressive bacterial synergistic gangrene (*Staph. aureus* + micro-aerophilic streptococcus) (‘Meleney’s gangrene’, primarily confined to skin)
- Pyomyositis (discrete abscesses within individual muscle groups)
- Clostridial myonecrosis (gas gangrene)
- Anaerobic streptococcal myonecrosis (non-clostridial infection mimicking gas gangrene)
- Group A streptococcal necrotising myositis

---

**Fig. 13.4** Excision following necrotising fasciitis in an injection drug-user.
Acute diarrhoea (p. 857), sometimes with vomiting, is the predominant symptom in infective gastroenteritis (Box 13.11). Acute diarrhoea may also be a symptom of other infectious and non-infectious diseases (Box 13.12). Stress, whether psychological or physical, can also produce loose stools.

The World Health Organization (WHO) estimates that there are more than 1000 million cases of acute diarrhoea annually in developing countries, with 3–4 million deaths, half of these in infants and children. In developed countries, diarrhoea remains an important problem and the elderly are most vulnerable (Box 13.13). The majority of episodes are due to infections spread by the faecal–oral route and transmitted either on fomites, or contaminated hands, or in food or water. Measures such as the provision of clean drinking water, appropriate disposal of human and animal sewage, and the application of simple principles of food hygiene can all limit gastroenteritis.

The clinical features of food-borne gastroenteritis vary. Some organisms (Bacillus cereus, Staph. aureus and Vibrio cholerae) elute exotoxins which cause vomiting and/or so-called ‘secretory’ diarrhoea (watery diarrhoea without blood or faecal leucocytes, reflecting small bowel dysfunction). In general, the time from ingestion to the onset of symptoms is short and, other than dehydration, little systemic upset occurs. Other organisms, such as Shigella spp., Campylobacter spp. and enterohaemorrhagic E. coli (EHEC), may directly invade the mucosa of the small bowel or produce cytotoxins that cause mucosal ulceration, typically affecting the terminal small bowel and colon. The incubation period is longer and more systemic upset occurs, with prolonged bloody diarrhoea. Salmonella spp. are capable of invading enterocytes, and of causing both a secretory response and invasive disease with systemic features. This is seen with Salmonella typhi and S. paratyphi (enteric fever), and, in the immunocompromised host, with non-typhoidal Salmonella spp.

**Clinical assessment**

The history should address foods ingested (Box 13.14), duration and frequency of diarrhoea, presence of blood or steatorrhoea, abdominal pain and tenesmus, and whether other people have been affected. Fever and bloody diarrhoea suggest an invasive, colitic, dysenteric process. An incubation period of less than 18 hours suggests toxin-mediated food poisoning, and longer than 5 days suggests diarrhoea caused by protozoa or helminths. Person-to-person spread suggests certain infections, such as shigellosis or cholera.

Examination includes assessment of the degree of dehydration by skin turgor, pulse and blood pressure measurement. The urine output and ongoing stool losses should be monitored.
13.14 Foods associated with infectious illness, including gastroenteritis

| Raw seafood          | • Norovirus           | • Hepatitis A |
| Raw eggs             | • Vibrio spp.         |               |
| Undercooked meat or poultry | • Salmonella spp.    | • EHEC        |
|                      | • Campylobacter spp.  | • C. perfringens |
| Unpasteurised milk or juice | • Salmonella spp.    | • EHEC        |
|                      | • Campylobacter spp.  | • Y. enterocolitica |
| Unpasteurised soft cheeses | • Salmonella spp.    | • Y. enterocolitica |
|                      | • Campylobacter spp.  | • L. monocytogenes |
| Home-made canned goods | • C. botulinum       |               |
|                      | • C. perfringens      |               |
|                      | • L. monocytogenes    |               |

**Investigations**

These include stool inspection for blood and microscopy for leukocytes, and also an examination for ova, cysts and parasites if the history indicates former tropical residence or travel. Stool culture should be performed and *C. difficile* toxin sought. FBC and serum electrolytes indicate the degree of inflammation and dehydration. In a malarious area, a blood film for malaria parasites should be obtained. Blood and urine cultures and a chest X-ray may identify alternative sites of infection, particularly if the clinical features suggest a syndrome other than gastroenteritis.

**Management**

All patients with acute, potentially infective diarrhoea should be appropriately isolated to minimise person-to-person spread of infection. If the history suggests a food-borne source, public health measures must be implemented to identify the source and to establish whether other linked cases exist (p. 147).

**Fluid replacement**

Replacement of fluid losses in diarrhoeal illness is crucial and may be life-saving.

Although normal daily fluid intake in an adult is only 1–2 L, there is considerable additional fluid movement in and out of the gut in secretions (Fig. 22.7, p. 843). Altered gut resorption with diarrhea can result in substantial fluid loss, e.g. 10–20 L of fluid may be lost in 24 hours in cholera. The fluid lost in diarrhoea is isotonic, so both water and electrolytes need to be replaced. Absorption of electrolytes from the gut is an active process requiring energy. Infected mucosa is capable of very rapid fluid and electrolyte transport if carbohydrate is available as an energy source. Oral rehydration solutions (ORS) therefore contain sugars, as well as water and electrolytes (Box 13.15). ORS can be just as effective as intravenous replacement fluid, even in the management of cholera. In mild to moderate gastroenteritis, adults should be encouraged to drink fluids and, if possible, continue normal dietary food intake. If this is impossible, e.g. due to vomiting, intravenous fluid administration will be required. In very sick patients, or those with cardiac or renal disease, monitoring of urine output and central venous pressure may be necessary. The volume of fluid replacement required should be estimated based on the following considerations.

- **Replacement of established deficit.** After 48 hours of moderate diarrhoea (6–10 stools per 24 hours), the average adult will be 2–4 L depleted from diarrhoea alone. Associated vomiting will compound this. Adults with this symptomatology should therefore be given rapid replacement of 1–1.5 L, either orally (ORS) or by intravenous infusion (normal saline), within the first 2–4 hours of presentation. Longer symptomatology or more persistent/severe diarrhoea rapidly produces fluid losses comparable to diabetic ketoacidosis and is a metabolic emergency requiring active intervention.

- **Replacement of ongoing losses.** The average adult’s diarrhoeal stool accounts for a loss of 200 mL of isotonic fluid. Stool losses should be carefully charted and an estimate of ongoing replacement fluid calculated. Commercially available rehydration sachets are conveniently produced to provide 200 mL of ORS; one sachet per diarrhoea stool is an appropriate estimate of supplementary replacement requirements.

- **Replacement of normal daily requirement.** The average adult has a daily requirement of 1–1.5 L of fluid in addition to the calculations above. This will be increased substantially in fever or a hot environment.

**Antimicrobial agents**

In non-specific gastroenteritis, antibiotics have been shown to shorten symptoms by only 1 day in an illness usually lasting 1–3 days. This benefit, when related to the potential for the development of antimicrobial resistance or side-effects, does not justify treatment, except if there is systemic involvement, a host with immunocompromise or significant comorbidity.

Evidence suggests that, in EHEC infections, the use of antibiotics may make the complication of haemolytic uraemic syndrome (HUS; p. 495) more likely due to
increased toxin release. Antibiotics should therefore not be used in this condition.

Conversely, antibiotics are indicated in *Sh. dysenteriae* infection and in invasive salmonellosis – in particular, typhoid fever. Antibiotics may also be advantageous in cholera epidemics, reducing infectivity and controlling the spread of infection.

**Antidiarrhoeal, antimotility and antisecretory agents**

These agents are not usually recommended in acute infective diarrhoea. Loperamide, diphenoxylate and opiates are potentially dangerous in dysentery in childhood, causing intussusception. Antisecretory agents, such as bismuth and chlorpromazine, may be effective but can cause significant sedation. They do not reduce stool fluid losses, although the stools may appear more bulky. Adsorbents, such as kaolin or charcoal, have little effect.

**Non-infectious causes of food poisoning**

Whilst acute food poisoning and gastroenteritis are most frequently caused by bacteria or their toxins, a number of non-infectious causes must be considered in the differential diagnosis.

**Plant toxins**

Legumes and beans produce oxidants which are toxic to people with glucose-6-phosphate dehydrogenase (G6PD) deficiency (p. 1029). Consumption produces headache, nausea and fever, progressing to potentially severe haemolysis, haemoglobinuria and jaundice (favism). Red kidney beans, if incompletely cooked, cause acute abdominal pain and diarrhoea from their lectin content. Adequate cooking abolishes this.

Alkaloids develop in potato tubers exposed to light, causing green discoloration. Ingestion induces acute vomiting and anticholinesterase-like activity.

Fungi and mushrooms of the *Psilocybe* spp. produce hallucinogens. Many fungal species induce a combination of gastroenteritis and cholineretic symptoms of blurred vision, salivation, sweating and diarrhoea. *Amanita phalloides* (‘death cap’) causes acute abdominal cramps and diarrhoea, followed by inexorable hepato-renal failure, often fatal.

**Chemical toxins**

**Paralytic shellfish toxin**

Saxitoxin from dinoflagellates, responsible for ‘red tides’, is concentrated in bivalve molluscs, e.g. mussels, clams, oysters, cockles and scallops. Consumption produces gastrointestinal symptoms within 30 minutes, followed by perioral paraesthesia and even respiratory paralysis. The UK water authorities ban the harvesting of molluscs at times of the year associated with excessive dinoflagellate numbers.

**Ciguatera fish poisoning**

Warm-water coral reef fish acquire ciguatoxin from dinoflagellates in their food chain. Consumption produces gastrointestinal symptoms 1–6 hours later, with associated paraesthesiae of the lips and extremities, distorted temperature sensation, myalgia and progressive flaccid paralysis. Autonomic dysfunction with hypotension may occur. In the South Pacific and Caribbean, there are 50000 cases per year, with a case fatality of 0.1%. The gastrointestinal symptoms resolve rapidly but the neuropathic features may persist for months.

**Scombrototoxic fish poisoning**

Under poor storage conditions, histidine in scombroid fish – tuna, mackerel, bonito, skipjack and the canned dark meat of sardines – may be converted by bacteria to histamine and other chemicals. Consumption produces symptoms within minutes, with flushing, burning, sweating, urticaria, pruritus, headache, colic, nausea and vomiting, diarrhoea, bronchospasm and hypotension. Management is with salbutamol and antihistamines. Occasionally, intravenous fluid replacement is required.

**Heavy metals**

Thallium and cadmium can cause acute vomiting and diarrhoea resembling staphylococcal enterotoxin poisoning.

**Antimicrobial-associated diarrhoea**

Antimicrobial-associated diarrhoea (AAD) is a common complication of antimicrobial therapy, especially with broad-spectrum agents. It is most common in the elderly but can occur at all ages. Although the specific mechanism is unknown in most AAD, *Clostridium difficile* is implicated in 20–25% of cases and is the most common cause amongst patients with evidence of colitis. Infection is diagnosed by detection of *C. difficile* toxins and is usually treated with metronidazole or vancomycin (p. 343). *C. perfringens* is a rarer cause which usually remains undiagnosed, and *Klebsiella oxytoca* is an occasional cause of antibiotic-associated haemorrhagic colitis.

**Infections acquired in the tropics**

Recent decades have seen unprecedented increases in long-distance business and holiday travel, as well as extensive migration. Although certain diseases retain their relatively fixed geographical distribution, being dependent on specific vectors or weather conditions, many travel with their human hosts and some may then be transmitted to other people. This means that the pattern of infectious diseases seen in each country changes constantly, and travel history and information on countries previously lived in, particularly during childhood, are crucial.

In general, the diversity of infectious diseases is greater in tropical than in temperate countries, and people in temperate countries have immunity to a narrower range of infections, reflecting less exposure in childhood and less ongoing boosting of immunity later in life, so that the most common travel-associated infections are those which are acquired by residents of temperate countries during visits to the tropics. In addition, those who have lived in tropical areas may lose immunity when they move to temperate countries and become susceptible when visiting their homeland.

Most travel-associated infections can be prevented. Pre-travel advice is tailored to the destination and the traveller (Box 13.16). It includes avoidance of insect bites (using at least 20% diethyltoluamide (DEET)), sun protection (sunscreen with a sun protection factor (SPF) of at least 15), food and water hygiene (‘Boil it, cook it, peel
it or forget it’), how to respond to travellers’ diarrhoea (seek medical advice if bloody or lasts more than 48 hours) and, if relevant, safe sex (condom use).

**Fever in travellers recently in the tropics**

Presentation with unexplained fever is common in travellers who are visiting or have recently travelled to tropical areas. Frequent final diagnoses in such patients are malaria, typhoid fever, viral hepatitis and dengue fever. Travellers to West Africa may have viral haemorrhagic fevers (VHF), such as Lassa fever, Crimean–Congo haemorrhagic fever, Marburg and Ebola (see Box 13.39, p. 325). Those to South-east Asia may have avian influenza (H5N1), which requires special isolation precautions.

**Clinical assessment**

The approach to unexplained fever is described above and key questions are listed in Box 13.17. Medicines purchased in some countries may have reduced efficacy, e.g. for malaria prophylaxis. Consult reliable up-to-date sources about resistance to antimalarial drugs in the country visited. Vaccinations against yellow fever and hepatitis A and B are sufficiently effective to virtually exclude these infections. Oral and injectable typhoid vaccinations are 70–90% effective.

The differential diagnosis is guided by the clinical scenario, presence of specific exposures (Box 13.18) and the incubation period (Box 13.19). *Falciparum* malaria

---

**Box 13.16 How to assess health needs in travellers before departure**

- Destination
- Personal details, including previous travel experience
- Dates of trip
- Itinerary and purpose of trip
- Personal medical history, including pregnancy, medication and allergies (e.g. to eggs, vaccines, antibiotics)
- Past vaccinations
  - Travel-related? Typhoid, yellow fever, hepatitis A, hepatitis B, meningococcal ACW135Y, rabies, Japanese B encephalitis, tick-borne encephalitis
- Malaria prophylaxis: questions influencing the choice of antimalarial drugs are destination, past experience with antimalarials, history of epilepsy or psychiatric illness

*Further information is available at www.fitfortravel.nhs.uk/

**Box 13.17 How to obtain a history from travellers to the tropics with fever**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Factors to ascertain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countries visited and dates of travel</td>
<td>Relate travel to known outbreaks of infection or antimicrobial resistance</td>
</tr>
<tr>
<td>Determine the environment visited</td>
<td>Travel to rural environments, forests, rivers or lakes</td>
</tr>
<tr>
<td>Clarify where the person slept</td>
<td>Sleeping in huts, use of bed nets, sleeping on the ground</td>
</tr>
<tr>
<td>Establish what he/she was doing</td>
<td>Exposure to people with medical illness, animals, soil, lakes, rivers</td>
</tr>
<tr>
<td>History of insect bites</td>
<td>Type of insect responsible, circumstances (location, time of day etc.), preventive measures</td>
</tr>
<tr>
<td>Dietary history</td>
<td>Ingestion of uncooked foods, salads and vegetables, meats (especially if undercooked), shellfish, molluscs, unpasteurised dairy products, unbotiled water and sites at which food prepared</td>
</tr>
<tr>
<td>Sexual history</td>
<td>History of sexual intercourse with commercial sex workers, local population</td>
</tr>
<tr>
<td>Malaria prophylaxis</td>
<td>Type of prophylaxis</td>
</tr>
<tr>
<td>Vaccination history</td>
<td>Receipt of pre-travel vaccines and appropriateness to area visited</td>
</tr>
<tr>
<td>History of any treatments received while abroad</td>
<td>Receipt of medicines, local remedies, blood transfusions or surgical procedures</td>
</tr>
</tbody>
</table>

---

**Box 13.18 Specific exposures and causes of fever in the tropics**

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Infection or disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mosquito bite</td>
<td>Malaria, dengue fever, Chikungunya, filariasis, tularaemia</td>
</tr>
<tr>
<td>Tsetse fly bite</td>
<td>African trypanosomiasis</td>
</tr>
<tr>
<td>Tick bite</td>
<td>Rickettsial infections, including typhus, Lyme disease, tularaemia, Crimean–Congo haemorrhagic fever, Kyasanur forest disease, babesiosis, tick-borne encephalitis</td>
</tr>
<tr>
<td>Louse bite</td>
<td>Typhus</td>
</tr>
<tr>
<td>Flea bite</td>
<td>Plague</td>
</tr>
<tr>
<td>Sandfly bite</td>
<td>Leishmaniasis, arbovirus infection</td>
</tr>
<tr>
<td>Reduvid bug</td>
<td>Q fever, brucellosis, anthrax, plague, tularaemia, viral haemorrhagic fevers, rables</td>
</tr>
<tr>
<td>Fresh-water swimming</td>
<td>Schistosomiasis, leptospirosis, <em>Naegleria fowleri</em></td>
</tr>
<tr>
<td>Exposure to soil</td>
<td>Inhalation: dimorphic fungi</td>
</tr>
<tr>
<td></td>
<td>Inhalation or inoculation: <em>Burkholderia pseudomallei</em></td>
</tr>
<tr>
<td></td>
<td>Inoculation (most often when barefoot): hookworms, <em>Strongyloides stercoralis</em></td>
</tr>
<tr>
<td>Raw or undercooked fruit and vegetables</td>
<td>Enteric bacterial infections, hepatitis A or E virus, <em>Fasciola hepatica</em>, <em>Toxocara spp.</em>, <em>Echinococcus granulosus</em> (hydatid disease), <em>Entamoeba histolytica</em></td>
</tr>
<tr>
<td>Undercooked pork</td>
<td><em>Taenia solium</em> (cysticercosis)</td>
</tr>
<tr>
<td>Crustaceans or molluscs</td>
<td>Paragonimiasis, gnathostomiasis, <em>Angiostrongylus cantonensis</em> infection, hepatitis A virus, cholera</td>
</tr>
<tr>
<td>Unpasteurised dairy products</td>
<td>Brucellosis, salmonellosis, abdominal tuberculosis, listeriosis</td>
</tr>
<tr>
<td>Untreated water</td>
<td>Enteric bacterial infections, giardiasis, <em>Cryptosporidium</em> spp. (chronic in immunocompromised), hepatitis A or E virus</td>
</tr>
</tbody>
</table>
Clinical examination is summarised on page 294. Particular attention should be paid to the skin, throat, eyes, nail beds, lymph nodes, abdomen and heart. Patients may be unaware of tick bites or eschars (p. 350). Body temperature should be measured at least twice daily.

**Investigations and management**

Initial investigations should start with thick and thin blood films for malaria parasites, FBC, urinalysis and chest X-ray if indicated. Box 13.20 lists diagnoses and investigations to consider in unexplained acute fever.

Management is directed at the underlying cause. In patients with suspected VHF (p. 324), strict infection control measures with isolation and barrier nursing are implemented to prevent contact with the patient’s body fluids. The risk of VHF should be determined using epidemiological risk factors and clinical signs (Fig. 13.5), and further management undertaken as described on page 324.

**Diarrhoea acquired in the tropics**

Gastrointestinal illness is the most common infection amongst visitors to the tropics, with *Salmonella* ssp., *Campylobacter* ssp. and *Cryptosporidium* spp. infections prevalent worldwide (Box 13.21). Typhoid, paratyphoid, *Shigella* spp. and *Entamoeba histolytica* (amoebiasis) are usually encountered in visitors to the Indian subcontinent or sub-Saharan and southern Africa.

The approach to patients with acute diarrhoea is described on page 306. The benefits of treating travellers’ diarrhoea with antimicrobials are marginal (Box 13.22). The differential diagnosis of diarrhoea persisting for more than 14 days is wide (Box 22.21, p. 857). Parasitic and bacterial causes, tropical malabsorption, inflammatory bowel disease and neoplasia should all be considered. Box 13.23 lists causes encountered particularly in visitors to the tropics. The work-up should include tests for parasitic causes of chronic diarrhoea,
**13.21 Most common causes of travellers’ diarrhoea**

- Enterotoxigenic *E. coli* (ETEC)
- *Shigella* spp.
- *Campylobacter jejuni*
- *Salmonella* spp.
- *Plesiomonas shigelloides*
- Non-cholera *Vibrio* spp.
- *Aeromonas* spp.

**13.22 Antimicrobials in travellers’ diarrhoea**

‘Antimicrobials reduce the duration of acute non-bloody diarrhoea.’


For further information: [www.cochrane.org/cochrane-reviews](http://www.cochrane.org/cochrane-reviews)

**13.23 Causes of chronic diarrhoea acquired in the tropics**

- *Giardia lamblia*
- Strongyloidsis
- Enteropathic *E. coli*
- HIV enteropathy
- Intestinal flukes
- Tropical sprue
- Chronic intestinal schistosomiasis
- Chronic calcific pancreatitis
- Hypolactasia (primary and secondary)

The response to parasite infections is often different when travellers to and residents of endemic areas are compared. Travellers often have recent and light infections associated with eosinophilia. Residents have often been infected for a long time, have evidence of chronic pathology and no longer have eosinophilia.

**Clinical assessment**

A history of travel to known endemic areas for schistosomiasis, onchocerciasis and the filariases will indicate possible causes. Assessment should establish how long patients have spent in endemic areas and the history should address all the elements in Box 13.17.

Physical signs or symptoms that suggest a parasitic cause for eosinophilia include transient rashes (schistosomiasis or strongyloidiasis), fever (Katayama syndrome – p. 377), pruritus (onchocerciasis) or migrating subcutaneous swellings (loiasis, gnathostomiasis) (see Box 13.24). Paragonimiasis can give rise to haemoptysis and the migratory phase of intestinal nematodes or lymphatic filariasis may cause cough, wheezing and transient pulmonary infiltrates. Schistosomiasis induces transient respiratory symptoms with infiltrates in the acute stages and, when eggs reach the pulmonary vasculature in chronic infection, can result in shortness of breath with features of right heart failure due to pulmonary hypertension. Fever and hepatosplenomegaly are seen in schistosomiasis, *Fasciola hepatica* infection and e.g. CMV or disseminated *Mycobacterium avium* complex infections.
**13.24 Parasite infections that cause eosinophilia**

<table>
<thead>
<tr>
<th>Infestation</th>
<th>Pathogen</th>
<th>Clinical syndrome with eosinophilia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongyloidiasis</td>
<td>Strongyloides stercoralis</td>
<td>Larva currens</td>
</tr>
<tr>
<td>Soil-transmitted helminthias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hookworm</td>
<td>Necator americanus</td>
<td>Anaemia</td>
</tr>
<tr>
<td>Ascariasis</td>
<td>Ascaris lumbricoide</td>
<td>Löffler’s syndrome</td>
</tr>
<tr>
<td>Toxocariasis</td>
<td>Toxocara canis</td>
<td>Visceral larva migrants</td>
</tr>
<tr>
<td>Schistosomiasis</td>
<td>Schistosoma haematobium</td>
<td>Katayama fever</td>
</tr>
<tr>
<td></td>
<td>S. mansoni, S. japonicum</td>
<td>Chronic infection</td>
</tr>
<tr>
<td>Filariasis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loiasis</td>
<td>Loa loa</td>
<td>Skin nodules</td>
</tr>
<tr>
<td>Wuchereria bancrofti</td>
<td>W. bancrofti</td>
<td>Lymphangitis, lymphadenopathy, orchitis, intermittent bouts of cellulitis, lymphoedema and elephantiasis</td>
</tr>
<tr>
<td>Brugia malayi</td>
<td>B. malayi</td>
<td>Brugian elephantias similar but typically less severe than that caused by W. bancrofti</td>
</tr>
<tr>
<td>Mansonella perstans</td>
<td>M. perstans</td>
<td>Asymptomatic infection, occasionally subconjunctival nodules</td>
</tr>
<tr>
<td>Onchocerciasis</td>
<td>Onchocerca volvulus</td>
<td>Visual disturbance, dermatitis</td>
</tr>
<tr>
<td>Other nematode infections</td>
<td>Trichinella spiralis</td>
<td>Myositis</td>
</tr>
<tr>
<td>Cestode infections</td>
<td>Taenia saginata, T. solium</td>
<td>Usually asymptomatic; eosinophilia associated with migratory phase</td>
</tr>
<tr>
<td></td>
<td>Echinococcus granulosus</td>
<td>Lesions in liver or other organ; eosinophilia associated with leakage from cyst</td>
</tr>
<tr>
<td>Liver flukes</td>
<td>Fasciola hepatica</td>
<td>Hepatic symptoms; eosinophilia associated with migratory phase</td>
</tr>
<tr>
<td></td>
<td>Clonorchis sinensis</td>
<td>As for fasciolias</td>
</tr>
<tr>
<td></td>
<td>Opisthorchis felineus</td>
<td>As for fasciolias</td>
</tr>
<tr>
<td>Lung fluke</td>
<td>Paragonimus westermani</td>
<td>Lung lesions</td>
</tr>
</tbody>
</table>

toxocarisis (visceral larva migrants). Intestinal worms, such as Ascaris lumbricoide and Strongyloides stercoralis, can cause abdominal symptoms, including intestinal obstruction and diarrhoea. In the case of heavy infestation with Ascaris, this may be due to fat malabsorption and there may be associated nutritional deficits. Schistosoma haematobium can cause haematuria or haematospermia. Toxocara spp. can give rise to choroidal lesions with visual field defects. Angiostrongylus cantonensis and gnathostomiasis induce eosinophilic meningitis, and the hyperinfection syndrome caused by Strongyloides stercoralis in immunocompromised hosts induces meningitis due to Gram-negative bacteria. Myositis is a feature of trichinellosis and cysticercosis, while periorbital oedema is found in trichinellosis.

**Investigations**

The diagnosis of a parasitic infestation requires direct visualisation of adult worms, larvae or ova. Serum antibody detection may not distinguish between active and past infection and is often unhelpful in those born in endemic areas. Radiological investigations may provide circumstantial evidence of parasite infestation. Box 13.25 describes initial investigations for eosinophilia.

**Management**

A specific diagnosis guides therapy. In the absence of a specific diagnosis, many clinicians will give an empirical course of praziquantel if the individual has been potentially exposed to schistosomiasis, or with albendazole/ivermectin if strongylodiasis or intestinal nematodes are likely causes.

**13.25 Initial investigation of eosinophilia**

<table>
<thead>
<tr>
<th>Investigation</th>
<th>Pathogens sought</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stool microscopy</td>
<td>Ova, cysts and parasites</td>
</tr>
<tr>
<td>Terminal urine</td>
<td>Ova of Schistosoma haematobium</td>
</tr>
<tr>
<td>Duodenal aspirate</td>
<td>Filariform larvae of Strongyloides, liver fluke ova</td>
</tr>
<tr>
<td>Day bloods</td>
<td>Microfilariae Brugia malayi, Loa loa</td>
</tr>
<tr>
<td>Night bloods</td>
<td>Microfilariae Wuchereria bancrofti</td>
</tr>
<tr>
<td>Skin snips</td>
<td>Onchocerca volvulus</td>
</tr>
<tr>
<td>Slit lamp examination</td>
<td>Onchocerca volvulus</td>
</tr>
<tr>
<td>Serology</td>
<td>Schistosomiasis, filariasis, strongylodiasis, hydatid, trichinosis etc.</td>
</tr>
</tbody>
</table>

**Skin conditions acquired in the tropics**

Community-based studies in the tropics consistently show that skin infections (bacterial and fungal), scabies and eczema are the most common skin problems (Box 13.26). Scabies and eczema are discussed on pages 1280 and 1283. Cutaneous leishmaniasis and onchocerciasis have defined geographical distributions (pp. 365 and 374). In travellers, secondarily infected insect bites, pyoderma, cutaneous larva migrants and non-specific dermatitis are common.
13.26 Rash in tropical travellers/residents

<table>
<thead>
<tr>
<th>Maculopapular rash</th>
<th>Petechial or purpuric rash</th>
<th>Vesicular rash</th>
<th>Urticarial rash</th>
<th>Nodules or plaques</th>
<th>Migratory linear rash</th>
<th>Migratory papules/nodules</th>
<th>Thickened skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dengue</td>
<td>• Spirillum minus</td>
<td>• Leptospirosis</td>
<td>• Strongyloides stercoralis</td>
<td>• Cutaneous larva migrans (dog hookworms)</td>
<td>• Myiasis (larvae of Tumbu or botfly)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV-1</td>
<td>• Rickettsial infections</td>
<td>• Rickettsia spotted fevers</td>
<td>• Fascioliasis</td>
<td>Strongyloides stercoralis</td>
<td>Tungiasis (Tunga penetrans)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typhoid</td>
<td>• Measles</td>
<td>• Malaria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 13.6 Examples of skin lesions in patients with fever in the tropics. A Subcutaneous nodule due to hookworm infection. B Emerging larva after treatment with petroleum jelly. C Eschar of scrub typhus. D Rat bite fever.

During the investigation of skin lesions, enquiry should be made about habitation, activities undertaken and regions visited (see Box 13.17). Examples of skin lesions in tropical disease are shown in Figure 13.6. Skin biopsies are helpful in diagnosing aetiology. Culture of biopsy material may be needed to diagnose bacterial, fungal, parasitic and mycobacterial infections.

Infections in adolescence

Particular issues of relevance in adolescent patients are shown in Box 13.27.

Infections in pregnancy

Box 13.28 shows some of the infections encountered in pregnancy.
### Infections during pregnancy

<table>
<thead>
<tr>
<th>Infection</th>
<th>Consequence</th>
<th>Prevention and management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubella</td>
<td>Congenital malformation</td>
<td>Vaccination of non-immune mothers</td>
</tr>
<tr>
<td>Cytomegalovirus</td>
<td>Neonatal infection, congenital malformation</td>
<td>Limited prevention strategies</td>
</tr>
<tr>
<td>Varicella zoster virus</td>
<td>Neonatal infection, congenital malformation, severe infection in mother</td>
<td>VZ immune globulin (see Box 13.34, p. 318), or aciclovir if exposure &gt; 4 days previously</td>
</tr>
<tr>
<td>Herpes simplex virus</td>
<td>Congenital or neonatal infection</td>
<td>Aciclovir and consideration of caesarean section for mothers who shed HSV from genital tract at time of delivery. Aciclovir for infected neonates</td>
</tr>
<tr>
<td>Hepatitis B virus</td>
<td>Chronic infection of neonate</td>
<td>Hepatitis B immune globulin and active vaccination of newborn</td>
</tr>
<tr>
<td>HIV-1</td>
<td>Chronic infection of neonate</td>
<td>Antiretrovirals for mother and infant and consideration of caesarean section if HIV-1 viral load detectable. Avoidance of breastfeeding</td>
</tr>
<tr>
<td>Parvovirus B19</td>
<td>Congenital infection</td>
<td>Avoid individuals with acute infection if pregnant</td>
</tr>
<tr>
<td>Measles</td>
<td>More severe infection in mother and neonate</td>
<td>Immunisation of mother</td>
</tr>
<tr>
<td>Dengue</td>
<td>Neonatal dengue if mother has infection &lt; 5 wks prior to delivery</td>
<td>Vector (mosquito) control</td>
</tr>
<tr>
<td>Syphilis</td>
<td>Congenital malformation</td>
<td>Serological testing in pregnancy with prompt treatment of infected mothers</td>
</tr>
<tr>
<td><em>Neisseria gonorrhoeae</em> and <em>Chlamydia trachomatis</em></td>
<td>Neonatal conjunctivitis (ophthalmia neonatorum, p. 422)</td>
<td>Treatment of infection in mother and neonate</td>
</tr>
<tr>
<td>Listeriosis</td>
<td>Neonatal meningitis or bacteraemia, bacteraemia or PUO in mother</td>
<td>Avoidance of unpasteurised cheeses and other dietary sources</td>
</tr>
<tr>
<td>Brucellosis</td>
<td>Possibly increased incidence of fetal loss</td>
<td>Avoidance of unpasteurised dairy products</td>
</tr>
<tr>
<td>Group B streptococcal infection</td>
<td>Neonatal meningitis and sepsis. Sepsis in mother after delivery</td>
<td>Risk- or screening-based antimicrobial prophylaxis in labour (recommendations vary between countries)</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
<td>Congenital malformation</td>
<td>Diagnosis and prompt treatment of cases, avoidance of undercooked meat while pregnant</td>
</tr>
<tr>
<td>Malaria</td>
<td>Fetal loss, intrauterine growth retardation, severe malaria in mother</td>
<td>Avoidance of insect bites. Intermittent preventative treatment during pregnancy to decrease incidence in high-risk countries</td>
</tr>
</tbody>
</table>

### Viral Infections

#### Systemic viral infections with exanthem

Childhood exanthems are characterised by fever and widespread rash. Maternal antibody gives protection for the first 6–12 months of life and infection occurs thereafter. Comprehensive immunisation programmes have dramatically reduced the number of paediatric infections but incomplete uptake results in infections in later life.

**Measles**

The WHO has set the objective of eradicating measles globally using the live attenuated vaccine. However, vaccination of more than 95% of the population is required to prevent outbreaks. Natural illness produces life-long immunity.

**Clinical features**

Infection is by respiratory droplets with an incubation period of 6–19 days. A prodromal illness, 1–3 days before the rash, occurs, with upper respiratory symptoms, conjunctivitis and the presence of the pathognomonic Koplik’s spots, small white spots surrounded by erythema on the buccal mucosa (Fig. 13.7A). As natural antibody develops, the maculopapular rash appears, spreading from the face to the extremities (Fig. 13.7B). Generalised lymphadenopathy and diarrhoea are common. Complications are more common in older children and adults, and include otitis media, bacterial pneumonia, transient hepatitis and clinical encephalitis (approximately 0.1% of cases). A rare late complication is subacute sclerosing panencephalitis (SSPE), which occurs up to 7 years after infection. Diagnosis is clinical (although this has become unreliable in areas where measles is no longer common) and by detection of antibody (serum IgM, seroconversion or salivary IgM).

Measles is a serious disease in the malnourished, vitamin-deficient or immunocompromised, in whom the typical rash may be missing and persistent infection with a giant cell pneumonitis or encephalitis may occur. In tuberculosis infection, measles suppresses cell-mediated immunity and may exacerbate disease; for this reason, measles vaccination should be deferred until after commencing antituberculous treatment. Measles does not cause congenital malformation but may be more severe in pregnant women.
Mortality clusters at the extremes of age, averaging 1:1000 in developed countries and up to 1:4 in developing countries. Death usually results from a bacterial superinfection, occurring as a complication of measles, most often pneumonia, diarrhoeal disease or noma/cancrum oris, a gangrenous stomatitis. Death may also result from complications of measles encephalitis.

**Management and prevention**

Normal immunoglobulin attenuates the disease in the immunocompromised (regardless of vaccination status) and in non-immune pregnant women, but must be given within 6 days of exposure. Vaccination can be used in outbreaks and vitamin A may improve the outcome in uncomplicated disease. Antibiotic therapy is reserved for bacterial complications. All children aged 12–15 months should receive measles vaccination (as combined measles, mumps and rubella (MMR), a live attenuated vaccine), and a further MMR dose at age 4 years.

**Rubella (German measles)**

Rubella causes exanthem in the non-immunised.

**Clinical features**

Rubella is spread by respiratory droplet, with infectivity from up to 10 days before to 2 weeks after the onset of the rash. The incubation period is 15–20 days. In childhood, most cases are subclinical, although clinical features may include fever, maculopapular rash spreading from the face, and lymphadenopathy. Complications are rare but include thrombocytopenia and hepatitis. Encephalitis and haemorrhage are occasionally reported. In adults, arthritis involving hands or knees is relatively common, especially in women.

If transplacental infection takes place in the first trimester or later, persistence of the virus is likely and severe congenital disease may result (Box 13.29). Even if normal at birth, the infant has an increased incidence of other diseases developing later, e.g. diabetes mellitus.

**Diagnosis**

Laboratory confirmation of rubella is required if there has been contact with a pregnant woman. This is achieved either by detection of rubella IgM in serum or by IgG seroconversion. In the exposed pregnant woman, absence of rubella-specific IgG confirms the potential for congenital infection.

**Prevention**

All children should be immunised with MMR, as above for measles. In view of the risks of congenital rubella syndrome, all women of child-bearing age should also be tested for rubella and vaccinated if seronegative.

**Parvovirus B19**

Parvovirus B19 causes exanthem and other clinical syndromes. Some 50% of children and 60–90% of adults are seropositive. Most infections are spread by the respiratory route, although spread via contaminated blood is also possible. The virus has particular tropism for red cell precursors.

**Clinical features**

Many infections are subclinical. Clinical manifestations result after an incubation period of 14–21 days (Box 13.30). The classic exanthem (erythema infectiosum) is preceded by a prodromal fever and coryzal symptoms. A ‘slapped cheek’ rash is characteristic but the rash is very variable (Fig. 13.8). In adults, polyarthropathy is common. Infected individuals have a transient block in erythropoiesis for a few days, which is of no clinical consequence, except in individuals with increased red cell turnover due to haemoglobinopathy or haemolytic anaemia. These individuals develop an acute anaemia which may be severe (transient aplastic crisis; p. 1032). Erythropoiesis usually recovers spontaneously after 10–14 days. Immunocompromised individuals,
including those with congenital immunodeficiency or AIDS, can develop a more sustained block in erythropoiesis in response to the chronic viraemia that results from their inability to clear the infection. Infection during the first two trimesters of pregnancy can result in intrauterine infection and impact on fetal bone marrow; it causes 10–15% of non-immune (non-Rhesus-related) hydrops fetalis, a rare complication of pregnancy.

**Diagnosis**

IgM to parvovirus B19 suggests recent infection but may persist for months and false positives occur. Seroconversion to IgG positivity confirms infection but in isolation a positive IgG is of little diagnostic utility. Detection of parvovirus B19 DNA in blood is particularly useful in immunocompromised patients. Giant pronormoblasts or haemophagocytosis may be demonstrable in the bone marrow.

**Management**

Infection is usually self-limiting. Symptomatic relief for arthritic symptoms may be required. Severe anaemia requires transfusion. Persistent viraemia in immunocompromised hosts may require immunoglobulin therapy to clear the virus.

Pregnant women should avoid contact with cases of parvovirus B19 infection; if they are exposed, serology should be performed to establish whether they are non-immune.

Passive prophylaxis with normal immunoglobulin has been suggested for non-immune pregnant women exposed to infection but there are limited data to support this recommendation. The pregnancy should be closely monitored by ultrasound scanning, so that hydrops fetalis can be treated by fetal transfusion.

**Human herpesvirus 6 and 7**

Human herpesvirus 6 (HHV-6) is a lymphotrophic virus that causes a childhood viral exanthem (exanthem subitum), rare cases of an infectious mononucleosis-like syndrome and infection in the immunocompromised host. Infection is almost universal, with approximately 95% of children acquiring this virus by 2 years of age. Transmission is via saliva.

HHV-7 is very closely related to HHV-6, and is believed to be responsible for a proportion of cases of exanthem subitum. Like HHV-6, HHV-7 causes an almost universal infection in childhood, with subsequent latent infection and occasional infection in the immunocompromised host.

**Clinical features**

Exanthem subitum is also known as roseola infantum or sixth disease (Box 13.31). A high fever is followed by a maculopapular rash as the fever resolves. Fever and/or febrile convulsions may also occur without a rash. Rarely, older children or adults may develop an infectious mononucleosis-like illness, hepatitis or rash. In the immunocompromised, infection is rare but can cause fever, rash, hepatitis, pneumonitis, cytopenia or encephalitis.

**Diagnosis and management**

Exanthem subitum is usually a clinical diagnosis but can be confirmed by antibody and/or DNA detection. The disease is self-limiting. Treatment with ganciclovir or foscarnet is used in immunocompromised hosts infected with HHV-6.

**Chickenpox (varicella)**

Varicella zoster virus (VZV) is a dermatropic and neurotropic virus that produces primary infection, usually in childhood, which may reactivate in later life. VZV is spread by aerosol and direct contact. It is highly
infectious to non-immune individuals. Disease in children is usually well tolerated. Manifestations are more severe in adults, pregnant women and the immunocompromised.

**Clinical features**

The incubation period is 11–20 days, after which a vesicular eruption begins (Fig. 13.9), often on mucosal surfaces first, followed by rapid dissemination in a centrifugal distribution (most dense on trunk and sparse on limbs). New lesions occur every 2–4 days and each crop is associated with fever. The rash progresses from small pink macules to vesicles and pustules within 24 hours. Infectivity lasts from up to 4 days (but usually 48 hours) before the lesions appear until the last vesicles crust over. Due to intense itching, secondary bacterial infection from scratching is the most common complication of primary chickenpox. Self-limiting cerebellar ataxia and encephalitis are rare complications.

Adults, pregnant women and the immunocompromised are at increased risk of visceral involvement, which presents as pneumonitis, hepatitis or encephalitis. Pneumonitis can be fatal and is more likely to occur in smokers. Maternal infection in early pregnancy carries a 3% risk of neonatal damage with developmental abnormalities of eyes, CNS and limbs. Chickenpox within 5 days of delivery leads to severe neonatal varicella with visceral involvement and haemorrhage.

**Diagnosis**

Diagnosis is primarily clinical, by recognition of the rash. If necessary, this can be confirmed by detection of antigen (direct immunofluorescence) or DNA (PCR) of aspirated vesicular fluid. Serology is used to identify seronegative individuals at risk of infection.

**Management and prevention**

The benefits of antivirals for uncomplicated primary VZV infection in children are marginal and treatment is not required (Box 13.32). Antivirals are, however, used for uncomplicated chickenpox when the patient presents within 24–48 hours of onset of vesicles, in all patients with complications, and in those who are immunocompromised, including pregnant women, regardless of duration of vesicles (Box 13.33). More severe disease, particularly in immunocompromised hosts, requires initial parenteral therapy. Immunocompromised patients may have prolonged viral shedding and may require prolonged treatment until all lesions crust over.

**EBM 13.32 Aciclovir for chickenpox/shingles**

‘Aciclovir shortens symptoms in chickenpox by an average of 1 day. In shingles, aciclovir reduces pain by 10 days and the risk of post-herpetic neuralgia by 8%. Aciclovir is therefore cost-effective in shingles but not chickenpox.’

Human VZ immunoglobulin (VZIG) is used to attenuate infection in people who have had significant contact with VZV, are susceptible to infection (i.e. have no history of chickenpox or shingles and are seronegative for VZV IgG) and are at risk of severe disease (e.g. immunocompromised, steroid-treated or pregnant).

### 13.33 Therapy for herpes simplex and varicella zoster virus infection

<table>
<thead>
<tr>
<th>Disease state</th>
<th>Treatment options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary genital HSV</td>
<td>Famciclovir 250 mg 3 times daily for 7–10 days Valaciclovir 1 g twice daily for 7–10 days Oral aciclovir 200 mg 5 times daily or 400 mg 3 times daily for 7–10 days</td>
</tr>
<tr>
<td>Severe and preventing oral intake</td>
<td>Aciclovir 5 mg/kg 3 times daily IV until patient can tolerate oral therapy</td>
</tr>
<tr>
<td>Recurrent genital HSV-1 or 2</td>
<td>Oral aciclovir 200 mg 5 times daily or 400 mg 3 times daily for 5 days Famciclovir 125 mg twice daily for 5 days Valaciclovir 500 mg twice daily for 3–5 days or 2 g twice daily for 1 day. Shorter durations increasingly favoured</td>
</tr>
<tr>
<td>Primary or recurrent oral HSV</td>
<td>Usually no treatment If required, usually short duration, e.g. valaciclovir 2 g twice daily for 1 day</td>
</tr>
<tr>
<td>Mucocutaneous HSV infection in immunocompromised host</td>
<td>Aciclovir 5 mg/kg 3 times daily IV for 7–10 days Oral aciclovir 400 mg 4 times daily for 7–10 days Famciclovir 500 mg 3 times daily for 7–10 days Valaciclovir 1 g twice daily for 7–10 days</td>
</tr>
<tr>
<td>Chickenpox in adult or child</td>
<td>Oral aciclovir 800 mg 5 times daily for 5 days Famciclovir 500 mg 3 times daily for 5 days Valaciclovir 1 g twice daily for 5 days</td>
</tr>
<tr>
<td>Immunocompromised host/pregnant woman</td>
<td>Aciclovir 5 mg/kg 3 times daily IV until patient is improving, then complete therapy with oral therapy until all lesions crusting over</td>
</tr>
<tr>
<td>Shingles</td>
<td>Treatment and doses as for chickenpox but duration typically 7–10 days</td>
</tr>
<tr>
<td>Visceral involvement (non-CNS) in HSV</td>
<td>Aciclovir IV 5 mg/kg 3 times daily for 14 days</td>
</tr>
<tr>
<td>Visceral involvement (non-CNS) in VZV</td>
<td>Aciclovir IV 5 mg/kg 3 times daily for 7 days</td>
</tr>
<tr>
<td>Severe complications (encephalitis, disseminated infection)</td>
<td>Aciclovir IV 10 mg/kg 3 times daily (up to 20 mg/kg in neonates) for 14–21 days</td>
</tr>
<tr>
<td>HSV disease suppression</td>
<td>Aciclovir 400 mg twice daily Famciclovir 250 mg twice daily Valaciclovir 500 mg daily</td>
</tr>
</tbody>
</table>

### 13.34 Indications for varicella zoster immunoglobulin (VZIG) in adults

An adult should satisfy all three of the following conditions:

1. **Significant contact**
   - Contact with chickenpox (any time from 48 hrs before the rash until crusting of lesions) or zoster (exposed, disseminated or, with immunocompromised contacts, localised zoster; between development of the rash until crusting) defined as:
     - Prolonged household contact, sharing a room for ≥ 15 mins or face-to-face contact (includes direct contact with zoster lesions)
     - Hospital contact with chickenpox in another patient, health-care worker or visitor
     - Intimate contact (e.g. touching) with person with shingles lesions
     - Newborn whose mother develops chickenpox no more than 5 days before delivery or 2 days after delivery

2. **Susceptible contact**
   - Individual with no history of chickenpox, ideally confirmed by negative test for VZV IgG

3. **Predisposition to severe chickenpox**
   - Immunocompromised due to disease (e.g. acute leukaemia, HIV, other primary or secondary immunodeficiency)
   - Medically immuno suppressed (e.g. following solid organ transplant; current or recent (< 6 mths) cytotoxic chemotherapy or radiotherapy; current or recent (< 3 mths) high-dose corticosteroids; haematopoietic stem cell transplant)
   - Pregnant (any stage)
   - Infants: newborn whose mother has had chickenpox as above; premature infants < 28 wks

(Box 13.34). Ideally, VZIG should be given within 7 days of exposure, but it may attenuate disease even if given up to 10 days afterwards. Susceptible contacts who develop severe chickenpox after receiving VZIG should be treated with aciclovir.

A live, attenuated VZV vaccine is available and in routine use in the USA and other countries, but in the UK its use has been restricted to non-immune healthcare workers and household contacts of immunocompromised individuals. Children receive one dose after 1 year of age and a second dose at 4–6 years of age; seronegative adults receive two doses at least 1 month apart. The vaccine may also be used prior to planned iatrogenic immunosuppression, e.g. before transplant.

**Shingles (herpes zoster)**

After initial infection, VZV persists in latent form in the dorsal root ganglion of sensory nerves and can reactivate in later life.

**Clinical features**

Burning discomfort occurs in the affected dermatome, where discrete vesicles appear 3–4 days later. This is associated with a brief viraemia, which can produce distant satellite ‘chickenpox’ lesions. Occasionally, paraesthesia occurs without rash (‘zoster sine herpete’). Severe disease, a prolonged duration of rash, multiple dermal involvement or recurrence suggests underlying immune deficiency, including HIV. Chickenpox may be contracted from a case of shingles but not vice versa.
Although thoracic dermatomes are most commonly involved (Fig. 13.9B), the ophthalmic division of the trigeminal nerve is also frequently affected; vesicles may appear on the cornea and lead to ulceration. This condition can lead to blindness and urgent ophthalmology review is required. Geniculate ganglion involvement causes the Ramsay Hunt syndrome of facial palsy, ipsilateral loss of taste and buccal ulceration, plus a rash in the external auditory canal. This may be mistaken for Bell’s palsy (p. 1163). Bowel and bladder dysfunction occur with sacral nerve root involvement. The virus occasionally causes cranial nerve palsy, myelitis or encephalitis. Granulomatous cerebral angiitis is a cerebrovascular complication that leads to a stroke-like syndrome in association with shingles, especially in an ophthalmic distribution.

Post-herpetic neuralgia causes troublesome persistence of pain for 1–6 months or longer, following healing of the rash. It is more common with advanced age.

Management

Early therapy with aciclovir or related agents has been shown to reduce both early- and late-onset pain, especially in patients over 65 years (see Box 13.32). Post-herpetic neuralgia requires aggressive analgesia, along with agents such as amitriptyline 25–100 mg daily or gabapentin (commencing at 300 mg daily and building slowly to 300 mg twice daily or more). Capsaicin cream (0.075%) may be helpful. Although controversial, corticosteroids have not been demonstrated to reduce post-herpetic neuralgia to date.

Enteroviral exanthems

Coxackie or echovirus infections can lead to a maculopapular eruption or roseola-like rash that occurs after fever falls. Enteroviral infections are discussed further under viral infections of the skin (see below).

Systemic viral infections without exanthem

Other systemic viral infections present with features other than a rash suggestive of exanthem. Rashes may occur in these conditions but differ from those seen in exanthems or are not the primary presenting feature.

Mumps

Mumps is a systemic viral infection characterised by swelling of the parotid glands. Infection is endemic worldwide and peaks at 5–9 years of age. Vaccination has reduced the incidence in children but incomplete coverage and waning immunity with time have led to outbreaks in young adults. Infection is spread by respiratory droplets.

Clinical features

The median incubation period is 19 days, with a range of 15–24 days. Classical tender parotid enlargement, which is bilateral in 75%, follows a prorome of pyrexia and headache (Fig. 13.10). Meningitis complicates up to 10% of cases. The CSF reveals a lymphocytic pleocytosis or, less commonly, neutrophils. Rare complications include encephalitis, transient hearing loss, labyrinthitis, electrocardiographic abnormalities, pancreatitis and arthritis.

Approximately 25% of post-pubertal males with mumps develop epididymo-orchitis but, although testicular atrophy occurs, sterility is unlikely. Oophoritis is less common. Abortion may occur if infection takes place in the first trimester of pregnancy. Complications may occur in the absence of parotitis.

Diagnosis

The diagnosis is usually clinical. In atypical presentations without parotitis, serology for mumps-specific IgM or IgG seroconversion (four-fold rise in IgG convalescent titre) confirms the diagnosis. Virus can also be cultured from urine in the first week of infection or detected by PCR in urine, saliva or CSF.

Management and prevention

Treatment is with analgesia. There is no evidence that corticosteroids are of value for orchitis. Mumps vaccine is one of the components of the combined MMR vaccine.

Influenza

Influenza is an acute systemic viral infection that primarily affects the respiratory tract; it carries a significant mortality. It is caused by influenza A virus or, in milder form, influenza B virus. Infection is seasonal, and variation in the haemagglutinin (H) and neuraminidase (N) glycoproteins on the surface of the virus leads to disease of variable intensity each year. Minor changes in haemagglutinin are known as ‘genetic drift’, whereas a switch in the haemagglutinin or neuraminidase antigen is termed ‘genetic shift’. Nomenclature of influenza strains is based on these glycoproteins, e.g. H1N1, H3N2 etc. Genetic shift results in the circulation of a new influenza strain within a community to which few people are immune, potentially initiating an influenza epidemic or pandemic in which there is a high attack rate and there may be increased disease severity.

Clinical features

After an incubation period of 1–3 days, uncomplicated disease leads to fever, malaise and cough. Viral pneumonia may occur, although pulmonary complications
are most often due to superinfection with *Strep. pneumoniae*, *Staph. aureus* or other bacteria. Rare extrapulmonary manifestations include myositis, myocarditis, pericarditis and neurological complications (Reye’s syndrome in children, encephalitis or transverse myelitis).

Mortality is greatest in the elderly, those with medical comorbidities and pregnant women. Recently, polymorphisms in the gene encoding an antiviral protein, interferon-induced transmembrane protein 3 (IFITM3), have been associated with more severe influenza.

**Diagnosis**

Acute infection is diagnosed by viral antigen or RNA detection in a nasopharyngeal sample. The disease may also be diagnosed retrospectively by serology.

**Management and prevention**

Management involves early microbiological identification of cases and good infection control, with an emphasis on hand hygiene and preventing dissemination of infection by coughing and sneezing. Administration of neuraminidase inhibitor, oral oseltamivir (75 mg twice daily) or inhaled zanamivir (10 mg twice daily) for 5 days, can reduce the severity of symptoms if started within 48 hours of symptom onset (or possibly later in immunocompromised individuals). These agents have superseded routine use of amantadine and rimantadine. Antiviral drugs can also be used as prophylaxis in high-risk individuals during the ‘flu’ season. Resistance can emerge to all of these agents and so updated local advice should be followed.

Prevention relies on seasonal vaccination of the elderly and of individuals with chronic medical illnesses which place them at increased risk of the complications of influenza, such as chronic cardiopulmonary diseases or immune compromise, as well as their health-care workers. The vaccine composition changes each year to cover the ‘predicted’ seasonal strains but vaccination may fail when a new pandemic strain emerges.

**Avian influenza**

Avian influenza is caused by transmission of avian influenza A viruses to humans. Avian viruses, such as H5N1, possess alternative haemagglutinin antigens to seasonal influenza viruses. Most cases have had contact with sick poultry, predominantly in South-east Asia, and person-to-person spread has been limited to date. Infections with H5N1 viruses have been severe, with enteric features and respiratory failure. Treatment depends on the resistance pattern but often involves oseltamivir. Vaccination against seasonal ‘flu’ does not adequately protect against avian influenza. There is a concern that adaptation of an avian strain to allow effective person-to-person transmission is likely to lead to a global pandemic of life-threatening influenza.

**Swine influenza**

Occasional cases of influenza are transmitted from pigs to humans. Re-association of swine, avian and human influenza strains can occur in pigs. Sometimes this can lead to an outbreak of swine ‘flu’ in humans, as occurred in 2009, when an outbreak of H1N1 influenza spread around the world from Mexico.

### Infectious mononucleosis and Epstein–Barr virus

Infectious mononucleosis (IM) is a clinical syndrome characterised by pharyngitis, cervical lymphadenopathy, fever and lymphocytosis. It is most often caused by Epstein–Barr virus (EBV) but other infections can produce a similar clinical syndrome (Box 13.35).

EBV is a gamma herpesvirus. In developing countries, subclinical infection in childhood is virtually universal. In developed countries, primary infection may be delayed until adolescence or early adult life. Under these circumstances, about 50% of infections result in typical IM. The virus is usually acquired from asymptomatic excreters via saliva, either by droplet infection or environmental contamination in childhood, or by kissing among adolescents and adults. EBV is not highly contagious and isolation of cases is unnecessary.

**Clinical features**

EBV infection has a prolonged and undetermined incubation period, followed in some cases by a prodrome of fever, headache and malaise. This is succeeded by IM with severe pharyngitis, which may include tonsillar exudates and non-tender anterior and posterior cervical lymphadenopathy. Palatal petechiae, periobital oedema, splenomegaly, inguinal or axillary lymphadenopathy, and macular, petechial or erythema multiforme rashes may occur. In most cases, fever resolves over 2 weeks, and fatigue and other abnormalities settle over a further few weeks. Complications are listed in Box 13.36. Death is rare but can occur due to respiratory obstruction, haemorrhage from splenic rupture or thrombocytopenia, or encephalitis.

The diagnosis of EBV infection outside the usual age in adolescence and young adulthood is more challenging. In children under 10 years the illness is mild and short-lived, but in adults over 30 years of age it can be severe and prolonged. In both groups, pharyngeal symptoms are often absent. EBV may present with jaundice, as a PUO or with a complication.

**Long-term complications of EBV infection**

Lymphoma complicates EBV infection in immunocompromised hosts, and some forms of Hodgkin’s disease are EBV-associated (p. 1042). The endemic form of Burkitt’s lymphoma complicates EBV infection in areas of sub-Saharan Africa where *falciparum* malaria is endemic. Nasopharyngeal carcinoma is a geographically restricted tumour seen in China and Alaska that is associated with EBV infection. X-linked lymphoproliferative (Duncan’s) syndrome is a familial lymphoproliferative disorder that follows primary EBV infection in boys without any other history of immunodeficiency; it is due to mutation of the SAP gene, causing failure of T-cell and NK-cell activation and inability to contain EBV infection.
Complications of Epstein–Barr virus 

**Common**
- Severe pharyngeal oedema
- Antibiotic-induced rash (80–96% with ampicillin)
- Hepatitis (80%)
- Prolonged post-viral fatigue (10%)  
  Jaundice (< 10%)

**Uncommon**
- Cranial nerve palsies
- Polynuertis
- Haemolytic anaemia
- Transverse myelitis
- Meningoencephalitis
- Thrombocytopenia
- Abnormalities on urinalysis
- Interstitial nephritis
- Myocarditis
- Pericarditis
- ECG abnormalities

**Rare**
- Ruptured spleen
- Respiratory obstruction
- Agranulocytosis
- X-linked lymphoproliferative syndrome
- Nasopharyngeal carcinoma
- Burkitt’s lymphoma
- Hodgkin’s disease (certain subtypes only)
- Primary CNS lymphoma
- Lymphoproliferative disease in immunocompromised

**EBV-associated malignancy**

Investigations

Atypical lymphocytes are common in EBV infection but also occur in other causes of IM, acute retroviral syndrome with HIV infection, viral hepatitis, mumps and rubella (Fig. 13.11A). A ‘heterophile’ antibody is present during the acute illness and convalescence, which is detected by the Paul–Bunnell or ‘Monospot’ test. Sometimes antibody production is delayed, so an initially negative test should be repeated. However, many children and 10% of adolescents with IM do not produce heterophile antibody at any stage.

Specific EBV serology confirms the diagnosis. Acute infection is characterised by IgM antibodies against the viral capsid, antibodies to EBV early antigen and the initial absence of antibodies to EBV nuclear antigen (anti-EBNA). Seroconversion of anti-EBNA at approximately 1 month after the initial illness may confirm the diagnosis in retrospect. CNS infections may be diagnosed by detection of viral DNA in cerebrospinal fluid.

**Management**

Treatment is largely symptomatic. If a throat culture yields a β-haemolytic streptococcus, penicillin should be given. Administration of ampicillin or amoxicillin in this condition commonly causes an itchy macular rash and should be avoided (Fig. 13.11B). When pharyngeal oedema is severe, a short course of corticosteroids, e.g. prednisolone 30 mg daily for 5 days, may help. Current antiviral drugs are not active against EBV.

Return to work or school is governed by physical fitness rather than laboratory tests; contact sports should be avoided until splenomegaly has resolved because of the danger of splenic rupture. Unfortunately, about 10% of patients with IM suffer a chronic relapsing syndrome.

**Cytomegalovirus**

Cytomegalovirus (CMV), like EBV, circulates readily among children. A second period of virus acquisition occurs among teenagers and young adults, peaking between the ages of 25 and 35 years, rather later than with EBV infection. CMV infection is persistent, and is characterised by subclinical cycles of active virus replication and by persistent low-level virus shedding. Most post-childhood infections are therefore acquired from asymptomatic excreters who shed virus in saliva, urine, semen and genital secretions. Sexual transmission and oral spread are common among adults, but infection may also be acquired by women caring for children with asymptomatic infections.

**Clinical features**

Most post-childhood CMV infections are subclinical, although some young adults develop an IM-like syndrome and some have a prolonged influenza-like illness lasting 2 weeks or more. Physical signs resemble those of IM, but in CMV infections hepatomegaly is more common, while lymphadenopathy, splenomegaly, pharyngitis and tonsillitis occur less often. Jaundice is uncommon and usually mild. Complications include meningoencephalitis, Guillain–Barré syndrome, autoimmune haemolytic anaemia, thrombocytopenia, myocarditis and skin eruptions, such as ampicillin-induced rash. Immunocompromised patients can develop hepatitis, oesophagitis, colitis, pneumonitis, retinitis, encephalitis and polyradiculitis.
**Infectious Disease**

Women who develop a primary CMV infection during pregnancy have about a 40% chance of passing CMV to the fetus, causing congenital infection and disease at any stage of gestation. Features include petechial rashes, hepatosplenomegaly and jaundice; 10% of infected infants will have long-term CNS sequelae, such as microcephaly, cerebral calcifications, chorioretinitis and deafness. Infections in the newborn usually are asymptomatic or have features of an IM-like illness, although some studies suggest that subtle sequelae affecting hearing or mental development may occur.

**Investigations**

Atypical lymphocytosis is not as prominent as in EBV infection and heterophile antibody tests are negative. LFTs are often abnormal, with an alkaline phosphatase level raised out of proportion to transaminases. Serological diagnosis depends on the detection of CMV-specific IgM antibody plus a four-fold rise or seroconversion of IgG. In the immunocompromised, antibody detection is unreliable and detection of CMV in an involved organ by PCR, culture or histopathology establishes the diagnosis. A positive culture of CMV in the blood may be useful in transplant populations but not in HIV-positive individuals, since in HIV infection CMV reactivates at regular intervals, but these episodes do not correlate well with episodes of clinical disease. Detection of CMV in urine is not helpful in diagnosing infection, except in neonates, since CMV is intermittently shed in the urine throughout life following infection.

**Management**

Only symptomatic treatment is required in the immunocompetent patient. Immunosuppressed individuals are treated with ganciclovir 5 mg/kg IV twice daily or with oral valganciclovir 900 mg twice daily for at least 14 days. Foscarnet or cidofovir is also used in CMV treatment of immunocompromised patients who are resistant to or intolerant of ganciclovir-based therapy.

**Dengue**

Dengue is a febrile illness caused by a flavivirus transmitted by mosquitoes. It is endemic in Asia, the Pacific, Africa and the Americas (Fig. 13.12). Approximately 50 million infections occur annually and dengue is the most rapidly spreading mosquito-borne viral illness. The principal vector is the mosquito *Aedes aegypti*, which breeds in standing water; collections of water in containers, water-based air coolers and tyre dumps are a good environment for the vector in large cities. *Aedes albopictus* is a vector in some South-east Asian countries. There are four serotypes of dengue virus, all producing a similar clinical syndrome; type-specific immunity is life-long, but immunity against the other serotypes lasts only a few months. Dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS) occur in individuals who are immune to one dengue virus serotype and are then infected with another. Prior immunity results in increased uptake of virus by cells expressing the antibody Fc receptor and increased T-cell activation with resultant cytokine release, causing capillary leak and disseminated intravascular coagulation (DIC, pp. 201 and 1055).

**Clinical features**

Clinical features of dengue fever are listed in Box 13.37. Asymptomatic infections are common, particularly in children, but the disease is more severe in infants and the elderly. The initial febrile phase is frequently followed by a rash as the fever settles. Laboratory features include leucopenia, neutropenia, thrombocytopenia and elevated alanine aminotransferase (ALT) or aspartate aminotransferase (AST). Many symptomatic infections run an uncomplicated course, but complications or a protracted convalescence may ensue.

### Box 13.37 Clinical features of dengue fever

**Incubation period**
- 2–7 days

**Prodrome**
- 2 days of malaise and headache

**Acute onset**
- Fever, backache, arthralgia, headache, generalised pains (‘break-bone fever’), pain on eye movement, lacrimation, scleral injection, anorexia, nausea, vomiting, pharyngitis, upper respiratory tract symptoms, relative bradycardia, prostration, depression, hyperaesthesia, dysgeusia, lymphadenopathy

**Fever**
- Continuous or ‘saddle-back’, with break on 4th or 5th day and then defervescence; usually lasts 7–8 days

**Rash**
- Initial flushing faint macular rash in first 1–2 days. Maculopapular, scarlet morbilliform blanching rash from days 3–5 on trunk, spreading centrifugally and sparing palms and soles, onset often with fever defervescence. May desquamate on resolution or give rise to petechiae on extensor surfaces

**Convalescence**
- Slow and may be associated with prolonged fatigue syndrome, arthralgia or depression

**Complications**
- Dengue haemorrhagic fever and disseminated intravascular coagulation
- Dengue shock syndrome
- Hepatitis, cerebral haemorrhage or oedema, encephalitis, cranial nerve palsies, rhabdomyolysis, myocarditis
- Vertical transmission if infection within 5 wks of delivery
The period 3–7 days after onset of fever is termed the ‘critical’ phase, during which signs of DHF or DSS may develop. In mild forms, petechiae occur in the arm when a blood pressure cuff is inflated to a point between systolic and diastolic blood pressure and left for 5 minutes (the positive ‘tourniquet test’) – a non-specific test of capillary fragility and thrombocytopenia. As the extent of capillary leak increases, there may be a raised haematocrit, tachycardia and hypotension, pleural effusions and ascites. This may progress to metabolic acidosis and multi-organ failure, including acute respiratory distress syndrome (ARDS, p. 192). Minor (petechiae, ecchymoses, epistaxis) or major (gastrointestinal or cerebrovascular) haemorrhage may occur.

### Diagnosis

In endemic areas, mild dengue must be distinguished from other viral infections. The WHO recently revised its clinical classification of dengue and is evaluating the usefulness of these categories in guiding diagnosis and treatment (Box 13.38). The diagnosis can be confirmed by seroconversion of IgM or a fourfold rise in IgG antibody titres. Serological tests may detect cross-reacting antibodies against other flaviviruses, including yellow fever vaccine. IgM/IgG ratios may be used to distinguish primary from secondary infection. Isolation of dengue virus from blood or detection of dengue virus RNA by PCR (p. 139) is available in specialist laboratories. Commercial enzyme-linked immunosorbent assay (ELISA) kits to detect the NS1 viral antigen, although less sensitive than PCR, are becoming more widely available in endemic areas.

### Management and prevention

Treatment is supportive, emphasising fluid replacement and appropriate management of shock and organ dysfunction. With intensive care support, mortality rates are 1% or less. Aspirin should be avoided due to bleeding risk. Corticosteroids have not been shown to help. No existing antivirals are effective.

Breeding places of Aedes mosquitoes should be abolished and the adults destroyed by insecticides. There is no licensed vaccine available.

### Yellow fever

Yellow fever is a haemorrhagic fever of the tropics, caused by a flavivirus. It is a zoonosis of monkeys in West and Central African, and South and Central American tropical rainforests, where it may cause devastating epidemics (see Fig. 13.12). Transmission is by tree-top mosquitoes *Aedes africanus* (Africa) and *Haemagogus* spp. (America). The infection is introduced to humans either by infected mosquitoes when trees are felled, or by monkeys raiding human settlements. In towns, yellow fever may be transmitted between humans by *Aedes aegypti*, which breeds efficiently in small collections of water. The distribution of this mosquito is far wider than that of yellow fever, and more widespread infection is a continued threat.

Yellow fever causes approximately 200,000 infections each year, mainly in sub-Saharan Africa, and the number is increasing. Overall mortality is around 15%, although this varies widely. Humans are infectious during the viraemic phase, which starts 3–6 days after the bite of the infected mosquito and lasts for 4–5 days.

### Clinical features

After an incubation period of 3–6 days, yellow fever is often a mild febrile illness lasting less than 1 week, with headache, myalgia, conjunctival erythema and bradycardia. This is followed by fever resolution (defervescence), but in some cases, fever recurs after a few hours to days. In more severe disease, fever recrudescence is associated with lower back pain, abdominal pain and somnolence, prominent nausea and vomiting, bradycardia and jaundice. Liver damage and DIC lead to bleeding with petechiae, mucosal haemorrhages and gastrointestinal bleeding. Shock, hepatic failure, renal failure, seizures and coma may ensue.
**INFECTION DISEASE**

**Diagnosis**

The differential diagnosis includes malaria, typhoid, viral hepatitis, leptospirosis, haemorrhagic fevers and aflatoxin poisoning. Diagnosis of yellow fever can be confirmed by viral isolation from blood in the first 24 days of illness, the presence of IgM or a fourfold rise in IgG antibody titre. Leucopenia is characteristic. Liver biopsy should be avoided in life due to the risk of fatal bleeding. Post-mortem features, such as acute mid-zonal necrosis and Councilman bodies with minimal inflammation in the liver, are suggestive but not specific. Immunohistochemistry for viral antigens improves specificity.

**Management and prevention**

Treatment is supportive, with meticulous attention to fluid and electrolyte balance, urine output and blood pressure. Blood transfusions, plasma expanders and peritoneal dialysis may be necessary. Patients should be isolated, as their blood and body products may contain virus particles.

A single vaccination with a live attenuated vaccine gives full protection for at least 10 years. Potential side-effects include hypersensitivity, encephalitis and systemic features of yellow fever (viscerotrophic disease) caused by the attenuated virus. Vaccination is not recommended in people who are significantly immunosuppressed. The risk of vaccine side-effects must be balanced against the risk of infection for less immunocompromised hosts, pregnant women and older patients. An internationally recognised certificate of vaccination is sometimes necessary when crossing borders.

**Viral haemorrhagic fevers**

Viral haemorrhagic fevers (VHF) are zoonoses caused by several different viruses (Box 13.39). They are geographically restricted and occur in rural settings or in health-care facilities. All of these viral illnesses, except Ebola and Marburg, have mildest self-limiting forms.

Serological surveys have shown that Lassa fever is widespread in West Africa and may lead to up to 500 000 infections annually. Mortality overall may be low, as 80% of cases are asymptomatic, but in hospitalised cases mortality averages 15%. Ebola outbreaks have occurred at a rate of approximately one per year, involving up to a few hundred cases. The largest outbreaks have been in the Democratic Republic of Congo, Uganda and Sudan. Marburg has been documented less frequently, with outbreaks in the Democratic Republic of Congo and Uganda, but the largest outbreak to date involved 163 cases in Angola in 2005. Mortality rates of Ebola and Marburg are high.

VHF have extended into Europe, with an outbreak of Congo–Crimean haemorrhagic fever in Turkey in 2006, and cases of haemorrhagic fever with renal syndrome in the Balkans and Russia. These conditions remain very rare in the UK, with about one case of Lassa fever arriving in the country every 2 years.

Kyasanur forest disease is a tick-borne VHF currently confined to a small focus in Karnataka, India; there are about 500 cases annually. Monkeys are the principal hosts, but with forest felling, there are fears that this disease will increase.

New outbreaks and new agents are identified sporadically. In 2008, Lujo virus, a novel arenavirus, caused an outbreak of VHF involving a woman from Zambia and several health-care workers associated with her care, and had 80% fatality. Details on recent disease outbreaks can be found at the WHO website (www.who.int).

**Clinical features**

VHF present with non-specific fever, malaise, body pains, sore throat and headache. On examination, conjunctivitis, throat injection, an erythematous or petechial rash, haemorrhage, lymphadenopathy and bradycardia may be noted. The viruses cause endothelial dysfunction with the development of capillary leak. Bleeding is due to endothelial damage and platelet dysfunction. Hypovolaemic shock and ARDS may develop (p. 192).

Haemorrhage is a late feature of VHF and most patients present with earlier features. In Lassa fever, joint and abdominal pain is prominent. A macular blanching rash may be present but bleeding is unusual, occurring in only 20% of hospitalised patients. Encephalopathy may develop and deafness affects 30% of survivors.

The clue to the viral aetiology comes from the travel and exposure history. Travel to an outbreak area, activity in a rural environment and contact with sick individuals or animals within 21 days all increase the risk of VHF. Enquiry should be made about insect bites, hospital visits and attendance at ritual funerals (Ebola virus infection). For Lassa fever, retrosternal pain, pharyngitis and proteinuria have a positive predictive value of 80% in West Africa.

**Investigations and management**

Non-specific findings include leucopenia, thrombocytopaenia and proteinuria. In Lassa fever, an AST > 150 U/L is associated with a 50% mortality. It is important to exclude other causes of fever, especially malaria, typhoid and respiratory tract infections. Most patients suspected of having a VHF in the UK turn out to have malaria.

The diagnosis of VHF must be considered in all febrile individuals who present within 21 days of leaving an endemic area or who present with haemorrhage or organ failure. A febrile patient from an endemic area within the incubation period, who has specific epidemiological risk factors (see Fig. 13.5, p. 311) or who has signs of organ failure or haemorrhage, should be treated as being at high risk of VHF; appropriate infection control measures must be implemented and the patient transferred to a centre with biosafety level (BSL) 4 facilities. Individuals with a history of travel within 21 days and fever, but without the relevant epidemiological features or signs of VHF, are classified as medium-risk and should have an initial blood sample tested to exclude malaria. If this is negative, relevant specimens (blood, throat swab, urine and pleural fluid, if available) are collected and sent to an appropriate reference laboratory for nucleic acid detection (PCR), virus isolation, and serology. If patients are still felt to be at significant risk of VHF or if infection is confirmed, they should be transferred to a specialised high-security infectious disease unit. All further laboratory tests should be performed at BSL4. Transport requires an ambulance with BSL3 facilities.
### 13.39 Viral haemorrhagic fevers

<table>
<thead>
<tr>
<th>Disease</th>
<th>Reservoir</th>
<th>Transmission</th>
<th>Incubation period</th>
<th>Geography</th>
<th>Mortality rate</th>
<th>Clinical features of severe disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lassa fever</td>
<td>Multimammate rats</td>
<td>Urine from rat</td>
<td>6–21 days</td>
<td>West Africa</td>
<td>15%</td>
<td>Haemorrhage, shock, encephalopathy, ARDS (responds to ribavirin) Deafness in survivors</td>
</tr>
<tr>
<td>Ebola fever</td>
<td>Undefined (?bats)</td>
<td>Body fluids from patients Handling infected primates</td>
<td>2–21 days</td>
<td>Central Africa Outbreak as far north as Sudan</td>
<td>25–90%</td>
<td>Haemorrhage, hepatic and renal failure</td>
</tr>
<tr>
<td>Marburg fever</td>
<td>Undefined</td>
<td>Body fluids from patients Handling infected primates</td>
<td>3–9 days</td>
<td>Central Africa Outbreak in Angola</td>
<td>25–90%</td>
<td>Haemorrhage, diarrhoea, encephalopathy, orchitis</td>
</tr>
<tr>
<td>Yellow fever</td>
<td>Monkeys</td>
<td>Mosquitoes</td>
<td>3–6 days</td>
<td>See Fig. 13.12</td>
<td>~15%</td>
<td>Hepatic failure, renal failure, haemorrhage</td>
</tr>
<tr>
<td>Dengue</td>
<td>Humans</td>
<td>Aedes aegypti</td>
<td>2–7 days</td>
<td>See Fig. 13.12</td>
<td>&lt; 10%²</td>
<td>Haemorrhage, shock</td>
</tr>
<tr>
<td>Crimea–Congo haemorrhagic fever</td>
<td>Small vertebrates Domestic and wild animals</td>
<td>lxooe tick</td>
<td>1–3 days up to 9 days 3–6 days up to 13 days</td>
<td>Africa, Asia, Eastern Europe</td>
<td>30%</td>
<td>Encephalopathy, haemorrhage, hepatic or renal failure, ARDS</td>
</tr>
<tr>
<td>Rift Valley fever</td>
<td>Domestic livestock</td>
<td>Contact with animals, mosquito or other insect bites</td>
<td>2–6 days</td>
<td>Africa, Arabian peninsula</td>
<td>1%</td>
<td>Haemorrhage, blindness, meningoencephalitis (complications only in a minority)</td>
</tr>
<tr>
<td>Kyasanur fever</td>
<td>Monkeys</td>
<td>Ticks</td>
<td>3–8 days</td>
<td>Karnataka State, India</td>
<td>5–10%</td>
<td>Haemorrhage, pulmonary oedema, neurological features, iridokeratitis in survivors</td>
</tr>
<tr>
<td>Bolivian and Argentinian haemorrhagic fever (Junin and Machupo viruses)</td>
<td>Rodents (Calomys spp.)</td>
<td>Urine, aerosols Body fluids from case (rare)</td>
<td>5–19 days (3–6 days for parenteral)</td>
<td>South America</td>
<td>15–30%</td>
<td>Haemorrhage, shock, cerebellar signs (may respond to ribavirin)</td>
</tr>
<tr>
<td>Haemorrhagic fever with renal syndrome (Hantaan fever)</td>
<td>Rodents</td>
<td>Aerosols from faeces</td>
<td>5–42 days (typically 14 days)</td>
<td>Northern Asia, northern Europe, Balkans</td>
<td>5%</td>
<td>Acute renal impairment, cerebrovascular accidents, pulmonary oedema, shock (hepatic failure and haemorrhagic features only in some variants)</td>
</tr>
</tbody>
</table>

¹All potentially have circulatory failure. ²Mortality of uncomplicated and haemorrhagic dengue fever, respectively.

In addition to general supportive measures, ribavirin is given intravenously (100 mg/kg, then 25 mg/kg daily for 3 days and 12.5 mg/kg daily for 4 days) when Lassa fever or South American haemorrhagic fevers are suspected.

**Prevention**

Ribavirin has been used as prophylaxis in close contacts in Lassa fever but there are no formal trials of its efficacy.

### 13. Viral infections of the skin

#### Herpes simplex virus 1 and 2

Herpes simplex viruses (HSV) cause recurrent mucocutaneous infection; HSV-1 typically involves the mucocutaneous surfaces of the head and neck (Fig. 13.13), whilst HSV-2 predominantly involves the genital mucosa (pp. 415 and 418), although there is overlap (see Box 13.31, p. 317). The seroprevalence of HSV-1 is 30–100%, varying by socioeconomic status, while that of HSV-2 is 20–60%. Infection is acquired by inoculation of viruses shed by an infected individual on to a mucosal surface in a susceptible person. The virus infects sensory and autonomic neurons and establishes latent infection in the nerve ganglia. Primary infection is followed by episodes of reactivation throughout life.

**Clinical features**

Primary HSV-1 or 2 infection is more likely to be symptomatic later in life, causing gingivostomatitis, pharyngitis or painful genital tract lesions. The primary
attack may be associated with fever and regional lymphadenopathy.

**Recurrence**

Recurrent attacks occur throughout life, most often in association with concomitant medical illness, menstruation, mechanical trauma, immunosuppression, psychological stress or, for oral lesions, ultraviolet light exposure. HSV reactivation in the oral mucosa produces the classical ‘cold sore’ or ‘herpes labialis’. Prodromal hyperaesthesia is followed by rapid vesiculation, pustulation and crusting. Recurrent HSV genital disease is a common cause of recurrent painful ulceration (pp. 415 and 418). An inoculation lesion on the finger gives rise to a paronychia termed a ‘whitlow’ in contacts of patients with herpetic lesions (Fig. 13.13B). It was formerly seen in health-care workers and dentists, but is prevented by protective gloves.

**Complications**

Disseminated cutaneous lesions can occur in individuals with underlying dermatological diseases, such as eczema (eczema herpeticum) (Fig. 13.13C). Herpes keratitis presents with pain and blurring of vision; characteristic dendritic ulcers are visible on slit-lamp examination and may produce corneal scarring and permanent visual impairment.

Primary HSV-2 can cause meningitis or transverse myelitis. HSV is the leading cause of sporadic viral encephalitis (p. 1205); this serious complication may occur following either primary or secondary disease, usually with HSV-1. A haemorrhagic necrotising temporal lobe cerebritis produces temporal lobe epilepsy and altered consciousness/coma. Without treatment, mortality is 80%. HSV is also implicated in the pathogenesis of Bell’s palsy with a lower motor neuron VII nerve palsy, although antivirals have not been demonstrated to improve outcome.

Neonatal HSV disease is usually associated with primary infection of the mother at term (see Box 13.28, p. 314). In excess of two-thirds of cases develop disseminated disease with cutaneous lesions, hepatitis, pneumonia, and frequently encephalitis.

Immunocompromised hosts can develop visceral disease with oesophagitis, hepatitis, pneumonitis, encephalitis or retinitis.

**Diagnosis**

Differentiation from other vesicular eruptions is achieved by demonstration of virus in vesicular fluid, usually by direct immunofluorescence or PCR. HSV encephalitis is diagnosed by a positive PCR for HSV in CSF. Serology is of limited value, only confirming whether an individual has had previous infection.

**Management**

The acyclic antivirals are the treatment of choice for HSV infection (see Box 13.33, p. 318). Therapy of localised disease must commence in the first 48 hours of clinical disease (primary or recurrent); thereafter it is unlikely to influence clinical outcome. Oral lesions in an immunocompetent individual may be treated with topical aciclovir. All severe manifestations should be treated, regardless of the time of presentation. Suspicion of HSV encephalopathy is an indication for immediate empirical antiviral therapy. Aciclovir resistance is encountered occasionally in immunocompromised hosts, in which case foscarnet is the treatment of choice.

**Human herpesvirus 8**

Human herpesvirus 8 (HHV-8) (see Box 13.31, p. 317) causes Kaposi’s sarcoma in both AIDS-related and endemic non-AIDS-related forms (p. 397). HHV-8 is spread via saliva, and men who have sex with men have increased incidence of infection. Seroprevalence varies widely, being highest in sub-Saharan Africa. HHV-8 also causes two rare haematological malignancies: primary effusion lymphoma and multicentric Castleman’s disease. Current antivirals are not effective.

**Enterovirus infections**

**Hand, foot and mouth disease**

This systemic infection is usually caused by Coxsackie viruses or occasionally echoviruses. It affects children and occasionally adults, resulting in local or household outbreaks, particularly in the summer months. A relatively mild illness with fever and lymphadenopathy develops after an incubation period of approximately 10 days; 2-3 days later, a painful papular or vesicular rash appears on palmarplantar surfaces of hands and feet, with associated oral lesions on the buccal mucosa and tongue that ulcerate rapidly. A papular
erythematous rash may appear on buttocks and thighs. Antiviral treatment is not available and management consists of symptom relief with analgesics.

**Herpangina**

This infection, caused by Coxsackie viruses, primarily affects children and teenagers in the summer months. It is characterised by a small number of vesicles at the soft/hard palate junction, often associated with high fever, an extremely sore throat and headache. The lesions are short-lived, rupturing after 2–3 days and rarely persisting for more than 1 week. Treatment is with analgesics if required. Culture of the virus from vesicles or DNA detection by PCR differentiates herpangina from HSV.

**Poxviruses**

These DNA viruses are rare but potentially important pathogens.

**Smallpox (variola)**

This severe disease, which has high mortality, was eradicated worldwide by a global vaccination programme. Interest in the disease has re-emerged due to its potential as a bioweapon. The virus is spread by the respiratory route or contact with lesions, and is highly infectious.

The incubation period is 7–17 days. A prodrome with fever, headache and prostration leads, in 1–2 days, to the rash, which develops through macules and papules to vesicles and pustules, worst on the face and distal extremities. Lesions in one area are all at the same stage of development with no cropping (unlike chickenpox). Vaccination can lead to a modified course of disease with milder rash and lower mortality.

If a case of smallpox is suspected, national public health authorities must be contacted. Electron micrography (like Fig. 13.14) and DNA detection tests (PCR) are used to confirm smallpox or, using specific primers, an alternative poxvirus.

**Monkeypox**

Despite the name, the animal reservoirs for this virus are probably small squirrels and rodents. It causes a rare zoonotic infection in communities in the rainforest belt of Central Africa, producing a vesicular rash that is indistinguishable from smallpox, but differentiated by the presence of lymphadenopathy. Little person-to-person transmission occurs. Outbreaks outside Africa have been linked to importation of African animals as exotic pets. Diagnosis is by electron micrography and/or DNA detection (PCR).

**Cowpox**

Humans in contact with infected cows develop large vesicles, usually on the hands or arms and associated with fever and regional lymphadenitis. The reservoir is thought to be wild rodents, and the virus also produces symptomatic disease in cats and a range of other animals.

**Vaccinia virus**

This laboratory strain is the basis of the existing vaccine to prevent smallpox. Widespread vaccination is no longer recommended due to the likelihood of local spread from the vaccination site (potentially life-threatening in those with eczema (eczema vaccinatum) or immune deficiency) and of encephalitis. However, vaccination may still be recommended for key medical staff.

**Other poxviruses: orf and molluscum contagiosum**

See page 1279 and Figure 13.14.

**Gastrointestinal viral infections**

**Norovirus (Norwalk agent)**

Norovirus is the most common cause of infectious gastroenteritis in the UK and causes outbreaks in closed communities, such as long-stay hospital wards, cruise ships and military camps. Food handlers may also transmit this virus, which is relatively resistant to decontamination procedures. The incubation period is 24–48 hours. High attack rates and prominent vomiting are characteristic. Diagnosis is by electron microscopy, antigen or DNA detection (PCR) in stool samples. The virus is highly infectious and cases should be isolated and environmental surfaces cleaned with detergents and disinfected with bleach.

**Astrovirus**

Astroviruses cause diarrhoea in small children and occasionally in immunocompromised adults.

**Rotavirus**

Rotaviruses are the major cause of diarrhoeal illness in young children worldwide and cause 10–20% of deaths due to gastroenteritis in developing countries. There are winter epidemics in developed countries, particularly in nurseries. Adults are less often infected but those in close contact with cases may develop disease. The virus infects enterocytes, causing decreased surface absorption. The incubation period is 48 hours and patients present with watery diarrhoea, vomiting, fever and abdominal pain. Dehydration is prominent. Diagnosis is aided by commercially available enzyme immunoassay kits, which require fresh or refrigerated stool samples. Immunity develops to natural infection. Monovalent and multivalent vaccines have been licensed in many
countries and have now demonstrated efficacy in large trials in Africa and the Americas. Increased rates of intussusception were observed with early rotavirus vaccines, but the benefits of the recently licensed vaccines outweigh this risk.

**Hepatitis viruses (A–E)**

See Chapter 23.

**Other viruses**

Adenoviruses are frequently identified from stool culture and implicated as a cause of diarrhoea in children. They have also been linked to cases of intussusception.

### Respiratory viral infections

These infections are described on page 681.

Adenoviruses, rhinoviruses and enteroviruses (Coxsackie viruses and echoviruses) often produce non-specific symptoms. Parainfluenza and respiratory syncytial viruses cause upper respiratory tract disease, croup and bronchiolitis in small children and pneumonia in the immunocompromised. Respiratory syncytial virus also causes pneumonia in nursing home residents and may be associated with nosocomial pneumonia. In recent years, metapneumovirus and bocavirus have been identified as causes of upper and occasionally lower respiratory tract infection. They may also cause pneumonia in immunosuppressed individuals, such as recipients of allogeneic haematopoietic stem cell transplants. The severe acute respiratory syndrome (SARS) caused by the SARS coronavirus emerged as a major respiratory pathogen during an outbreak in 2002–2003, with 8000 cases and 10% mortality (p. 683). In 2012, a novel coronavirus, distantly related to the SARS coronavirus, caused several deaths connected with pneumonia and acute renal failure in patients originating from the Middle East.

### Viral infections with neurological involvement

See also page 1205.

**Japanese B encephalitis**

This flavivirus is an important cause of endemic encephalitis in Japan, China, Russia, South-east Asia, India and Pakistan; outbreaks also occur elsewhere. There are 10000–20000 cases reported to the WHO annually. Pigs and aquatic birds are the virus reservoirs and transmission is by mosquitoes. Exposure to rice paddies is a recognised risk factor.

**Clinical features**

The incubation period is 4–21 days. Most infections are subclinical in childhood and 1% or less of infections lead to encephalitis. Initial systemic illness with fever, malaise and anorexia is followed by photophobia, vomiting, headache and changes in brainstem function. Neurological features other than encephalitis include meningitis, seizures, cranial nerve palsies, flaccid or spastic paralysis, and extrapyramidal features. Mortality with neurological disease is 25%. Most children die from respiratory failure with infection of brainstem nuclei. Approximately 50% of survivors are left with neurological sequelae.

**Investigations, management and prevention**

Other infectious causes of encephalitis should be excluded (p. 1205). There is neutrophilia and often hyponatraemia. CSF analysis reveals lymphocytosis and elevated protein. Serological testing may be helpful and there is a CSF antigen test.

Treatment is supportive, anticipating and treating complications. Vaccination for travellers to endemic areas during the monsoon period is effective prophylaxis. Some endemic countries include this vaccination in their childhood schedules.

**West Nile virus**

This flavivirus has emerged as an important cause of neurological disease in an area that extends from Australia, India and Russia through Africa and Southern Europe and across to North America. The disease has an avian reservoir and a mosquito vector. The elderly are at increased risk of neurological disease.

**Clinical features**

Most infections are asymptomatic. After 2–6 days’ incubation, a mild febrile illness and arthralgia constitute the most common clinical presentation. A prolonged incubation may be seen in immunocompromised individuals. Children may develop a maculopapular rash. Neurological disease is seen in 1% and is characterised by encephalitis, meningitis or asymmetric flaccid paralysis with 10% mortality.

**Diagnosis and management**

Diagnosis is by serology or detection of viral RNA in blood or CSF. Serological tests may show cross-reactivity with other flaviviruses, including vaccine strains. Treatment is supportive.

**Enterovirus 71**

Enterovirus 71 has caused outbreaks around the globe of enteroviral disease with hand, foot and mouth disease (p. 326) and aseptic meningitis. Some cases have been complicated by encephalitis with flaccid paralysis or by brainstem involvement and death. The virus can be isolated from vesicle fluid, stool or CSF, and viral RNA can be detected in CSF by reverse transcription (RT-)PCR.

**Nipah virus encephalitis**

In 1999, a newly discovered paramyxovirus in the Hendra group, the Nipah virus, caused an epidemic of encephalitis amongst Malaysian pig farmers. Infection is through direct contact with pig secretions. Mortality is around 30%. Antibodies to the Hendra virus are present in 76% of cases.

**Human T-cell lymphotropic virus type I**

Human T-cell lymphotropic virus type I (HTLV-1) is a retrovirus which causes chronic infection with development of adult T-cell leukaemia/lymphoma or HTLV-I-associated myelopathy (HAM) in a subset of those infected (p. 1043). It is found mainly in Japan, the Caribbean, Central and South America, and the Seychelles.
HAM or tropical spastic paraparesis occurs in less than 5% of those with chronic infection, and presents with gait disturbance, spasticity of the lower extremities, urinary incontinence, impotence and sensory disturbance. Myositis and uveitis may also occur with HTLV-1 infection. Serology confirms the diagnosis. Treatment is usually supportive for asymptomatic patients but can include zidovudine and interferon-alpha for leukaemia. The role of antivirals in other settings including HAM is being investigated.

**Viral infections with rheumatological involvement**

Rheumatological syndromes characterise a variety of viral infections ranging from exanthems, such as rubella and parvovirus B19 (p. 315), to blood-borne viruses, such as HBV and HIV-1.

**Chikungunya virus**

Chikungunya is an alphavirus that causes fever, rash and arthropathy. It is found principally in Africa and Asia, including India. Humans and non-human primates are the main reservoir and the main vector is the *Aedes aegypti* mosquito. Cases occur in epidemics on a background of sporadic cases. In 2007, an outbreak extended as far north as Italy.

The incubation period is 2–12 days. A period of fever may be followed by an afebrile phase and then recrudescence of fever. Children may develop a maculopapular rash. Adults are susceptible to arthritis, which causes early morning pain and swelling, most often in the small joints. Arthritis can persist for months and may become chronic in individuals who are positive for human leucocyte antigen (HLA)-B27. Related alphaviruses causing similar syndromes include Sindbis virus (Scandinavia and Africa), *O’nyong-nyong* virus (Central Africa), Ross River virus (Australia) and Mayaro virus (Caribbean and South America).

Diagnosis is by serology but cross-reactivity between alphaviruses occurs. Treatment is symptomatic.

**PRION DISEASES**

Prions (p. 134) cause transmissible spongiform encephalopathies in humans, sheep, cows and cats (Box 13.40 and p. 1211). The prion protein is not inactivated by cooking or conventional sterilisation, and transmission is thought to occur by consumption of infected CNS tissue or by inoculation (e.g. via depth EEG electrodes, corneal grafts, cadaveric dura mater grafts and pooled cadaveric growth hormone preparations). The same diseases can occur in an inherited form, due to mutations in the PrP gene.

The apparent transmission of bovine spongiform encephalopathy (BSE) to humans following an outbreak of BSE in the UK beginning in the late 1980s has caused great concern, leading to precautionary measures in the UK, such as leucodepletion of all blood used for transfusion, and the mandatory use of disposable surgical instruments wherever possible for tonsillectomy, appendicectomy and ophthalmological procedures.

**BACTERIAL INFECTIONS**

**Bacterial infections of the skin, soft tissues and bones**

Most infections of the skin, soft tissues and bone are caused by either staphylococci (mainly *Staph. aureus*) or streptococci (mainly *Strep. pyogenes*). Clinical manifestations are also described in Chapters 25 and 27.

**Staphylococcal infections**

Staphylococci are usually found colonising the anterior nares and skin. Traditionally, staphylococci were divided into two groups according to their ability to produce coagulase, an enzyme that converts fibrinogen to fibrin in rabbit plasma, causing it to clot. *Staph. aureus* is coagulase-positive, and most other species coagulase-negative. The coagulase test is now less commonly undertaken, with identification of *Staph. aureus* often achieved by other methods.

*Staph. aureus* is the main cause of staphylococcal infections. *Staph. intermedium* is another coagulase-positive staphylococcus, which causes infection following dog bites. Among coagulase-negative organisms, *Staph. epidermidis* is the predominant commensal organism of the skin, and can cause severe infections in those with central venous catheters or implanted prosthetic materials. *Staph. saprophyticus* is part of the normal vaginal flora and causes urinary tract infections in sexually active young women. Others implicated in human infections include *Staph. lugdunensis, Staph. schleiferi, Staph. haemolyticus* and *Staph. caprae*. Coagulase-negative staphylococci are not usually identified to species level.

Staphylococci are particularly dangerous if they gain access to the blood stream, having the potential to cause disease in many sites (Fig. 13.15). In any patient with staphylococcal bacteraemia, especially injection drug-users, the possibility of endocarditis must be considered (p. 625). Growth of *Staph. aureus* in blood cultures should not be dismissed as a ‘contaminant’ unless all possible underlying sources have been excluded and repeated
blood culture is negative. Any evidence of spreading cellulitis indicates the urgent need for an antistaphylococcal antibiotic, such as flucloxacillin. This is particularly true for mid-facial cellulitis, which can result in cavernous sinus thrombophlebitis.

In addition, Staph. aureus can cause severe systemic disease due to the effects of toxin produced at superficial sites in the absence of tissue invasion by bacteria.

**Skin infections**

Staphylococcal infections cause ecthyma, folliculitis, furuncles, carbuncles, bullous impetigo and the scalded skin syndrome (pp. 1275–1276). They may also be involved in necrotising infections of the skin and subcutaneous tissues (p. 305).

**Wound infections**

Many wound infections are caused by staphylococci, which may significantly prolong post-operative hospital stays (Fig. 13.16A). Prevention involves careful attention to hand hygiene, skin preparation and aseptic technique, and the use of topical and systemic antibiotic prophylaxis.

Treatment is by drainage of any abscesses plus adequate dosage of antistaphylococcal antibiotics. These should be instituted early, particularly if prosthetic implants of any kind have been inserted.

**Cannula-related infection**

Staphylococcal infection associated with cannula sepsis (Fig. 13.16B and p. 304) and thrombophlebitis is an important and, unfortunately, extremely common reason for morbidity following hospital admission. The Visual Infusion Phlebitis (VIP) score is a useful way of monitoring cannulae (Box 13.41). Staphylococci have a predilection for plastic, rapidly forming a biofilm which remains as a source of bacteraemia as long as the plastic is in situ. Local poultice application may relieve symptoms but cannula removal and antibiotic treatment with flucloxacillin (or a glycopeptide if MRSA is suspected) are necessary if there is any suggestion of spreading infection.

**Meticillin-resistant Staph. aureus**

Resistance to meticillin, due to a penicillin-binding protein mutation, has been recognised in *Staph. aureus* for more than 30 years. The recognition of resistance to vancomycin/teicoplanin (glycopeptides) in either glycopeptide intermediate *Staph. aureus* (GISA) or,
rarely, vancomycin-resistant (VRSA) strains threatens
the ability to manage serious infections produced by
such organisms. Metcillin-resistant Staph. aureus
(MRSA) is now a major worldwide health care-acquired
pathogen, accounting for up to 40% of staphylococcal
bacteraemia in developed countries. Community-
acquired MRSA (c-MRSA) currently accounts for 50% of
all MRSA infections in the USA. These organisms have
also acquired other virulence factors, such as Panton-
Valentine leukocidin (PVL), which can cause rapidly
fatal infection in young people. Clinicians must be aware
of the potential danger of these infections and be pre-
tpared to take whatever appropriate infection control
measures are locally advised (p. 145).

Treatment options for MRSA are shown in Box 6.17
(p. 151). Treatment should always be based on the
results of antimicrobial susceptibility testing, since resist-
ance to all these agents occurs. Milder MRSA infections
may be treated with clindamycin, tetracyclines or
cotrimoxazole. Glycopeptides, linezolid and daptomycin
are reserved for treatment of more severe infec-
tions. PVL-producing Staph. aureus infections should be
treated with protein-inhibiting antibiotics (clindamycin,
linezolid).

Staphylococcal toxic shock syndrome

Staphylococcal toxic shock syndrome (TSS) is a serious
and life-threatening disease associated with infection by
Staph. aureus, which produces a specific toxin (toxic
shock syndrome toxin 1, TSST1). It was commonly seen
in young women in association with the use of highly
absorbent intravaginal tampons but can occur with any
Staph. aureus infection involving a relevant toxin-
producing strain. The toxin acts as a ‘superantigen’, trig-
gerating significant T-helper cell activation and massive
cytokine release.

TSS has an abrupt onset with high fever, generalised
systemic upset (myalgia, headache, sore throat and
vomiting), a generalised erythematous blanching rash
resembling scarlet fever, and hypotension. It rapidly
progresses over a matter of hours to multisystem
involvement with cardiac, renal and hepatic compro-
mise, leading to death in 10–20%. Recovery is accompa-
nied by 7–10 days by desquamation (Fig. 13.17).

The diagnosis is clinical and may be confirmed in
menstrual cases by vaginal examination, the finding of a
retained tampon, and microbiological examination
by Gram stain demonstrating typical staphylococci. Sub-
sequent culture and demonstration of toxin production
are confirmatory.

Management

Treatment is with immediate and aggressive fluid resus-
citation and an intravenous antistaphylococcal antibiotic
(flucloxacillin or vancomycin), usually with the addition
of a protein synthesis inhibitor (e.g. clindamycin) to
inhibit toxin production. Intravenous immunoglobulin
is occasionally added in the most severe cases on the
basis of efficacy in streptococcal toxic shock. Women
who recover should be advised not to use tampons for
at least 1 year and should also be warned that, due to an
inadequate antibody response to TSST1, the condition
can recur.

Streptococcal infections

Streptococci are nasopharyngeal and gut commensals,
which appear as Gram-positive cocci in chains (Fig. 6.3,
p. 136). They are classified by the haemolysis they
produce on blood agar (Fig. 6.4, p. 136) and by their
serotypes (Box 13.42). Some streptococci (e.g. Strep.
milleri group) defy simple classification.

Skin presentations of streptococcal infections

Group A streptococci (GAS) are the major cause of cellu-
litis, erysipelas and impetigo (pp. 1275 and 1277). Groups C and G streptococci cause cellulitis, in elderly,
diabetic or immunocompromised patients in particular.
Group B streptococcal (GBS) infection is an increasing
problem at the extremes of age.

Streptococcal scarlet fever

Group A (or occasionally groups C and G) streptococci
causing pharyngitis, tonsillitis or other infection may
lead to scarlet fever, if the infecting strain produces a
streptococcal pyrogenic exotoxin. Common in school-
age children, scarlet fever can occur in young adults who
have contact with young children. A diffuse erythema-
tous rash occurs, which blanches on pressure (Fig.
13.18A), classically with circumoral pallor. The tongue,
initially coated, becomes red and swollen (‘strawberry
tongue’ – Fig. 13.18B). The disease lasts about 7 days, the
rash disappearing in 7–10 days, followed by a fine
desquamation. Residual petechial lesions in the antecu-
bital fossa may be seen (‘Pastia’s sign’ – Fig. 13.18C).

Treatment involves active therapy for the underlying
infection (benzylpenicillin or orally available penicillin)
plus symptomatic measures.

Streptococcal toxic shock syndrome

This is associated with severe group A (or occasionally
group C or G) streptococcal skin infections producing
one of a variety of toxins, such as pyogenic exotoxin A.
Like staphylococcal toxic shock syndrome toxin (see
above), these act as superantigens, stimulating a dra-
matic cytokine response. Initially, an influenza-like
illness occurs with, in 50% of cases, signs of localised
infection, most often involving the skin and soft tissues.
A faint erythematous rash, mainly on the chest, rapidly

Fig. 13.17 Full-thickness desquamation after staphylococcal
toxic shock syndrome.
### 13.42 Streptococcal and related infections

<table>
<thead>
<tr>
<th>β-haemolytic group A (Strep. pyogenes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Skin and soft tissue infection (including erysipelas, impetigo, necrotising fasciitis)</td>
</tr>
<tr>
<td>• Streptococcal toxic shock syndrome</td>
</tr>
<tr>
<td>• Puerperal sepsis</td>
</tr>
<tr>
<td>• Scarlet fever</td>
</tr>
<tr>
<td>• Glomerulonephritis</td>
</tr>
<tr>
<td>• Rheumatic fever</td>
</tr>
<tr>
<td>• Bone and joint infection</td>
</tr>
<tr>
<td>• Tonsillitis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β-haemolytic group B (Strep. agalactiae)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Neonatal infections, including meningitis</td>
</tr>
<tr>
<td>• Septicaemia</td>
</tr>
<tr>
<td>• Female pelvic infections</td>
</tr>
<tr>
<td>• Cellulitis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β-haemolytic group C (various zoonotic streptococci)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Septicaemia</td>
</tr>
<tr>
<td>• Cellulitis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α, β- or non-haemolytic group D enterococci (E. faecalis/faecium)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Endocarditis</td>
</tr>
<tr>
<td>• Intra-abdominal infections</td>
</tr>
<tr>
<td>• Urinary tract infection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α, β- or non-haemolytic group D streptococci (Strep. bovis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Endocarditis</td>
</tr>
<tr>
<td>• Septicaemia</td>
</tr>
<tr>
<td>• Liver abscess</td>
</tr>
<tr>
<td>• Brain abscess</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β-haemolytic group G streptococci</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Septicaemia</td>
</tr>
<tr>
<td>• Cellulitis</td>
</tr>
<tr>
<td>• Liver abscess</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α-haemolytic viridans group (Strep. mitis, sanguis, mutans, salivarius)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Septicaemia in immunosuppressed patients</td>
</tr>
<tr>
<td>• Endocarditis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>α-haemolytic optochin-sensitive (Strep. pneumoniae)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pneumonia</td>
</tr>
<tr>
<td>• Meningitis</td>
</tr>
<tr>
<td>• Endocarditis</td>
</tr>
<tr>
<td>• Otitis media</td>
</tr>
<tr>
<td>• Septicaemia</td>
</tr>
<tr>
<td>• Spontaneous bacterial peritonitis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anaerobic streptococci (Peptostreptococcus spp.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Peritonitis</td>
</tr>
<tr>
<td>• Dental infections</td>
</tr>
<tr>
<td>• Liver abscess</td>
</tr>
<tr>
<td>• Pelvic inflammatory disease</td>
</tr>
</tbody>
</table>

**N.B.** All streptococci can cause septicaemia.

---

**Treponematoses**

**Syphilis**

This disease is described on page 419.

**Endemic treponematoses**

**Yaws**

Yaws is a granulomatous disease, mainly involving the skin and bones; it is caused by *Treponema pertenue*, morphologically and serologically indistinguishable from the causative organisms of syphilis and pinta. It is important to establish the geographical origin and sexual history of patients to exclude a false-positive syphilis serology due to the endemic treponemal infections. Between 1950 and 1960, WHO campaigns treated over 60 million people and eradicated yaws from many areas, but the disease has persisted patchily throughout the tropics; there was resurgence in the 1980s and 1990s in West and Central Africa and the South Pacific.

Organisms are transmitted by bodily contact from a patient with infectious yaws through minor abrasions of the skin of another patient, usually a child. After an incubation period of 3–4 weeks, a proliferative granuloma containing numerous treponemes develops at the site of inoculation. This primary lesion is followed by secondary eruptions. In addition, there may be hypertrophic periosteal lesions of many bones, with underlying cortical rarefaction. Lesions of late yaws are characterised by destructive changes which closely resemble the osteitis and gummas of tertiary syphilis and which heal with much scarring and deformity. Investigations and management are outlined in Box 13.43.

The disease disappears with improved housing and cleanliness. In few fields of medicine have
chemotherapy and improved hygiene achieved such dramatic success as in the control of yaws.

Pinta and bejel

These two treponemal infections occur in poor rural populations with low standards of domestic hygiene, but are found in separate parts of the world. They have features in common, notably that they are transmitted by contact, usually within the family and not sexually, and in the case of bejel, through common eating and drinking utensils. Their diagnosis and management are as for yaws (see Box 13.43).

- **Pinta.** Pinta is probably the oldest of the human treponemal infections. It is found only in South and Central America, where its incidence is declining. The infection is confined to the skin. The early lesions are scaly papules or dyschromic patches on the skin. The late lesions are often depigmented and disfiguring.
- **Bejel.** Bejel is the Middle Eastern name for nonvenereal syphilis, which has a patchy distribution across sub-Saharan Africa, the Middle East, Central Asia and Australia. It has been eradicated from Eastern Europe. Transmission is most commonly from the mouth of the mother or child and the primary mucosal lesion is seldom seen. The early and late lesions resemble those of secondary and tertiary syphilis (pp. 419–420) but cardiovascular and neurological disease is rare.

Tropical ulcer

Tropical ulcer is due to a synergistic bacterial infection caused by a fusobacterium (*F. ulcerans*, an anaerobe) and *Treponema vincentii*. It is common in hot, humid regions. The ulcer is most common on the lower legs and develops as a papule that rapidly breaks down to a sharply defined, painful ulcer. The base of the ulcer has a foul slough. Penicillin and metronidazole are useful in the early stages but rest, elevation and dressings are the mainstays of treatment.

Buruli ulcer

This ulcer is caused by *Mycobacterium ulcerans* and occurs widely in tropical rainforests. In 1999, a survey in Ghana found 6500 cases; there are an estimated 10000 cases in West Africa as a whole.

The initial lesion is a small subcutaneous nodule on the arm or leg. This breaks down to form a shallow, necrotic ulcer with deeply undermined edges, which extends rapidly. Healing may occur after 6 months, but granuloma formation and the accompanying fibrosis cause contractures and deformity. Clumps of acid-fast bacilli can be detected in the ulcer floor.

A combination of rifampicin and streptomycin can cure the infection. Infected tissue should be removed surgically. Health campaigns in Ghana have successfully focused on early removal of the small, pre-ulcerative nodules.

### Systemic bacterial infections

#### Brucellosis

Brucellosis is an enzootic infection (i.e. endemic in animals). Although six species of *Brucella* Gram-negative bacilli are known, only four are important to humans: *B. melitensis* (goats, sheep and camels in Europe, especially the Mediterranean basin, the Middle East, Africa, India, Central Asia and South America), *B. abortus* (cattle, mainly in Africa, Asia and South America), *B. suis* (pigs in South Asia) and *B. canis* (dogs). *B. melitensis* causes the most severe disease; *B. suis* is often associated with abscess formation.

Infected animals may excrete *Brucella* spp. in their milk for prolonged periods, and human infection is acquired by ingesting contaminated dairy products (especially unpasteurised milk), uncooked meat or offal. Animal urine, faeces, vaginal discharge and uterine products may act as sources of infection through abraded skin or via splashes and aerosols to the respiratory tract and conjunctiva.

#### Clinical features

*Brucella* spp. are intracellular organisms that survive for long periods within the reticulo-endothelial system. This explains many of the clinical features, including the chronicity of disease and tendency to relapse, even after antimicrobial therapy.

Acute illness is characterised by a high swinging temperature, rigours, lethargy, headache, joint and muscle pains, and scrotal pain. Occasionally, there is delirium, abdominal pain and constipation. Physical signs are non-specific, e.g. enlarged lymph nodes. Enlargement of the spleen may lead to hypersplenism and thrombocytopenia.

Localised infection (Fig. 13.19), which occurs in about 30% of patients, is more likely if diagnosis and treatment are delayed.

#### Diagnosis

Definitive diagnosis depends on the isolation of the organism. Blood cultures are positive in 75–80% of infections caused by *B. melitensis* and 50% of those caused by *B. abortus*. Bone marrow culture should not be used routinely but may increase the diagnostic yield, particularly if antibiotics have been given before specimens are taken. CSF culture in neurobrucellosis is positive in about 30% of cases. The laboratory should be alerted to a suspected diagnosis of brucellosis, as the organism has a propensity for infecting laboratory workers and must be cultured at an enhanced containment level.

Serology may also aid diagnosis. In endemic areas, a single high antibody titre of more than 1/320 or a fourfold rise in titre is needed to support a diagnosis of acute infection. The test usually takes several weeks to become positive.
INFECTIONOUS DISEASE

Fig. 13.19 Clinical features of brucellosis.

13.44 Treatment of brucellosis

Adults with non-localised disease

- Doxycycline 100 mg twice daily orally for 6 wks plus gentamicin 5 mg/kg IV once daily for 7 days
- Doxycycline 100 mg twice daily plus rifampicin 600–900 mg orally once daily for 6 wks

Bone disease

- Doxycycline 100 mg twice daily plus rifampicin 600–900 mg once daily orally for 6 wks plus gentamicin 5 mg/kg IV once daily for 7 days
- Ciprofloxacin 750 mg twice daily orally plus rifampicin 600–900 mg orally once daily for 3 mths

Neurobrucellosis

- Doxycycline 100 mg twice daily plus rifampicin 600–900 mg orally once daily for 6 wks plus ceftriaxone 2 g IV twice daily until CSF is clear (though susceptibility should be confirmed because sensitivity to third-generation cephalosporins varies amongst strains)

Endocarditis

- Almost always needs surgical intervention plus
- Doxycycline 100 mg twice daily, rifampicin 600–900 mg orally once daily and co-trimoxazole 5 mg/kg of trimethoprim component for 6 mths plus gentamicin 5 mg/kg IV once daily for 2–4 wks

Pregnancy

- Rifampicin 600–900 mg orally once daily and co-trimoxazole 5 mg/kg of trimethoprim component for 4 wks, but caution in last week of pregnancy due to displacement of bilirubin from albumin by drugs and risk of kernicterus to the fetus

positive but should eventually detect 95% of acute infections.

Management

Aminoglycosides show synergistic activity with tetracyclines against brucellae. Treatment regimens for different forms of brucellosis are outlined in Box 13.44.

Borrelia infections

Borrelia are flagellated spirochaetal bacteria which infect humans after bites from ticks or lice. They cause a variety of human infections worldwide (Box 13.45).

Lyme disease

Lyme disease (named after the town of Old Lyme in Connecticut, USA) is caused by B. burgdorferi, which occurs in the USA, Europe, Russia, China, Japan and Australia. In Europe, two additional genospecies are also encountered, B. afzelii and B. garinii. The reservoir of infection is ixodid (hard) ticks that feed on a variety of large mammals, particularly deer. Birds may spread ticks over a wide area. The organism is transmitted to humans via the bite of infected ticks; larval, nymphal and adult forms are all capable of spreading infection.

Ehrlichiosis is a common co-infection with Lyme disease (Anaplasma phagocytophila, human granulocytic anaplasmosis (HGA); Ehrlichia chaffeensis, human monocytic ehrlichiosis (HME)).

Clinical features

There are three stages of disease. Progression may be arrested at any stage.

- Early localised disease. The characteristic feature is a skin reaction around the site of the tick bite, known as erythema migrans (Fig. 13.20). Initially, a red
abnormalities may occur, but are rare in the UK. Acrodermatitis chronica atrophicans is an uncommon late complication seen more frequently in Europe than North America. Doughy, patchy discoloration occurs on the peripheries, eventually leading to shiny atrophic skin. The lesions are easily mistaken for those of peripheral vascular disease. In patients from an endemic area or with risk factors, who have facial nerve palsy, Lyme disease should be considered.

**Diagnosis**

The diagnosis of early Lyme borreliosis is often clinical. Culture from biopsy material is not generally available, has a low yield, and may take longer than 6 weeks. Antibody detection is frequently negative early in the course of the disease, but sensitivity increases to 90–100% in disseminated or late disease. Immunofluorescence or ELISA can give false-positive reactions in a number of conditions, including other spirochaetal infections, infectious mononucleosis, rheumatoid arthritis and systemic lupus erythematosus (SLE). Immunoblot (Western blot) techniques are more specific and, although technically demanding, should be used to confirm the diagnosis. Microorganism DNA detection by PCR has been applied to blood, urine, CSF, and biopsies of skin and synovium.

**Management**

Recent evidence suggests that asymptomatic patients with positive antibody tests should not be treated. However, erythema migrans always requires therapy because organisms may persist and cause progressive disease, even if the skin lesions resolve. Standard therapy consists of a 14-day course of doxycycline (200 mg daily) or amoxicillin (500 mg 3 times daily). Some 15% of patients with early disease will develop a mild Jarisch–Herxheimer reaction (JHR) during the first 24 hours of therapy (p. 421). In pregnant women and small children, or in those allergic to amoxicillin and doxycycline, 14-day treatment with cefuroxime axetil (500 mg twice daily) or erythromycin (250 mg 4 times daily) may be used.

Disseminated disease and arthritis require therapy for a minimum of 28 days. Arthritis may respond poorly, and prolonged or repeated courses may be necessary. Neuroborreliosis is treated with parenteral β-lactam...
antibiotics for 3–4 weeks; the cephalosporins may be superior to penicillin in this situation.

Prevention
Protective clothing and insect repellents should be used in tick-infested areas. Since the risk of borrelial transmission is lower in the first few hours of a blood feed, prompt removal of ticks is advisable. Unfortunately, larval and nymphal ticks are tiny and may not be noticed. Where risk of transmission is high, a single 200 mg dose of doxycycline, given within 72 hours of exposure, has been shown to prevent erythema migrans. A recombinant OspA vaccine was developed but withdrawn due to side-effects.

Louse-borne relapsing fever
The human body louse, Pediculus humanus, causes itching. Borreliae (B. recurrensis) are liberated from infected lice when they are crushed during scratching, which also inoculates the borreliae into the skin. The disease occurs worldwide, with epidemic relapsing fever most often seen in Central/East Africa and South America.

The borreliae multiply in the blood, where they are abundant in the febrile phases, and invade most tissues, especially the liver, spleen and meninges. Hepatitis and thrombocytopenia are common.

Clinical features
Onset is sudden with fever. The temperature rises to 39.5–40.5°C, accompanied by a tachycardia, headache, generalised aching, injected conjunctivae (Fig. 13.21) and, frequently, a petechial rash, epistaxis and herpes labialis. As the disease progresses, the liver and spleen frequently become tender and palpable, and jaundice is common. There may be severe serosal and intestinal haemorrhage, mental confusion and meningism. The fever ends in crisis between the 4th and 10th days, often associated with profuse sweating, hypotension, and circulatory and cardiac failure. There may be no further fever but, in a proportion of patients, after an afebrile period of about 7 days, there are one or more relapses, which are usually milder and less prolonged. In the absence of specific treatment, the mortality rate is up to 40%, especially among the elderly and malnourished.

Investigations and management
The organisms are demonstrated in the blood during fever either by dark ground microscopy of a wet film or in Wright–Giemsa stained thick and thin films.

The problems of treatment are to eradicate the organism, to minimise the severe JHR which inevitably follows successful chemotherapy, and to prevent relapses. The safest treatment is procaine penicillin 300 mg IM, followed the next day by 0.5 g tetracycline. Tetracycline alone is effective and prevents relapse, but may give rise to a worse reaction. Doxycycline 200 mg once orally in place of tetracycline has the advantage of also being curative for typhus, which often accompanies epidemics of relapsing fever. JHR is best managed in a high-dependency unit with expert nursing and medical care.

The patient, clothing and all contacts must be freed from lice, as in epidemic typhus.

Tick-borne relapsing fever
Soft ticks (Ornithodoros spp.) transmit B. duttonii (and several other borrelia species) through saliva while feeding on their host. People sleeping in mud houses are at risk, as the tick hides in crevices during the day and feeds on humans during the night. Rodents are the reservoir in all parts of the world except East Africa, where humans are the reservoir. Clinical manifestations are similar to those seen with the louse-borne disease, but spirochaetes are detected in fewer patients on dark field microscopy. A 7-day course (due to a higher relapse rate than in louse-borne relapsing fever) of treatment with either tetracycline (500 mg 4 times daily) or erythromycin (500 mg 4 times daily) is needed.

Leptospirosis

Microbiology and epidemiology
Leptospirosis is one of the most common zoonotic diseases, favoured by a tropical climate and flooding during the monsoon but occurring worldwide. Leptospires are tightly coiled, thread-like organisms about 5–7 μm in length, which are actively motile; each end is bent into a hook. Leptospira interrogans is pathogenic for humans. The genus can be separated into more than 200 serovars (subtypes) belonging to 23 serogroups.

Leptospirosis appears to be ubiquitous in wildlife and in many domestic animals. The organisms persist indefinitely in the convoluted tubules of the kidney and are shed into the urine in massive numbers, but infection is asymptomatic in the host. The most frequent hosts are rodents, especially the common rat (Rattus norvegicus). Particular leptospiral serogroups are associated with characteristic animal hosts; for example, L. icterohaemorrhagiae is the classical parasite of rats and L. canicola of dogs. There is nevertheless considerable overlap in host–serogroup associations.

Leptospires can enter their human hosts through intact skin or mucous membranes, but entry is facilitated by cuts and abrasions. Prolonged immersion in contaminated water will also favour invasion, as the spirochaete can survive in water for months. Leptospirosis is common in the tropics and also in freshwater sports enthusiasts.

Clinical features
After a relatively brief bacteraemia, invading organisms are distributed throughout the body, mainly in kidneys, liver, meninges and brain. The incubation period averages 1–2 weeks. Four main clinical syndromes can be discerned and clinical features can involve multiple different organ systems (Fig. 13.22).
**Bacteraeic leptospirosis**

Bacteraemia with any serogroup can produce a non-specific illness with high fever, weakness, muscle pain and tenderness (especially of the calf and back), intense headache and photophobia, and sometimes diarrhoea and vomiting. Conjunctival congestion is the only notable physical sign. The illness comes to an end after about 1 week, or else merges into one of the other forms of infection.

**Aseptic meningitis**

Classically associated with *L. canicola* infection, this illness is very difficult to distinguish from viral meningitis. The conjunctivae may be congested but there are no other differentiating signs. Laboratory clues include a neutrophil leucocytosis, abnormal LFTs, and the occasional presence of albumin and casts in the urine.

**Icteric leptospirosis (Weil’s disease)**

Fewer than 10% of symptomatic infections result in severe icteric illness. Weil’s disease is a dramatic life-threatening event, characterised by fever, haemorrhages, jaundice and renal impairment. Conjunctival hyperaemia is a frequent feature. The patient may have a transient macular erythematous rash, but the characteristic skin changes are purpura and large areas of bruising. In severe cases there may be epistaxis, haematemesis and melaena, or bleeding into the pleural, pericardial or subarachnoid spaces. Thrombocytopenia, probably related to activation of endothelial cells with platelet adhesion and aggregation, is present in 50% of cases. Jaundice is deep and the liver is enlarged, but there is usually little evidence of hepatic failure or encephalopathy. Renal failure, primarily caused by impaired renal perfusion and acute tubular necrosis, manifests as oliguria or anuria, with the presence of albumin, blood and casts in the urine.

Weil’s disease may also be associated with myocarditis, encephalitis and aseptic meningitis. Uveitis and iritis may appear months after apparent clinical recovery.

**Pulmonary syndrome**

This syndrome has long been recognised in the Far East, and has been described during an outbreak of leptospirosis in Nicaragua. It is characterised by haemoptysis, patchy lung infiltrates on chest X-ray, and respiratory failure. Total bilateral lung consolidation and ARDS (p. 192) with multi-organ dysfunction may develop, with a high mortality (over 50%).

**Diagnosis**

A polymorphonuclear leucocytosis is accompanied in severe infection by thrombocytopenia and elevated blood levels of creatine kinase. In jaundiced patients, there is hepatitis and the prothrombin time may be prolonged. The CSF in leptospiral meningitis shows a variable cellular response, a moderately elevated protein level and normal glucose content. Acute renal failure due to interstitial nephritis is common.

In the tropics, dengue, malaria, typhoid fever, scrub typhus and hantavirus infection are important differential diagnoses.

Definitive diagnosis of leptospirosis depends upon isolation of the organism, serological tests or the detection of specific DNA. In general, however, it is probably under-diagnosed.

- Blood cultures are most likely to be positive if taken before the tenth day of illness. Special media are required and cultures may have to be incubated for several weeks.
- Leptospires appear in the urine during the second week of illness, and in untreated patients may be recovered on culture for several months.
INFECTIOUS DISEASE

- Serological tests are diagnostic if seroconversion or a fourfold increase in titre is demonstrated. The microscopic agglutination test (MAT) is the test of choice and can become positive by the end of the first week. IgM ELISA and immunofluorescent techniques are, however, easier to perform, while rapid immunochromatographic tests are specific but of only moderate sensitivity in the first week of illness.
- Detection of leptospiral DNA by PCR is possible in blood in early symptomatic disease, and in urine from the eighth day of illness and for many months thereafter.

Management and prevention

The general care of the patient is critically important. Blood transfusion for haemorrhage and careful attention to renal failure, the usual cause of death, are especially important. Renal failure is potentially reversible with adequate support, such as dialysis. The optimal antimicrobial regimen has not been established. Most infections are self-limiting. Therapy with either oral doxycycline (100 mg twice daily for 1 week) or intravenous penicillin (900 mg 4 times daily for 1 week) is effective but may not prevent the development of renal failure. Parenteral ceftriaxone (1 g daily) is as effective as penicillin. A Jarisch–Herschheimer reaction may occur during treatment but is usually mild. Uveitis is treated with a combination of systemic antibiotics and local corticosteroids. There is no role for the routine use of corticosteroids in the management of leptospirosis.

Trials in military personnel have shown that infection with *L. interrogans* can be prevented by taking prophylactic doxycycline 200 mg weekly.

Plague

Plague is caused by *Yersinia pestis*, a small Gram-negative bacillus that is spread between rodents by their fleas. If domestic rats become infected, infected fleas may bite humans. Hunters and trappers can contract plague from handling rodents. In the late stages of human plague, *Y. pestis* may be expectorated and spread between humans by droplets, causing ‘pneumonic plague’.

Epidemics of plague, such as the ‘Black Death’, have occurred since ancient times. It is often said that the first sign of plague is the appearance of dead rats. Plague foci are widely distributed throughout the world, including the USA; human cases are reported from about ten countries per year (Fig. 13.23).

*Y. pestis* is a potential bioweapon because of its capacity for mass production and aerosol transmission, and the high fatality rate associated with pneumonic plague.

Clinical features

Organisms inoculated through the skin are taken rapidly to the draining lymph nodes, where they elicit a severe inflammatory response that may be haemorrhagic. If the infection is not contained, septicaemia ensues and necrotic, purulent or haemorrhagic lesions develop in many organs. Oliguria and shock follow, and disseminated intravascular coagulation may result in widespread haemorrhage. Inhalation of *Y. pestis* causes alveolitis. The incubation period is 3–6 days, but shorter in pneumonic plague.

Bubonic plague

In this, the most common form of the disease, onset is usually sudden, with a rigor, high fever, dry skin and severe headache. Soon, aching and swelling at the site of the affected lymph nodes begin. The groin is the most common site of this ‘bubo’, made up of the swollen lymph nodes and surrounding tissue. Some infections are relatively mild but, in the majority of patients, toxemia quickly increases, with a rapid pulse, hypotension and mental confusion. The spleen is usually palpable.

Septicaemic plague

Those not exhibiting a bubo usually deteriorate rapidly and have a high mortality. The elderly are more prone to this form of illness. The patient is toxic and may have gastrointestinal symptoms, such as nausea, vomiting, abdominal pain and diarrhoea. DIC may occur, manifested by bleeding from various orifices or puncture sites, along with ecchymoses. Hypotension, shock, renal failure and ARDS may lead to further deterioration. Meningitis, pneumonia and expectation of blood-stained sputum containing *Y. pestis* may complicate septicaemic, or occasionally bubonic, plague.

Pneumonic plague

Following primary infection in the lung, the onset of disease is very sudden, with cough and dyspnoea. The patient soon expectorates copious blood-stained, frothy, highly infective sputum, becomes cyanosed and dies. Chest radiology reveals bilateral infiltrates which may be nodular and progress to an ARDS-like picture.

Investigations

The organism may be cultured from blood, sputum and bubo aspirates. For rapid diagnosis, Gram, Giemsa and Wayson’s stains (the latter containing methylene blue) are applied to smears from these sites. *Y. pestis* is seen as bipolar staining coccobacilli, sometimes referred to as having a ‘safety pin’ appearance. Smears are also subjected to antigen detection by immunofluorescence, using *Y. pestis* F1 antigen-specific antibodies. The diagnosis may be confirmed by seroconversion or a single high titre (> 128) of anti-F1 antibodies in serum. DNA detection by PCR is under evaluation.

Plague is a notifiable disease under international health regulations (p. 147).

Management

If the diagnosis is suspected on clinical and epidemiological grounds, treatment must be started as soon as, or even before, samples have been collected for laboratory

---

**Fig. 13.23** Foci of the transmission of plague. Reproduced by permission of WHO.
diagnosis. Streptomycin (1 g twice daily) or gentamicin (1 mg/kg 3 times daily) is the drug of choice. Tetracycline (500 mg 4 times daily) and chloramphenicol (12.5 mg/kg 4 times daily) are alternatives. Fluoroquinolones (ciprofloxacin and levofloxacin) may be as effective, but there is less clinical experience. Treatment may also be needed for acute circulatory failure, DIC and hypoxia.

**Prevention and infection control**

Rats and fleas should be controlled. In endemic areas, people should avoid handling and skinning wild animals. The patient should be isolated for the first 48 hours or until clinical improvement begins. Attendants must wear gowns, masks and gloves. Exposed symptomatic or asymptomatic people who have been in close contact with a patient with pneumonic plague should receive post-exposure antibiotic prophylaxis (doxycycline 100 mg or ciprofloxacin 500 mg twice daily) for 7 days.

A formalin-killed vaccine is available for those at occupational risk but offers little protection against pneumonic plague. A recombinant subunit vaccine (protein antigens F1 + V) is in development.

**Listeriosis**

*Listeria monocytogenes* is an environmental Gram-positive bacillus which can contaminate food. Outbreaks have been associated with raw vegetables, soft cheeses, undercooked chicken, fish, meat and pâtés. The bacterium demonstrates ‘cold enrichment’, outgrowing other contaminating bacteria during refrigeration. Although foodborne outbreaks of gastroenteritis have been reported in immunocompetent individuals, *Listeria* causes more significant invasive infection, especially in pregnancy, the elderly (over 55 years) and the immunocompromised.

In pregnancy, in addition to systemic symptoms of fever and myalgia, listeriosis causes chorioamnionitis, fetal deaths, abortions and neonatal infection. In other susceptible individuals, it causes systemic illness due to bacteraemia without focal symptoms. Meningitis, similar to other bacterial meningitis but with normal CSF glucose, is the next most common presentation; CSF usually shows increased neutrophils but occasionally only the mononuclear cells are increased (p. 1201).

**Investigations and management**

Diagnosis is made by blood and CSF culture. The organism grows readily in culture media.

The most effective regimen consists of a combination of an intravenous aminopenicillin (amoxicillin or ampicillin) plus an aminoglycoside. A sulfamethoxazole/trimethoprim combination can be used in those with penicillin allergy. Cephalosporins are of no use in this infection, as the organism is inherently resistant, an important consideration when empirically treating meningitis.

Proper treatment of foods before eating is the key to preventing listeriosis. Pregnant women are advised to avoid high-risk products, including soft cheeses.

**Typhoid and paratyphoid (enteric) fevers**

Typhoid and paratyphoid fevers, which are transmitted by the faecal–oral route, are important causes of fever in India, sub-Saharan Africa and Latin America. Elsewhere, they are relatively rare. Enteric fevers are caused by infection with *Salmonella typhi* and *S. paratyphi* A and B. After a few days of bacteraemia, the bacilli localise, mainly in the lymphoid tissue of the small intestine, resulting in typical lesions in the Peyer’s patches and follicles. These swell at first, then ulcerate and usually heal. After clinical recovery, about 5% of patients become chronic carriers (i.e. continue to excrete the bacteria after 1 year); the bacilli may live in the gallbladder for months or years and pass intermittently in the stool and, less commonly, in the urine.

**Clinical features**

**Typhoid fever**

Clinical features are outlined in Box 13.46. The incubation period is typically about 10–14 days but can be longer, and the onset may be insidious. The temperature rises in a stepladder fashion for 4 or 5 days with malaise, increasing headache, drowsiness and aching in the limbs. Constipation may be caused by swelling of lymphoid tissue around the ileocaecal junction, although in children diarrhoea and vomiting may be prominent early in the illness. The pulse is often slower than would be expected from the height of the temperature, i.e. a relative bradycardia.

At the end of the first week, a rash may appear on the upper abdomen and on the back as sparse, slightly raised, rose-red spots, which fade on pressure. It is usually visible only on white skin. Cough and epistaxis occur. Around the 7th–10th day, the spleen becomes palpable. Constipation is then succeeded by diarrhoea and abdominal distension with tenderness. Bronchitis and delirium may develop. If untreated, by the end of the second week the patient may be profoundly ill.

**Paratyphoid fever**

The course tends to be shorter and milder than that of typhoid fever and the onset is often more abrupt with acute enteritis. The rash may be more abundant and the intestinal complications less frequent.

**Complications**

These are given in Box 13.47. Haemorrhage from, or a perforation of, the ulcerated Peyer’s patches may occur at the end of the second week or during the third week of the illness. A drop in temperature to normal or subnormal levels may be falsely reassuring in patients with intestinal haemorrhage. Additional complications may

<table>
<thead>
<tr>
<th>13.46 Clinical features of typhoid fever</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First week</strong></td>
</tr>
<tr>
<td>• Fever</td>
</tr>
<tr>
<td>• Headache</td>
</tr>
<tr>
<td>• Myalgia</td>
</tr>
<tr>
<td>• Relative bradycardia</td>
</tr>
<tr>
<td>• Constipation</td>
</tr>
<tr>
<td>• Diarrhoea and vomiting in children</td>
</tr>
<tr>
<td><strong>End of first week</strong></td>
</tr>
<tr>
<td>• Rose spots on trunk</td>
</tr>
<tr>
<td>• Spleenomegaly</td>
</tr>
<tr>
<td>• Cough</td>
</tr>
<tr>
<td>• Abdominal distension</td>
</tr>
<tr>
<td>• Diarrhoea</td>
</tr>
<tr>
<td><strong>End of second week</strong></td>
</tr>
<tr>
<td>• Delirium, complications, then coma and death (if untreated)</td>
</tr>
</tbody>
</table>
Infections may involve almost any viscus or system because of the septicaemia present during the first week. Bone and joint infection is common in children with sickle-cell disease.

**Investigations**

In the first week, diagnosis may be difficult because, in this invasive stage with bacteraemia, the symptoms are those of a generalised infection without localising features. A white blood count may be helpful, as there is typically a leucopenia. Blood culture is the most important diagnostic method. The faeces contain the organism more frequently in the second and third weeks.

**Management**

Antibiotic therapy must be guided by in vitro sensitivity testing. Chloramphenicol (500 mg 4 times daily), ampicillin (750 mg 4 times daily) and co-trimoxazole (2 tablets or IV equivalent twice daily) are losing their effect due to resistance in many areas of the world, especially India and South-East Asia. The fluoroquinolones are the drugs of choice (e.g. ciprofloxacin 500 mg twice daily) if the organism is susceptible, but resistance is common, especially in the Indian subcontinent and also in the UK. Extended-spectrum cephalosporins (ceftriaxone and cefotaxime) are useful alternatives but have a slightly increased treatment failure rate. Azithromycin (500 mg once daily) is an alternative when fluoroquinolone resistance is present but has not been validated in severe disease. Treatment should be continued for 14 days. Pyrexia may persist for up to 5 days after the start of specific therapy. Even with effective chemotherapy, there is still a danger of complications, recrudescence of the disease and the development of a carrier state.

Chronic carriers were formerly treated for 4 weeks with ciprofloxacin but may require an alternative agent and duration, as guided by antimicrobial sensitivity testing. Cholecystectomy may be necessary.

**Prevention**

Improved sanitation and living conditions reduce the incidence of typhoid. Travellers to countries where enteric infections are endemic should be inoculated with one of the three available typhoid vaccines (two inactivated injectable and one oral live attenuated).

**Tularemia**

Tularemia is primarily a zoonotic disease of the northern hemisphere. It is caused by a highly infectious Gram-negative bacillus, *Francisella tularensis*. *F. tularensis* is passed transovarially (ensuring transmission from parent to progeny) in ticks, which allows persistence in nature without the absolute requirement for an infected animal reservoir. It is a potential weapon for bioterrorism. Wild rabbits, rodents, and domestic dogs or cats are some of the many potential reservoirs, and ticks, mosquitoes or other biting flies are the vectors.

Infection is introduced either through an arthropod or animal bite or via contact with infected animals, soil or water through skin abrasions. This results in the most common ‘ulceroglandular’ variety of the disease (70–80%), characterised by skin ulceration with regional lymphadenopathy. There is also a purely ‘glandular’ form. Alternatively, inhalation of the infected aerosols may result in pulmonary tularemia, presenting as pneumonia. Rarely, the portal of entry of infection may be the conjunctiva, leading to a nodular, ulcerated conjunctivitis with regional lymphadenopathy (an ‘oculoglandular’ form).

**Investigations and management**

Demonstration of a single high titre (≥ 1:160) or a fourfold rise in 2–3 weeks in the tularemia tube agglutination test confirms the diagnosis. Bacterial yield from the lesions is extremely poor. DNA detection methods to enable rapid diagnosis are in development.

Treatment consists of a 7–10-day course of parenteral aminoglycosides, streptomycin (7.5–10 mg/kg twice daily) or gentamicin (1.7 mg/kg 3 times daily). *F. tularensis* is not susceptible to most other antibiotics.

**Melioidosis**

Melioidosis is caused by *Burkholderia pseudomallei*, a saprophyte found in soil and water (rice paddy fields). Infection is by inoculation or inhalation, leading to bacteraemia, which is followed by the formation of abscesses in the lungs, liver and spleen. Patients with diabetes, renal stones, thalassaemia or severe burns are particularly susceptible. The disease is most common in South India, East Asia and northern Australia, and carries a significant mortality. Disease may present many years or decades after the initial exposure.

**Clinical features**

There is high fever, prostration and sometimes diarrhoea, with signs of pneumonia and enlargement of the liver and spleen. The chest X-ray resembles that of acute caseous tuberculosis. In more chronic forms, multiple abscesses occur in subcutaneous tissue and bone, and profound wasting is a major problem.

**Investigations and management**

Culture of blood, sputum or pus may yield *B. pseudomallei*. Indirect haemagglutination testing can be helpful in travellers; however, most people in endemic areas are seropositive.

In the acute illness, prompt treatment, without waiting for confirmation by culture, may be life-saving. Ceftazidime 100 mg/kg (2 g 3 times daily), imipenem 50 mg/kg (1 g 4 times daily) or meropenem (0.5–1 g 3 times daily) is given for 2–3 weeks. This is followed by maintenance therapy of doxycycline 200 mg daily, plus co-trimoxazole (sulfamethoxazole 1600 mg plus trimethoprim 320 mg twice daily) for a minimum of 12 weeks. Abscesses should be drained surgically.
**Actinomycete infections**

**Nocardiosis**

Nocardiosis is an uncommon Gram-positive bacterial infection caused by aerobic actinomycetes of the genus *Nocardia* found in the soil. Infection occurs most frequently by direct traumatic inoculation or occasionally via inhalation or ingestion. Nocardiosis can result in localised cutaneous ulcers or nodules, most often in the lower limbs. Chronic destructive infection in tropical countries can result in actinomycetoma, involving soft tissues with occasional penetration to the bone. Actinomycetoma may also be caused by other bacteria, and a similar clinical syndrome, termed eumycetoma, is caused by fungi (p. 382). Systemic *Nocardia* infection, most likely in immunocompromised individuals, results in suppurative disease with lung and brain abscesses.

On microscopy, *Nocardia* spp. appear as long, filamenous, branching Gram-positive rods which are also weakly acid-fast. They are easily grown in culture but require prolonged incubation.

Treatment is guided by susceptibility testing. Systemic infection typically requires combinations of ceftriaxone, meropenem, amikacin and co-trimoxazole, often for 6–12 months or longer. Abscesses are drained surgically when this is feasible. Localised cutaneous infection is usually treated with a single agent for 1–3 months.

Actinomycetoma is also treated with prolonged antibiotic combinations. There is no universal consensus on the most appropriate drug or combination. The usual combination is streptomycin and dapsone, with dapsone replaced by co-trimoxazole in cases with intolerance or refractory disease. Success has also been reported with co-trimoxazole plus amikacin, with rifampicin added in difficult cases and to prevent recurrence.

**Actinomyces israelii**

*Actinomyces israelii* can cause deep infection in the head and neck, and also suppurating disease in the pelvis associated with intrauterine contraceptive devices (IUCDs). Treatment is usually with penicillin or doxycycline.

**Gastrointestinal bacterial infections**

The differential diagnosis and approach to patients presenting with acute gastroenteritis are described on page 306.

**Staphylococcal food poisoning**

*Staph. aureus* transmission takes place via the hands of food handlers to foodstuffs such as dairy products, including cheese, and cooked meats. Inappropriate storage of these foods allows growth of the organism and production of one or more heat-stable enterotoxins which cause the symptoms.

Nausea and profuse vomiting develop within 1–6 hours. Diarrhoea may not be marked. The toxins that cause the syndrome act as ‘superantigens’, inducing a significant neutrophil leucocyteosis that may be clinically misleading. Superantigens are secreted proteins (enterotoxins) that exhibit highly potent lymphocyte-transforming (mitogenic) activity directed towards T lymphocytes. Most cases settle rapidly but severe dehydration can occasionally be life-threatening.

Antiemetics and appropriate fluid replacement are the mainstays of treatment. Suspect food should be cultured for staphylococci and demonstration of toxin production. The public health authorities should be notified if food vending is involved.

**Bacillus cereus food poisoning**

Ingestion of the pre-formed heat-stable exotoxins of *B. cereus* causes rapid onset of vomiting and some diarrhoea within hours of food consumption, which resolves within 24 hours. Fried rice and freshly made sauces are frequent sources; the organism grows and produces enterotoxin during storage (Fig. 13.24). If viable bacteria are ingested and toxin formation takes place within the gut lumen, then the incubation period is longer (12–24 hours) and watery diarrhoea and cramps are the predominant symptoms. The disease is self-limiting but can be quite severe.

Rapid and judicious fluid replacement and appropriate notification of the public health authorities are all that is required.

![Fig. 13.24 Bacillus cereus food poisoning.](image-url)
**INFECTIOUS DISEASE**

**Clostridium perfringens food poisoning**

Spores of *C. perfringens* are widespread in the guts of large animals and in soil. If contaminated meat products are incompletely cooked and stored in anaerobic conditions, *C. perfringens* spores germinate and viable organisms multiply to give large numbers. Subsequent reheating of the food causes heat-shock sporulation of the organisms, during which they release an enterotoxin. Symptoms (diarrhoea and cramps) occur some 6–12 hours following ingestion. The illness is usually self-limiting.

Clostridial enterotoxins are potent and most people who ingest them will be symptomatic. ‘Point source’ outbreaks, in which a number of cases all become symptomatic following ingestion, classically occur after school or canteen lunches where meat stews are served.

Clostridial necrotising enteritis (CNE) or pigbel is an often-fatal type of food poisoning caused by a β-toxin of *C. perfringens*, type C. The toxin is normally inactivated by certain proteases or by normal cooking. Pigbel is more likely in protein malnutrition or in the presence of trypsin inhibitors, either in foods such as sweet potatoes or during infection with *Ascaris* sp. roundworms.

**Campylobacter jejuni infection**

This infection is essentially a zoonosis, although contaminated water may be implicated, as the organism can survive for many weeks in fresh water. The most common sources of the infection are chicken, beef and contaminated milk products. There has been an association with pet puppies. *Campylobacter* infection is now the most common cause of bacterial gastroenteritis in the UK, accounting for some 100,000 cases per annum, most of which are sporadic.

The incubation period is 2–5 days. Colicky abdominal pain, which may be quite severe and mimic surgical pathology, occurs with nausea, vomiting and significant diarrhoea, frequently containing blood. The majority of *Campylobacter* infections affect fit young adults and are self-limiting after 5–7 days. About 10–20% will have prolonged symptomatology, occasionally meriting treatment with antibiotics such as erythromycin, as many organisms are resistant to ciprofloxacin.

Approximately 1% of cases will develop bacteraemia and possible distant foci of infection. *Campylobacter* spp. have been linked to Guillain–Barré syndrome and post-infectious reactive arthritis (pp. 1224 and 1107).

**Salmonella spp. infection**

*Salmonella* serotypes other than *S. typhi* and *S. paratyphi* (p. 339), of which there are more than 2000, are subdivided into five distinct subgroups which produce gastroenteritis. They are widely distributed throughout the animal kingdom. Two serotypes are most important worldwide: *S. enteritidis* phage type 4 and *S. typhimurium* dt104. The latter may be resistant to commonly used antibiotics such as ciprofloxacin. Some strains have a clear relationship to particular animal species, e.g. *S. arizonae* and pet reptiles. Transmission is by contaminated water or food, particularly poultry, egg products and related fast foods, direct person-to-person spread or the handling of exotic pets such as salamanders, lizards or turtles. The incidence of *Salmonella* enteritis is falling in the UK due to an aggressive culling policy in broiler chicken stocks, coupled with vaccination.

The incubation period of *Salmonella* gastroenteritis is 12–72 hours and the predominant feature is diarrhoea, sometimes with passage of blood. Vomiting may be present at the outset. Approximately 5% of cases are bacteraemic. Reactive (post-infective) arthritis occurs in approximately 2%.

Antibiotics are not indicated for uncomplicated *Salmonella* gastroenteritis (Box 13.48). However, evidence of bacteraemia is a clear indication for antibiotic therapy, as salmonellae are notorious for persistent infection and often colonise endothelial surfaces such as an atherosclerotic aorta or a major blood vessel. Mortality, as with other forms of gastroenteritis, is higher in the elderly (see Box 13.13, p. 306).

**Escherichia coli infection**

Many serotypes of *E. coli* are present in the human gut at any given time. Production of disease depends on either colonisation with a new or previously unrecognized strain, or the acquisition by current colonising bacteria of a particular pathogenicity factor for mucosal attachment or toxin production. Travel to unfamiliar areas of the world allows contact with different strains of endemic *E. coli* and the development of travellers’ diarrhoea. Enteropathogenic strains may be found in the gut of healthy individuals and, if these people move to a new environment, close contacts may develop symptoms.

At least five different clinico-pathological patterns of diarrhoea are associated with specific strains of *E. coli* with characteristic virulence factors.

**Enterotoxigenic E. coli**

Enterotoxigenic *E. coli* (ETEC) cause the majority of cases of travellers’ diarrhoea in developing countries, although there are other causes (see Box 13.21, p. 311). The organisms produce either a heat-labile or a heat-stable enterotoxin, causing marked secretory diarrhoea and vomiting after 1–2 days’ incubation. The illness is usually mild and self-limiting after 3–4 days. Antibiotics, such as ciprofloxacin, have been used to limit the duration of symptoms (see Box 13.22, p. 311) but are of questionable value.

**Enteroinvasive E. coli**

Illness caused by enteroinvasive *E. coli* (EIEC) is very similar to *Shigella* dysentery (p. 345) and is caused by invasion and destruction of colonic mucosal cells. No enterotoxin is produced. Acute watery diarrhoea, abdominal cramps and some scanty blood-staining of the stool are common. The symptoms are rarely severe and are usually self-limiting.

---

**EBM 13.48 Antibiotics in Salmonella gastroenteritis**

‘In otherwise healthy adults or children with non-severe *Salmonella* diarrhoea, antibiotics have no clinical benefit over placebo, increase side-effects and prolong *Salmonella* detection.’


For further information: [www.cochrane.org/cochrane-reviews](http://www.cochrane.org/cochrane-reviews)
Enteropathogenic E. coli

Enteropathogenic E. coli (EPEC) organisms are very important in infant diarrhoea. They are able to attach to the gut mucosa, inducing a specific ‘attachment and effacement’ lesion, and causing destruction of microvilli and disruption of normal absorptive capacity. The symptoms vary from mild non-bloody diarrhoea to quite severe illness, but without bacteraemia.

Enterohaemorrhagic E. coli

Enterohaemorrhagic E. coli (EAEC) strains adhere to the mucosa but also produce a locally active enterotoxin and demonstrate a particular ‘stacked brick’ aggregation to tissue culture cells when viewed by microscopy. They have been associated with prolonged diarrhoea in children in South America, South-East Asia and India.

Enterohaemorrhagic E. coli

A number of distinct ‘O’ serotypes of E. coli possess both the genes necessary for adherence (see ‘EPEC’ above) and plasmids encoding for two distinct enterotoxins (verotoxins) which are identical to the toxins produced by Shigella (‘shiga toxins 1 and 2’). E. coli O157:H7 is perhaps the best known of these verotoxin-producing E. coli (VTEC), but others, including types O126 and O11, are also implicated. In 2011, an outbreak of food-borne illness linked to fenugreek seeds occurred in Germany and was due to E. coli O104:H4, an EAEC strain that had acquired genes encoding shiga toxin 2a. Although the incidence of enterohaemorrhagic E. coli (EHEC) is considerably lower than that of Campylobacter and Salmonella infection, it is increasing in the developing world.

The reservoir of infection is in the gut of herbivores. The organism has an extremely low infecting dose (10–100 organisms). Runoff water from pasture lands where cattle have grazed, which is used to irrigate vegetable crops, as well as contaminated milk, meat products (especially hamburgers which have been incompletely cooked), lettuce, radish shoots and apple juice, have all been implicated as sources (Fig. 13.25).

The incubation period is between 1 and 7 days. Initial watery diarrhoea becomes frankly and uniformly blood-stained in 70% of cases and is associated with severe and often constant abdominal pain. There is little systemic upset, vomiting or fever.

Entorotoxins have both a local effect on the bowel and a distant effect on particular body tissues, such as glomerular apparatus, heart and brain. The potentially life-threatening haemolytic uraemic syndrome (HUS, p. 495) occurs in 10–15% of sufferers from this infection, arising 5–7 days after the onset of symptoms. It is most likely at the extremes of age, is heralded by a high peripheral leucocyte count, and may be induced, particularly in children, by antibiotic therapy.

HUS is treated by dialysis if necessary and may be averted by plasma exchange. Antibiotics should be avoided since they can stimulate toxin release.

Clostridium difficile infection

C. difficile is the most commonly diagnosed cause of antibiotic-associated diarrhoea (p. 308), and is an occasional constituent of the normal intestinal flora. C. difficile is capable of producing two toxins (A and B). C. difficile infection (CDI) usually follows antimicrobial therapy, which alters the composition of the gastrointestinal flora and may result in colonisation with C. difficile if the patient is exposed to C. difficile spores. The combination of toxin production and the ability to produce environmentally stable spores accounts for the clinical features and transmissibility of CDI. A hypervirulent strain of C. difficile, ribotype 027, has emerged, which produces more toxin than other C. difficile strains and thus more severe disease.

Clinical features

Disease manifestations range from diarrhoea to life-threatening pseudomembranous colitis. Around 80% of cases occur in people over 65 years of age, many of whom are frail with comorbid diseases. Symptoms usually begin in the first week of antibiotic therapy but can occur at any time up to 6 weeks after treatment has finished. The onset is often insidious, with lower abdominal pain and diarrhoea which may become profuse and watery. The presentation may resemble acute ulcerative colitis with bloody diarrhoea, fever and even toxic dilatation and perforation. Ileus is also seen in pseudomembranous colitis.

Investigations

C. difficile can be isolated from stool culture in 30% of patients with antibiotic-associated diarrhoea and over 90% of those with pseudomembranous colitis, but also from 5% of healthy adults and up to 20% of elderly patients in residential care. The diagnosis of CDI
therefore rests on detection of toxins A or B in the stool. Current practice in the UK is to screen stool from patients with a compatible clinical syndrome by detection either of glutamate dehydrogenase (GDH), an enzyme produced by *C. difficile*, or of *C. difficile* nucleic acid (e.g., by PCR); if screening is positive, a *C. difficile* toxin ELISA or a tissue culture cytotoxicity assay is performed.

The rectal appearances at sigmoidoscopy may be characteristic, with erythema, white plaques or an adherent pseudomembrane (Fig. 13.26). Appearances may also resemble those of ulcerative colitis. In some cases, the rectum is spared and abnormalities are observed in the proximal colon. Patients who are ill may require abdominal and erect chest X-rays to exclude perforation or toxic dilatation. CT may be useful when the diagnosis is in doubt.

**Management**

The precipitating antibiotic should be stopped and the patient should be isolated. Supportive therapy with intravenous fluids and resting of the bowel is often needed. CDI is treated with antibiotics. The options for first-line therapy are metronidazole (500 mg orally 3 times daily for 10 days) or vancomycin (125 mg orally 4 times daily for 7–10 days). Although vancomycin is more effective than metronidazole against hypervirulent *C. difficile* strains (e.g. ribotype 027), it is more expensive and may drive the emergence of vancomycin resistance in other organisms (e.g., enterococci, *Staph. aureus*). For these reasons, some authorities reserve its use for relapse (15–30% of patients), failure of initial response or severe infection. A new agent, fidaxomicin, is associated with a lower relapse rate than vancomycin. Intravenous immunoglobulin and/or corticosteroids are sometimes given in the most severe or refractory cases and faecal transplantation is also emerging as a therapy in relapsing patients. Surgical intervention is sometimes needed and needs to be considered early in severe cases.

**Yersinia enterocolitica infection**

This organism, commonly found in pork, causes mild to moderate gastroenteritis and can produce significant mesenteric adenitis after an incubation period of 3–7 days. It predominantly causes disease in children but adults may also be affected. The illness resolves slowly, with 10–30% of cases complicated by persistent arthritis or Reiter’s syndrome (p. 1107).

**Cholera**

Cholera, caused by *Vibrio cholerae* serotype O1, is the archetypal toxin-mediated bacterial cause of acute watery diarrhoea. The enterotoxin activates adenylate cyclase in the intestinal epithelium, inducing net secretion of chloride and water. *V. cholerae* O1 has two biotypes, classical and El Tor, and each of these has two distinct serotypes, Inaba and Ogawa. Following its origin in the Ganges valley, devastating epidemics have occurred, often in association with large religious festivals, and pandemics have spread worldwide. The seventh pandemic, due to the El Tor biotype, began in 1961 and spread via the Middle East to become endemic in Africa. In 1990, it reached Peru and spread throughout South and Central America. Since 2005, numbers of cases of cholera have been increasing. There are recurrent outbreaks and epidemics in Africa, often related to flooding. El Tor is more resistant to commonly used antimicrobials than classical *Vibrio*, and causes prolonged carriage in 5% of infections. A new classical toxigenic strain, serotype O139, established itself in Bangladesh in 1992 and started a new pandemic.

Infection spreads via the stools or vomit of symptomatic patients or of the much larger number of subclinical cases. Organisms survive for up to 2 weeks in fresh water and 8 weeks in salt water. Transmission is normally through infected drinking water, shellfish and food contaminated by flies, or on the hands of carriers.

**Clinical features**

Severe diarrhoea without pain or colic begins suddenly and is followed by vomiting. Following the evacuation of normal gut faecal contents, typical ‘rice water’ material is passed, consisting of clear fluid with flecks of mucus. Classical cholera produces enormous loss of fluid and electrolytes, leading to intense dehydration with muscular cramps. Shock and oliguria develop but mental clarity remains. Death from acute circulatory failure may occur rapidly unless fluid and electrolytes are replaced. Improvement is rapid with proper treatment.

The majority of infections, however, cause mild illness with slight diarrhoea. Occasionally, a very intense illness, ‘cholera sicca’, occurs, with loss of fluid into dilated bowel, killing the patient before typical gastrointestinal symptoms appear. The disease is more dangerous in children.

**Diagnosis and management**

Clinical diagnosis is easy during an epidemic. Otherwise, the diagnosis should be confirmed bacteriologically. Stool dark-field microscopy shows the typical ‘shooting star’ motility of *V. cholerae*. Rectal swab or stool cultures allow identification. Cholera is notifiable under international health regulations.

Maintenance of circulation by replacement of water and electrolytes is paramount (p. 307). Ringer-Lactate is the best fluid for intravenous replacement. Vomiting usually stops once the patient is rehydrated, and fluid should then be given orally up to 500 mL hourly. Early intervention with oral rehydration solutions that include resistant starch, based on either rice or cereal, shortens
the duration of diarrhoea and improves prognosis. Total fluid requirements may exceed 50 L over a period of 2–5 days. Accurate records are greatly facilitated by the use of a ‘cholera cot’, which has a reinforced hole under the patient’s buttocks, beneath which a graded bucket is placed.

Three days treatment with tetracycline 250 mg 4 times daily, a single dose of doxycycline 300 mg or ciprofloxacin 1 g in adults reduces the duration of excretion of *V. cholerae* and the total volume of fluid needed for replacement.

**Prevention**

Strict personal hygiene is vital and drinking water should come from a clean piped supply or be boiled. Flies must be denied access to food. Parenteral vaccination with a killed suspension of *V. cholerae* provides some protection. Oral vaccines containing killed *V. cholerae* and the B subunit of cholera toxin are available but are of limited efficacy.

In epidemics, public education and control of water sources and population movement are vital. Mass single-dose vaccination and treatment with tetracycline are valuable. Disinfection of discharges and soiled clothing, and scrupulous hand-washing by medical attendants reduce the danger of spread.

**Vibrio parahaemolyticus infection**

This marine organism produces a disease similar to enterotoxigenic *E. coli* (see above). It is acquired from raw seafood and is very common where ingestion of such food is widespread (e.g. Japan). After an incubation period of approximately 20 hours, explosive diarrhoea, abdominal cramps and vomiting occur. Systemic symptoms of headache and fever are frequent but the illness is self-limiting after 4–7 days. Rarely, a severe septicaemic illness arises; in this case, *V. parahaemolyticus* can be isolated using specific halophilic culture.

**Bacillary dysentery (shigellosis)**

Shigellae are Gram-negative rods, closely related to *E. coli*, that invade the colonic mucosa. There are four main groups: *Sh. dysenteriae*, *flexneri*, *boydii* and *sonnei*. In the tropics, bacillary dysentery is usually caused by *Sh. flexneri*, whilst in the UK most cases are caused by *Sh. sonnei*. Shigellae are often resistant to multiple antibiotics, especially in tropical countries. The organism only infects humans and its spread is facilitated by its low infecting dose of around 10 organisms.

Spread may occur via contaminated food or flies, but transmission by unwashed hands after defecation is by far the most important factor. Outbreaks occur in mental hospitals, residential schools and other closed institutions, and dysentery is a constant accompaniment of wars and natural catastrophes, which bring crowding and poor sanitation in their wake. *Shigella* infection may spread rapidly amongst men who have sex with men.

**Clinical features**

Disease severity varies from mild *Sh. sonnei* infections that may escape detection to more severe *Sh. flexneri* infections, while those due to *Sh. dysenteriae* may be fulminating and cause death within 48 hours.

In a moderately severe illness, the patient complains of diarrhoea, colicky abdominal pain and tenesmus. Stools are small, and after a few evacuations contain blood and purulent exudate with little faecal material. Fever, dehydration and weakness occur, with tenderness over the colon. Arthritis or iritis may occasionally complicate bacillary dysentery (Reiter’s syndrome, p. 1107), associated with HLA-B27.

**Management and prevention**

Oral rehydration therapy or, if diarrhoea is severe, intravenous replacement of water and electrolyte loss is necessary. Antibiotic therapy with ciprofloxacin (500 mg twice daily for 3 days) is effective in known shigellosis and appropriate in epidemics. The use of antidiarrhoeal medication should be avoided.

The prevention of faecal contamination of food and milk and the isolation of cases may be difficult, except in limited outbreaks. Hand-washing is very important.

**Respiratory bacterial infections**

Most of these infections are described in Chapter 19.

**Diphtheria**

Infection with *Corynebacterium diphtheriae* occurs most commonly in the upper respiratory tract and is usually spread by droplet infection. Infection may also complicate skin lesions, especially in those who misuse alcohol. The organisms remain localised at the site of infection but release of a soluble exotoxin damages the heart muscle and the nervous system.

Diphtheria has been eradicated from many parts of the world by mass vaccination using a modified exotoxin but remains important in areas where vaccination has been incomplete, e.g. in Russia and South-east Asia. The disease is notifiable in all countries of Europe and North America, and international guidelines have been issued by the WHO for the management of infection.

**Clinical features**

The average incubation period is 2–4 days. The disease begins insidiously with a sore throat (Box 13.49). Despite modest fever, there is usually marked tachycardia. The diagnostic feature is the ‘wash-leather’ elevated, greyish-green membrane on the tonsils. It has a well-defined edge, is firm and adherent, and is surrounded by a zone of inflammation. There may be swelling of the neck (‘bull neck’) and tender enlargement of the lymph nodes. In the mildest infections, especially in the presence of a high degree of immunity, a membrane may never appear and the throat is merely slightly inflamed.

### 13.49 Clinical features of diphtheria

<table>
<thead>
<tr>
<th>Acute infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membranous tonsillitis</td>
</tr>
<tr>
<td>or Nasal infection</td>
</tr>
<tr>
<td>or Laryngeal infection</td>
</tr>
<tr>
<td>or Skin/wound/conjunctival infection (rare)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laryngeal obstruction or paralysis</td>
</tr>
<tr>
<td>Myocarditis</td>
</tr>
<tr>
<td>Peripheral neuropathy</td>
</tr>
</tbody>
</table>
INFECTIOUS DISEASE

With anterior nasal infection there is nasal discharge, frequently blood-stained. In laryngeal diphtheria, a husky voice and high-pitched cough signal potential respiratory obstruction requiring urgent tracheostomy. If infection spreads to the uvula, fauces and nasopharynx, the patient is gravely ill.

Death from acute circulatory failure may occur within the first 10 days. Late complications arise as a result of toxin action on the heart or nervous system. About 25% of survivors of the early toxæmia may later develop myocarditis with arrhythmias or cardiac failure. These are usually reversible, with no permanent damage other than heart block in survivors.

Neurological involvement occurs in 75% of cases. After tonsillar or pharyngeal diphtheria, it usually starts after 10 days with palatal palsy. Paralysis of accommodation often follows, manifest by difficulty in reading small print. Generalised polynœuritis with weakness and paraesthesia may follow in the next 10–14 days. Recovery from such neuritis is always ultimately complete.

Management

A clinical diagnosis of diphtheria must be notified to the public health authorities and the patient sent urgently to a specialist infectious diseases unit. Treatment should begin once appropriate swabs have been taken before waiting for microbiological confirmation.

Diphtheria antitoxin is produced from hyperimmune horse serum. It neutralises circulating toxin but has no effect on toxin already fixed to tissues, so it must be injected intramuscularly without awaiting the result of a throat swab. However, reactions to this foreign protein include a potentially lethal immediate anaphylactic reaction (p. 91) and a ‘serum sickness’ with fever, urticaria and joint pains, which occurs 7–12 days after injection. A careful history of previous horse serum injections or allergic reactions should be taken and a small test injection of serum should be given half an hour before the full dose in every patient. Adrenaline (epinephrine) solution must be available to deal with any immediate type of reaction (0.5–1.0 mL of 1/1000 solution IM). An antihistamine is also given. In a severely ill patient, the risk of anaphylactic shock is outweighed by the mortal danger of diphtheritic toxæmia, and up to 100000 U of antitoxin are injected intravenously if the test dose has not given rise to symptoms. For disease of moderate severity, 16000–40000 U IM will suffice, and for mild cases 4000–8000 U.

Penicillin (1200 mg 4 times daily IV) or amoxicillin (500 mg 3 times daily) should be administered for 2 weeks to eliminate C. diptheria. Patients allergic to penicillin can be given erythromycin. Due to poor immunogenicity of primary infection, all sufferers should be immunised with diphtheria toxoid following recovery.

Patients must be managed in strict isolation and attended by staff with a clearly documented immunisation history until three swabs 24 hours apart are culture-negative.

Prevention

Active immunisation should be given to all children. If diphtheria occurs in a closed community, contacts should be given erythromycin, which is more effective than penicillin in eradicating the organism in carriers.

All contacts should also be immunised or given a booster dose of toxoid. Booster doses are required every 10 years to maintain immunity.

Pneumococcal infection

Strep. pneumoniae (the pneumococcus) is the leading cause of community-acquired pneumonia globally (p. 682) and one of the leading causes of infection-related mortality. Otitis media, meningitis and sinusitis are also frequently due to Strep. pneumoniae. Occasional patients present with bacteraemia without obvious focus. Asplenic individuals are at risk of fulminant pneumococcal disease with purpuric rash.

Increasing rates of penicillin resistance have been reported around the world for Strep. pneumoniae, although they remain relatively low in the UK. Strains with high-level resistance causing meningitis require treatment with glycopeptides rather than with penicillins or cephalosporins. Macrolide resistance is also increasing. Newer quinolones are also used but rates of resistance are rising.

Vaccination of infants with the protein conjugate pneumococcal vaccine decreases Strep. pneumoniae infection in infants and in their relatives. The polysaccharide pneumococcal vaccine is used in individuals predisposed to Strep. pneumoniae infection and the elderly, but only modestly reduces pneumococcal bacteraemia and does not prevent pneumonia. All asplenic individuals should receive vaccination against Strep. pneumoniae.

Anthrax

Anthrax is an endemic zoonosis in many countries; it causes human disease following inoculation of the spores of Bacillus anthracis. B. anthracis was the first recognised bacterial pathogen described by Koch and became the model pathogen for ‘Koch’s postulates’ (Box 6.1, p. 134). It is a Gram-positive organism with a central spore. The spores can survive for years in soil. Infection is commonly acquired from contact with animals, particularly herbivores. The ease of production of B. anthracis spores makes this infection a candidate for biological warfare or bioterrorism. B. anthracis produces a number of toxins which mediate the clinical features of disease.

Clinical features

These depend on the route of entry of the anthrax spores.

Cutaneous anthrax

This skin lesion is associated with occupational exposure to anthrax spores during processing of hides and bone products. It accounts for the vast majority of clinical cases. Animal infection is a serious problem in Africa, India, Pakistan and the Middle East.

Spores are inoculated into exposed skin. A single lesion develops as an irritable papule on an oedematous haemorrhagic base. This progresses to a depressed black eschar. Despite extensive oedema, pain is infrequent.

Gastrointestinal anthrax

This is associated with the ingestion of contaminated meat products. The caecum is the seat of the infection, which produces nausea, vomiting, anorexia and fever, followed in 2–3 days by severe abdominal pain and bloody diarrhoea. Toxaemia and death can develop rapidly thereafter.
Inhalational anthrax

This form of the disease is extremely rare but has been associated with bioterrorism. Without rapid and aggressive therapy at the onset of symptoms, the mortality is 50–90%. Fever, dyspnoea, cough, headache and symptoms of septicemia develop 3–14 days following exposure. Typically, the chest X-ray shows only widening of the mediastinum and pleural effusions, which are haemorrhagic. Meningitis may occur.

Management

*B. anthracis* can be cultured from skin swabs from lesions. Skin lesions are readily curable with early antibiotic therapy. Treatment is with ciprofloxacin (500 mg twice daily) until penicillin susceptibility is confirmed; the regimen can then be changed to benzylpenicillin with doses up to 2.4 g IV given 6 times daily or phenoxymethylpenicillin 500–1000 mg 4 times daily administered for 10 days. The addition of an aminoglycoside may improve the outlook in severe disease. In view of concerns about concomitant inhalational poisoning, particularly in the era of bioterrorism, a further 2-month course of ciprofloxacin 500 mg twice daily or doxycycline 100 mg twice daily orally is added. Prophylaxis with ciprofloxacin (500 mg twice daily) is recommended for anyone at high risk of exposure to anthrax spores.

**Bacterial infections with neurological involvement**

Infections affecting the CNS, including bacterial meningitis, botulism and tetanus, are described on page 1201.

**Mycobacterial infections**

Tuberculosis is predominantly, although by no means exclusively, a respiratory disease and is described on page 688.

**Leprosy**

Leprosy (Hansen’s disease) is a chronic granulomatous disease affecting skin and nerves, and is caused by *Mycobacterium leprae*, a slow-growing mycobacterium which cannot be cultured in vitro. The clinical manifestations are determined by the degree of the patient’s cell-mediated immunity (CMI, p. 78) towards *M. leprae* (Fig. 13.27). High levels of CMI with elimination of leprosy bacilli produces tuberculoid leprosy, whereas absent CMI results in lepromatous leprosy. The complications of leprosy are due to nerve damage, immunological reactions and bacillary infiltration. Leprosy patients are frequently stigmatised and using the word ‘leper’ is inappropriate.

**Epidemiology and transmission**

Some 4 million people have leprosy and around 750000 new cases are detected annually. About 70% of the world’s leprosy patients live in India, with the disease endemic in Brazil, Indonesia, Mozambique, Madagascar, Tanzania and Nepal.

Untreated lepromatous patients discharge bacilli from the nose. Infection occurs through the nose, followed by haematogenous spread to skin and nerve. The incubation period is 2–5 years for tuberculoid cases and 8–12 years for lepromatous cases. Leprosy incidence peaks at 10–14 years, and is more common in males and in those with close household exposure to leprosy cases.

**Pathogenesis**

*M. leprae* has a predilection for infecting Schwann cells and skin macrophages. In tuberculoid leprosy, effective CMI controls bacillary multiplication (‘paucibacillary’) but organised epithelioid granulomas are formed. In lepromatous leprosy, there is abundant bacillary multiplication (‘multibacillary’), e.g. in Schwann cells and perineurium. Between these two extremes is a continuum, varying from patients with moderate CMI (borderline tuberculoid) to patients with little cellular response (borderline lepromatous).

In addition, immunological reactions to the infection occur as the immune response develops and the antigenic stimulus from the bacilli varies, particularly in borderline patients. Delayed hypersensitivity reactions produce type 1 (reversal) reactions, while immune complexes contribute to type 2 (erythema nodosum leprous) reactions.

HIV/leprosy co-infected patients have typical lepromatous and tuberculoid leprosy skin lesions and typical leprosy histology and granuloma formation. Surprisingly, even with low circulating CD4 counts, tuberculoid leprosy may be observed and there is not an obvious shift to lepromatous leprosy.

**Clinical features**

Box 13.50 gives the cardinal features of leprosy. Types of leprosy are compared in Box 13.51.

• **Skin.** The most common skin lesions are macules or plaques. Tuberculoid patients have few,
13.50 **Cardinal features of leprosy**

- Skin lesions, typically anaesthetic at tuberculoid end of spectrum
- Thickened peripheral nerves
- Acid-fast bacilli on skin smears or biopsy

13.51 **Clinical characteristics of the polar forms of leprosy**

<table>
<thead>
<tr>
<th>Clinical and tissue-specific features</th>
<th>Tuberculoid</th>
<th>Lepromatous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skin and nerves</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number and distribution</td>
<td>One or a few sites, asymmetrical</td>
<td>Widely disseminated</td>
</tr>
<tr>
<td><strong>Skin lesions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definition</td>
<td>Common</td>
<td>Good</td>
</tr>
<tr>
<td>Elevation of margin</td>
<td>Never</td>
<td>Poor</td>
</tr>
<tr>
<td>Clarity of margin</td>
<td>Marked</td>
<td>Slight</td>
</tr>
<tr>
<td><strong>Colour</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark skin</td>
<td>Marked</td>
<td>Slight</td>
</tr>
<tr>
<td><strong>Light skin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypopigmentation</td>
<td>Marked</td>
<td>Slight</td>
</tr>
<tr>
<td>Slight erythema</td>
<td>Dry, scaly</td>
<td>Common</td>
</tr>
<tr>
<td>Smooth, shiny</td>
<td>Common</td>
<td>Impaired early</td>
</tr>
<tr>
<td>None</td>
<td>Impaired late</td>
<td>Impaired early</td>
</tr>
<tr>
<td>Impaired late</td>
<td>Impaired early</td>
<td>Impaired early</td>
</tr>
<tr>
<td><strong>Sweat and hair growth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss of sensation</td>
<td>Early and marked</td>
<td>Late</td>
</tr>
<tr>
<td><strong>Nerve enlargement and damage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late</td>
<td>Late</td>
<td>Late</td>
</tr>
<tr>
<td>Early and marked</td>
<td>Early and marked</td>
<td>Early and marked</td>
</tr>
<tr>
<td><strong>Bacilli (bacterial index)</strong></td>
<td>Absent (0)</td>
<td>Many (5 or 6+)</td>
</tr>
<tr>
<td><strong>Natural history</strong></td>
<td>Self-healing</td>
<td>Progressive</td>
</tr>
<tr>
<td><strong>Other tissues</strong></td>
<td>None</td>
<td>Upper respiratory mucosa, eye, testes, bones, muscle</td>
</tr>
<tr>
<td><strong>Reactions</strong></td>
<td>Cell-mediated (type 1)</td>
<td>Immune complexes (type 2)</td>
</tr>
</tbody>
</table>

**Fig. 13.28** Clinical features of leprosy.

**A** Tuberculoid leprosy. Single lesion with a well-defined active edge and anaesthesia within the lesion. **B** Lepromatous leprosy. Widespread nodules and infiltration, with loss of the eyebrows. This man also has early collapse of the nose. **C** Borderline tuberculoid leprosy with severe nerve damage. This boy has several well-defined, hypopigmented, macular, anaesthetic lesions. He has severe nerve damage affecting both ulnar and median nerves bilaterally and has sustained severe burns to his hands. **D** Reversal (type 1) reactions. Erythematous, oedematous lesions.

Hypopigmented lesions (Fig. 13.28A). In lepromatous leprosy, papules, nodules or diffuse infiltration of the skin occur. The earliest lesions are ill defined; gradually, the skin becomes infiltrated and thickened. Facial skin thickening leads to the characteristic leonine facies (Fig. 13.28B).

- Anaesthesia. In skin lesions, the small dermal sensory and autonomic nerve fibres are damaged, causing local sensory loss and loss of sweating within that area. Anaesthesia may also occur in the distribution of a damaged large peripheral nerve. A ‘glove and stocking’ sensory neuropathy is also common in lepromatous leprosy.

- Nerve damage. Peripheral nerve trunks are affected at ‘sites of predilection’. These are the ulnar (elbow), median (wrist), radial (humerus), radial cutaneous (wrist), common peroneal (knee), posterior tibial and sural nerves (ankle), facial nerve (zygomatic arch) and great auricular nerve (posterior triangle of the neck). Damage to peripheral nerve trunks produces characteristic signs with regional sensory loss and muscle dysfunction (Fig. 13.28C). All these nerves should be examined for enlargement and tenderness and tested for motor and sensory function. The CNS is not affected.

- Eye involvement. Blindness is a devastating complication for a patient with anaesthetic hands and feet. Eyelid closure is impaired when the facial
nerve is affected. Damage to the trigeminal nerve causes anaesthesia of the cornea and conjunctiva. The cornea is then susceptible to trauma and ulceration.

- **Other features.** Many organs can be affected. Nasal collapse occurs secondary to bacillary destruction of the bony nasal spine. Diffuse infiltration of the testes causes testicular atrophy and the acute orchitis that occurs with type 2 reactions. This results in azoospermia and hypogonadism.

### Leprosy reactions

Leprosy reactions (Box 13.52) are events superimposed on the cardinal features shown in Box 13.50.

- **Type 1 (reversal) reactions.** These occur in 30% of borderline patients (BT, BB or BL) and are delayed hypersensitivity reactions. Skin lesions become erythematous (Fig. 13.28D). Peripheral nerves become tender and painful, with sudden loss of nerve function. These reactions may occur spontaneously, after starting treatment and also after completion of multidrug therapy.

- **Type 2 (erythema nodosum leprosum, ENL) reactions.** These are partly due to immune complex deposition and occur in BL and LL patients who produce antibodies and have a high antigen load. They manifest with malaise, fever and crops of small pink nodules on the face and limbs. Iritis and episcleritis are common. Other signs are acute neuritis, lymphadenitis, orchitis, bone pain, dactylitis, arthritis and proteinuria. ENL may continue intermittently for several years.

### Borderline cases

In borderline tuberculoid (BT) cases, skin lesions are more numerous than in tuberculoid (TT) cases, and there is more severe nerve damage and a risk of type 1 reactions. In borderline leprosy (BB) cases, skin lesions are numerous and vary in size, shape and distribution; annular lesions are characteristic and nerve damage is variable. In borderline lepromatous (BL) cases, there are widespread small macules in the skin and widespread nerve involvement; both type 1 and type 2 reactions occur.

Pure neural leprosy (i.e. without skin lesions) occurs principally in India and accounts for 10% of patients. There is asymmetrical involvement of peripheral nerve trunks and no visible skin lesions. On nerve biopsy, all types of leprosy have been found.

### Investigations

The diagnosis is clinical, made by finding a cardinal sign of leprosy and supported by finding acid-fast bacilli in slit-skin smears or typical histology in a skin biopsy. Slit-skin smears are obtained by scraping dermal material on to a glass slide. The smears are then stained for acid-fast bacilli, the number counted per high-power field and a score derived on a logarithmic scale (0–6): the bacterial index (BI). Smears are useful for confirming the diagnosis and monitoring response to treatment. Neither serology nor PCR testing for *M. leprae* DNA is sensitive or specific enough for diagnosis.

### Management

The principles of treatment are outlined in Box 13.53. All leprosy patients should be given multidrug treatment (MDT) with an approved first-line regimen (Box 13.54).

---

13.52 Reactions in leprosy

<table>
<thead>
<tr>
<th>Lepra reaction type 1 (reversal)</th>
<th>Lepra reaction type 2 (erythema nodosum leprosum)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanism</strong></td>
<td>Cell-mediated hypersensitivity</td>
</tr>
<tr>
<td><strong>Clinical features</strong></td>
<td>Painful tender nerves, loss of function</td>
</tr>
<tr>
<td></td>
<td>Swollen skin lesions</td>
</tr>
<tr>
<td></td>
<td>New skin lesions</td>
</tr>
<tr>
<td><strong>Management</strong></td>
<td>Prednisolone 40 mg, reducing over 3–6 mths¹</td>
</tr>
<tr>
<td></td>
<td>Moderate: prednisolone 40 mg daily</td>
</tr>
<tr>
<td></td>
<td>Severe: thalidomide² or prednisolone 40–80 mg daily, reducing over 1–6 mths; local if eye involved³</td>
</tr>
</tbody>
</table>

1Indicated for any new impairment of nerve or eye function.
2Contraindicated in women who may become pregnant.
³1% hydrocortisone drops or ointment and 1% atropine drops.

13.53 Principles of leprosy treatment

- Stop the infection with chemotherapy
- Treat reactions
- Educate the patient about leprosy
- Prevent disability
- Support the patient socially and psychologically

13.54 Modified WHO-recommended multidrug therapy (MDT) regimens in leprosy

<table>
<thead>
<tr>
<th>Type of leprosy¹</th>
<th>Monthly supervised treatment</th>
<th>Daily self-administered treatment</th>
<th>Duration of treatment²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paucibacillary</td>
<td>Rifampicin 600 mg</td>
<td>Dapsone 100 mg</td>
<td>6 mths</td>
</tr>
<tr>
<td></td>
<td>Clofazimine 50 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ofloxacin 400 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minocycline 100 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multibacillary</td>
<td>Rifampicin 600 mg</td>
<td>Clofazimine 300 mg</td>
<td>12 mths</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paucibacillary</td>
<td>Dapsone 100 mg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>single-lesion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹Classification uses the bacillary index (BI) in slit-skin smears or, if BI is not available, the number of skin lesions:
- paucibacillary single-lesion leprosy (one skin lesion)
- paucibacillary (2–5 skin lesions)
- multibacillary (> 5 skin lesions).
²Studies from India have shown that multibacillary patients with an initial BI > 4 need longer treatment, for at least 24 mths.
INFECTIONOUS DISEASE

Rifampicin is a potent bactericidal for *M. leprae* but should always be given in combination with other antileprotics, since a single-step mutation can confer resistance. Dapsone is bacteriostatic. It commonly causes mild haemolysis and rarely anaemia. Clofazimine is a red, fat-soluble crystalline dye, weakly bactericidal for *M. leprae*. Skin discoloration (red to purple-black) and ichthyosis are troublesome side-effects, particularly on pale skins. New drugs that are bactericidal for *M. leprae* have been identified, notably the fluoroquinolones pefloxacin and ofloxacin, minocycline and clarithromycin. These agents are now established second-line drugs. Minocycline causes a grey pigmentation of skin lesions.

Although single-dose treatment is less effective than the conventional 6-month treatment for paucibacillary leprosy, it is an operationally attractive field regimen and has been recommended for use by the WHO.

Lepra reactions are treated as shown in Box 13.52. Chloroquine can also be used.

**Patient education**

Educating leprosy patients about their disease is vital. Patients should be reassured that, after 3 days of chemotherapy, they are not infectious and can lead a normal social life. It should be emphasised that gross deformities are not inevitable.

Patients with anaesthetic hands or feet need to take special care to avoid and treat burns and other minor injuries. Good footwear is important. Physiotherapy may be required to maintain range of movement of affected muscles and neighbouring joints.

**Prognosis**

Untreated, tuberculoid leprosy has a good prognosis; it may self-heal and peripheral nerve damage is limited. Lepromatous leprosy (LL) is a progressive condition with high morbidity if untreated.

After treatment, the majority of patients, especially those who have no nerve damage at the time of diagnosis, do well, with resolution of skin lesions. Borderline patients are at risk of developing type 1 reactions, which may result in devastating nerve damage.

**Prevention and control**

The previous strategy of centralised leprosy control campaigns has now been superseded by integrated programmes, with primary health-care workers in many countries now responsible for case detection and provision of MDT. It is not yet clear how successful this will be, especially in the time-consuming area of disability prevention.

BCG vaccination has been shown to give good but variable protection against leprosy; adding killed *M. leprae* to BCG does not enhance protection.

---

### Pathogenesis

The rickettsiae are intracellular Gram-negative organisms which parasitise the intestinal canal of arthropods. Infection is usually conveyed to humans through the skin from the excreta of arthropods, but the saliva of some biting vectors is infected. The organisms multiply in capillary endothelial cells, producing lesions in the skin, CNS, heart, lungs, liver, kidneys and skeletal muscles. Endothelial proliferation, associated with a perivascular reaction, may cause thrombosis and purpura. In epidemic typhus, the brain is the target organ; in scrub typhus, the cardiovascular system and lungs in particular are attacked. An eschar, a black necrotic crusted sore, is often found in tick- and mite-borne typhus (see Fig. 13.6C, p. 313). This is due to vasculitis following immunological recognition of the inoculated organism. Regional lymph nodes often enlarge.

### Spotted fever group

**Rocky Mountain spotted fever**

*Rickettsia rickettsii* is transmitted by tick bites. It is widely distributed and increasing in western and south-eastern states of the USA and also in Central and South America. The incubation period is about 7 days. The rash appears on about the third or fourth day of illness, looking at first like measles, but in a few hours a typical maculopapular eruption develops. The rash spreads in 24–48 hours from wrists, forearms and ankles to the back, limbs and chest, and then to the abdomen, where it is least pronounced. Larger cutaneous and subcutaneous haemorrhages may appear in severe cases. The liver and spleen become palpable. At the extremes of life, the mortality is 2–12%.

### Other spotted fevers

*R. conori* (boutonneuse fever) and *R. africae* (African tick fever) cause Mediterranean and African tick typhus, which also occurs on the Indian subcontinent. The incubation period is approximately 7 days. Infected ticks may be picked up by walking on grasslands, or dogs may bring ticks into the house. Careful examination might reveal a diagnostic eschar, and the maculopapular rash on the trunk, limbs, palms and soles. There may be delirium and meningeal signs in severe infections but recovery is usual. *R. africae* can be associated with multiple eschars. Some cases, particularly those with *R. africae*, present without rash (‘spotless spotted fever’). Other spotted fevers are shown in Box 13.55.

### Typhus group

**Scrub typhus fever**

Scrub typhus is caused by *Orientia tsutsugamushi* (formerly *Rickettsia tsutsugamushi*), transmitted by mites. It occurs in the Far East, Myanmar, Pakistan, Bangladesh, India, Indonesia, the South Pacific islands and Queensland, particularly where patches of forest cleared for plantations have attracted rats and mites.

In many patients, one eschar or more develops, surrounded by an area of cellulitis (see Fig. 13.6C, p. 313) and enlargement of regional lymph nodes. The incubation period is about 9 days.

Mild or subclinical cases are common. The onset of symptoms is usually sudden, with headache (often retro-orbital), fever, malaise, weakness and cough. In severe illness, the general symptoms increase, with

---

### Rickettsial and related intracellular bacterial infections

#### Rickettsial fevers

The rickettsial fevers are the most common tick-borne infections. It is important to ask potentially infected patients about contact with ticks, lice or fleas. There are two main groups of rickettsial fevers: spotted fevers and typhus (Box 13.55).
<table>
<thead>
<tr>
<th>Disease</th>
<th>Organism</th>
<th>Reservoir</th>
<th>Vector</th>
<th>Geographical area</th>
<th>Rash</th>
<th>Gangrene</th>
<th>Target organs</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spotted fever group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rocky Mountain</td>
<td><em>R. rickettii</em></td>
<td>Rodents, dogs, ticks</td>
<td><em>Ixodes</em></td>
<td>North, Central and South</td>
<td>Morbilliform</td>
<td>Often</td>
<td>Bronchi, myocardium, brain</td>
<td>2–12%²</td>
</tr>
<tr>
<td>spotted fever</td>
<td></td>
<td></td>
<td></td>
<td>America</td>
<td>Haemorrhagic</td>
<td></td>
<td>skin</td>
<td></td>
</tr>
<tr>
<td>Boutonneuse fever</td>
<td><em>R. conorii</em></td>
<td>Rodents, dogs, ticks</td>
<td><em>Ixodes</em></td>
<td>Mediterranean, Africa,</td>
<td>Maculopapular –</td>
<td>–</td>
<td>Skin, meninges</td>
<td>2.5%³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>South-west Asia, India</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siberian tick typhus</td>
<td><em>R. siberica</em></td>
<td>Rodents, birds, domestic</td>
<td>Various</td>
<td>Siberia, Mongolia,</td>
<td>Maculopapular –</td>
<td>–</td>
<td>Skin, meninges</td>
<td>Rare³</td>
</tr>
<tr>
<td></td>
<td></td>
<td>animals, ticks</td>
<td></td>
<td>northern China</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australian tick typhus</td>
<td><em>R. australis</em></td>
<td>Rodents, ticks</td>
<td>Ticks</td>
<td>Australia</td>
<td>Maculopapular –</td>
<td>–</td>
<td>Skin, meninges</td>
<td>Rare³</td>
</tr>
<tr>
<td>Oriental spotted</td>
<td><em>R. japonica</em></td>
<td>Rodents, dogs, ticks</td>
<td>Ticks</td>
<td>Japan</td>
<td>Maculopapular –</td>
<td>–</td>
<td>Skin, meninges</td>
<td>Rare³</td>
</tr>
<tr>
<td>fever</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African tick bite fever</td>
<td><em>R. africae</em></td>
<td>Cattle, game, ticks</td>
<td><em>Ixodes</em></td>
<td>South Africa</td>
<td>Can be spotless –</td>
<td>–</td>
<td>Skin, meninges</td>
<td>Rare³</td>
</tr>
</tbody>
</table>

| Typhus group            |                |                            |         |                           |                     |          |                            |             |
| Scrub typhus            | *Orientia      | Rodents                    | *Trombicula* | South-east Asia          | Maculopapular       | Unusual | Bronchi, myocardium, brain | Rare³       |
|                         | *tsutsugamushi*|                            | mite     |                           |                     |          | skin                       |             |
| Epidemic typhus         | *R. prowazekii*| Humans                     | Louse    | Worldwide                 | Morbilliform        | Often    | Brain, skin, bronchi,      | Up to 40%   |
|                         |                |                            |         |                           | Haemorrhagic        |          | myocardium                 |             |
| Endemic typhus          | *R. typhi*     | Rats                       | Flea     | Worldwide                 | Slight – – – – – – – – | –        | – – – – – – – – – – – – – – | – – – – – – – – |

¹Eschar at bite site and local lymphadenopathy. ²Highest in adult males. ³Except in infants, older people and the debilitated.

apathy and prostration. An erythematous maculopapular rash often appears on about the 5th–7th day and spreads to the trunk, face and limbs, including the palms and soles, with generalised painless lymphadenopathy. The rash fades by the 14th day. The temperature rises rapidly and continues as a remittent fever (i.e. the difference between maximum and minimum temperature exceeds 1°C), remaining above normal with sweating until it falls on the 12th–18th day. In severe infection, the patient is prostrate with cough, pneumonia, confusion and deafness. Cardiac failure, renal failure and haemorrhage may develop. Convalescence is often slow and tachycardia may persist for some weeks.

**Epidemic (louse-borne) typhus**

Epidemic typhus is caused by *R. prowazekii* and is transmitted by infected faeces of the human body louse, usually through scratching the skin. Patients suffering from epidemic typhus infect the lice, which leave when the patient is febrile. In conditions of overcrowding, the disease spreads rapidly. It is prevalent in parts of Africa, especially Ethiopia and Rwanda, and in the South American Andes and Afghanistan. Large epidemics have occurred in Europe, usually as a sequel to war. The incubation period is usually 12–14 days.

There may be a few days of malaise but the onset is more often sudden, with rigors, fever, frontal headaches, pains in the back and limbs, constipation and bronchitis. The face is flushed and cyanotic, the eyes are congested and the patient becomes confused. The rash appears on the 4th–6th day. In its early stages, it disappears on pressure but soon becomes petechial with subcutaneous mottling. It appears first on the anterior folds of the axillae, sides of the abdomen or backs of hands, then on the trunk and forearms. The neck and face are seldom affected. During the second week, symptoms increase in severity. Sores develop on the lips. The tongue becomes dry, brown, shrunk and tremulous. The spleen is palpable, the pulse feeble and the patient stuporous and delirious. The temperature falls rapidly at the end of the second week and the patient recovers gradually. In fatal cases, the patient usually dies in the second week from toxæmia, cardiac or renal failure, or pneumonia.

**Flea-borne (flea-borne) typhus**

Flea-borne or ‘endemic’ typhus caused by *R. typhi* is endemic worldwide. Humans are infected when the faeces or contents of a crushed flea, which has fed on an infected rat, are introduced into the skin. The incubation period is 8–14 days. The symptoms resemble those of a
mild louse-borne typhus. The rash may be scanty and transient.

**Investigation of rickettsial infection**

Routine blood investigations are not diagnostic but malaria must be excluded by blood film examination in most cases, and there is usually hepatitis and thrombocytopenia. Diagnosis is made on clinical grounds and response to treatment, and may be confirmed by antibody detection or PCR in specialised laboratories. Differential diagnoses include malaria, typhoid, meningococcal sepsis and leptospirosis.

**Management of rickettsial fevers**

The different rickettsial fevers vary greatly in severity but all respond to tetracycline 500 mg 4 times daily, doxycycline 200 mg daily or chloramphenicol 500 mg 4 times daily for 7 days. Louse-borne typhus and scrub typhus can be treated with a single dose of 200 mg doxycycline, repeated for 2–3 days to prevent relapse. Chloramphenicol- and doxycycline-resistant strains of *O. tsutsugamushi* have been reported from Thailand and patients here may need treatment with rifampicin.

Nursing care is important, especially in epidemic typhus. Sedation may be required for delirium and blood transfusion for haemorrhage. Relapsing fever and typhoid are common intercurrent infections in epidemic typhus, and pneumonia in scrub typhus. They must be sought and treated. Convalescence is usually protracted, especially in older people.

To prevent rickettsial infection, lice, fleas, ticks and mites need to be controlled with insecticides.

**Q fever**

Q fever occurs worldwide and is caused by the rickettsial-like organism *Coxiella burnetii*, an obligate intracellular organism that can survive in the extracellular environment. Cattle, sheep and goats are important reservoirs and the organism is transmitted by inhalation of aerosolised particles. An important characteristic of *C. burnetii* is its antigenic variation, called phase variation, due to a change of lipopolysaccharide (LPS). When isolated from animals or humans, *C. burnetii* expresses phase I antigen and is very infectious (a single bacterium is sufficient to infect a human). In culture, there is an antigenic shift to the phase II form, which is not infectious. This antigenic shift can be measured and is valuable for the differentiation of acute and chronic Q fever.

**Clinical features**

The incubation period is 3–4 weeks. The initial symptoms are non-specific with fever, headache and chills; in 20% of cases, a maculopapular rash occurs. Other presentations include pneumonia and hepatitis. Chronic Q fever may present with osteomyelitis, encephalitis and endocarditis.

**Investigations and management**

Diagnosis is usually serological and the stage of the infection can be distinguished by isotype tests and phase-specific antigens. Phase I and II IgM titres peak at 4–6 weeks. In chronic infections, IgG titres to phase I and II antigens may be raised.

Prompt treatment of acute Q fever with doxycycline reduces fever duration. Treatment of Q fever endocarditis is problematic, requiring prolonged therapy with doxycycline and rifampicin or ciprofloxacin with hydroxychloroquine; even then, organisms are not always eradicated. Valve surgery is often required (p. 629).

**Bartonellosis**

This group of diseases are caused by intracellular Gram-negative bacilli closely related to the rickettsiae, which have been discovered to be important causes of ‘culture-negative’ endocarditis. They are found in many domestic pets, such as cats, although for several the host is ill defined (Box 13.56). The principal human pathogens are *Bartonella quintana*, *B. henselae* and *B. bacilliformis*. Bartonella infections are associated with the following:

- **Trench fever.** This is a relapsing fever with severe leg pain and is caused by *B. quintana*. The disease is not fatal but is very debilitating.
- **Bacteremia and endocarditis in the homeless.** Endocarditis due to *B. quintana* or *henselae* is associated with severe damage to the heart valves.
- **Cat scratch disease.** *B. henselae* causes this common benign lymphadenopathy in children and young adults. A vesicle or papule develops on the head, neck or arms after a cat scratch. The lesion resolves spontaneously but there may be regional lymphadenopathy that persists for up to 4 months before also resolving spontaneously.
- **Bacillary angiomatosis.** This is an HIV-associated disease caused by *B. quintana* or *henselae* (p. 398).
- **Oroya fever and verruga peruana (Carrion’s disease).** This is endemic in areas of Peru. It is a biphasic disease caused by *B. bacilliformis* and is transmitted by sandflies of the genus *Phlebotomus*. Fever, haemolytic anaemia and microvascular thrombosis with end-organ ischaemia are features. It is frequently fatal if untreated.

**Investigations and management**

Bartonellae can be grown from the blood but this requires prolonged incubation using enriched media. Serum antibody detection is possible. *Bartonella* species are susceptible to β-lactams, rifampicin, erythromycin and tetracyclines. Antibiotic use is guided by clinical need. Cat scratch disease usually

### Box 13.56 Clinical diseases caused by *Bartonella* spp.

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Vector</th>
<th>Organism</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cats</td>
<td>Flea</td>
<td><em>B. henselae</em></td>
<td>Cat scratch disease, bacillary angiomatosis, endocarditis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undefined</td>
<td>Lice</td>
<td><em>B. quintana</em></td>
<td>Trench fever, bacillary angiomatosis, endocarditis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undefined</td>
<td>Sandfly</td>
<td><em>B. bacilliformis</em></td>
<td>Carrion's disease: Oroya fever and verruga peruana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undefined</td>
<td>Flea</td>
<td><em>B. rochalimae</em></td>
<td>Fever, rash, anaemia, splenomegaly</td>
</tr>
</tbody>
</table>
resolves spontaneously but Bartonella endocarditis requires valve replacement and combination antibiotic therapy.

**Chlamydial infections**

These are listed in Box 13.57 and are also described in Chapters 15 and 19.

**Trachoma**

Trachoma is a chronic keratoconjunctivitis caused by *Chlamydia trachomatis*, and is the most common cause of avoidable blindness. The classic trachoma environment is dry and dirty, causing children to have eye and nose discharges. Transmission occurs through flies, on fingers and within families. In endemic areas, the disease is most common in children.

**Pathology and clinical features**

The onset is usually insidious. Infection may be asymptomatic, lasts for years, may be latent over long periods and may recur. The conjunctiva of the upper lid is first affected with vascularisation and cellular infiltration. Early symptoms include conjunctival irritation and blepharospasm. The early follicles are characteristic (Fig. 13.29), but clinical differentiation from conjunctivitis due to other viruses may be difficult. Scarring causes inversion of the lids (entropion) so that the lashes rub against the cornea (trichiasis). The cornea becomes vascularised and opaque. The problem may not be detected until vision begins to fail.

**Box 13.57** Chlamydial infections

<table>
<thead>
<tr>
<th>Organism</th>
<th>Disease caused</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chlamydia trachomatis</em></td>
<td>Trachoma</td>
</tr>
<tr>
<td></td>
<td>Lymphogranuloma venereum (see Box 15.12, p. 424)</td>
</tr>
<tr>
<td></td>
<td>Cervicitis, urethritis, proctitis (p. 422)</td>
</tr>
<tr>
<td><em>Chlamydia psittaci</em></td>
<td>Psittacosis (Box 19.42, p. 683)</td>
</tr>
<tr>
<td><em>Chlamyphila (Chlamydia)</em></td>
<td>Atypical pneumonia (Box 19.42, p. 683)</td>
</tr>
<tr>
<td><em>pneumoniae</em></td>
<td>Acute/chronic sinusitis</td>
</tr>
</tbody>
</table>

**Investigations and management**

Intracellular inclusions may be demonstrated in conjunctival scrapings by staining with iodine or immunofluorescence. Chlamydia may be isolated in chick embryo or cell culture.

A single dose of azithromycin (20 mg/kg) has been shown to be superior to 6 weeks of tetracycline eye ointment twice daily for individuals in mass treatment programmes. Deformity and scarring of the lids, and corneal opacities, ulceration and scarring require surgical treatment after control of local infection.

**Prevention**

Personal and family cleanliness should be improved. Proper care of the eyes of newborn and young children is essential. Family contacts should be examined. The WHO is promoting the SAFE strategy for trachoma control (surgery, antibiotics, facial cleanliness and environmental improvement).

**PROTOZOAL INFECTIONS**

Protozoa are responsible for many important infectious diseases. They can be categorised according to whether they cause systemic or local infection. Trichomoniasis is described on page 417.

**Systemic protozoal infections**

**Malaria**

Malaria in humans is caused by *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and the predominantly simian parasite, *P. knowlesi*. It is transmitted by the bite of female anopheline mosquitoes and occurs throughout the tropics and subtropics at altitudes below 1500 metres (Fig. 13.30). Recent estimates have put the number of episodes of clinical malaria at 515 million cases per year, with two-thirds of these occurring in sub-Saharan Africa, especially amongst children and pregnant women. Following previous WHO-sponsored campaigns focusing on prevention and effective treatment, the incidence of malaria was greatly reduced between 1950 and 1960, but since 1970 there has been resurgence. Furthermore, *P. falciparum* has now become resistant to chloroquine and sulfadoxine-pyrimethamine, initially in South-east Asia and now throughout Africa. The WHO’s Millennium Development Goal 6 aims to halt the spread of the disease by 2015, and its ‘Roll Back Malaria’ campaign was designed to halve mortality by 2010 by utilising the ‘best evidence’ vector and disease control methods, such as artemisinin combination therapy (ACT).

**Fig. 13.29** Trachoma. Trachoma is characterised by hyperaemia and numerous pale follicles.

**Fig. 13.30** Distribution of malaria. (For up-to-date information see the Malaria Atlas Project (MAP): www.map.ox.ac.uk)
INFECTIONOUS DISEASE

Travellers are susceptible to malaria (p. 309). Due to increased travel, over 2000 cases are imported annually into the UK. Most are due to *P. falciparum*, usually from Africa, and of these 1% die because of late diagnosis. Immigrants returning home after visiting family and friends overseas but who have long-term residence in the UK are particularly at risk. They have lost their partial immunity and do not realise that they should be taking malaria prophylaxis. A few people living near airports in Europe have acquired malaria from accidentally imported mosquitoes.

**Pathogenesis**

*Life cycle of the malarial parasite*

The female anopheline mosquito becomes infected when it feeds on human blood containing gametocytes, the sexual forms of the malarial parasite (Figs 13.31 and 13.32). Development in the mosquito takes from 7 to 20 days, and results in sporozoites accumulating in the salivary glands and being inoculated into the human blood stream. Sporozoites disappear from human blood within half an hour and enter the liver. After some days, merozoites leave the liver and invade red blood cells, where further asexual cycles of multiplication take place, producing schizonts. Rupture of the schizont releases more merozoites into the blood and causes fever, the periodicity of which depends on the species of parasite.

*P. vivax* and *P. ovale* may persist in liver cells as dormant forms, hypnozoites, capable of developing into merozoites months or years later. Thus the first attack of clinical malaria may occur long after the patient has left the endemic area, and the disease may relapse after treatment if drugs that kill only the erythrocytic stage of the parasite are given.

*P. falciparum* and *P. malariae* have no persistent exo-erythrocytic phase but recrudescence of fever may result from multiplication of parasites in red cells which have not been eliminated by treatment and immune processes (Box 13.58).

**Pathology**

Red cells infected with malaria are prone to haemolysis. This is most severe with *P. falciparum*, which invades red cells of all ages but especially young cells; *P. vivax* and *P. ovale* invade reticulocytes, and *P. malariae*...
normoblasts, so that infections remain lighter. Anaemia may be profound and is worsened by dyserythropoiesis, splenomegaly and depletion of folate stores.

In *P. falciparum* malaria, red cells containing trophozoites adhere to vascular endothelium in post-capillary venules in brain, kidney, liver, lungs and gut by the formation of ‘knob’ proteins. They also form ‘rosettes’ and rouleaux with uninfected red cells. The vessels become congested, resulting in widespread organ damage which is exacerbated by rupture of schizonts, liberating toxic and antigenic substances (see Fig. 13.32).

*P. falciparum* has influenced human evolution, with the appearance of protective mutations such as sickle-cell (HbS; p. 1032), thalassaemia (p. 1034), G6PD deficiency (p. 1029) and HLA-B53. *P. falciparum* does not grow well in red cells that contain haemoglobin F, C or especially S. Haemoglobin S heterozygotes (AS) are protected against the lethal complications of malaria. *P. vivax* cannot enter red cells that lack the Duffy blood group; therefore many West Africans and African-Americans are protected.

**Clinical features**

The clinical features of malaria are non-specific and the diagnosis must be suspected in anyone returning from an endemic area who has features of infection.

**P. falciparum infection**

This is the most dangerous of the malarias and patients are either ‘killed or cured’. The onset is often insidious, with malaise, headache and vomiting. Cough and mild diarrhoea are also common. The fever has no particular pattern. Jaundice is common due to haemolysis and hepatic dysfunction. The liver and spleen enlarge and may become tender. Anaemia develops rapidly, as does thrombocytopenia.

A patient with *falciparum* malaria, apparently not seriously ill, may rapidly develop dangerous complications (Fig. 13.33 and Box 13.59). Cerebral malaria is manifested by confusion, seizures or coma, usually without localising signs. Children die rapidly without any special symptoms other than fever. Immunity is impaired in pregnancy and the parasite can preferentially bind to a placental protein known as chondroitin sulphate A. Abortion and intrauterine growth retardation from parasitisation of the maternal side of the placenta are frequent. Previous splenectomy increases the risk of severe malaria.

**P. vivax and *P. ovale* infection**

In many cases, the illness starts with several days of continued fever before the development of classical bouts of fever on alternate days. Fever starts with a rigor. The patient feels cold and the temperature rises to about 40°C. After half an hour to an hour, the hot or flush phase begins. It lasts several hours and gives way to profuse perspiration and a gradual fall in temperature. The cycle is repeated 48 hours later. Gradually, the spleen and liver enlarge and may become tender. Anaemia develops slowly. Relapses are frequent in the first 2 years after leaving the malarious area and infection may be acquired from blood transfusion.

**P. malariae infection**

This is usually associated with mild symptoms and bouts of fever every third day. Parasitaemia may persist for many years, with the occasional recrudescence of fever or without producing any symptoms. Chronic *P. malariae* infection causes glomerulonephritis and long-term nephrotic syndrome in children.

**Investigations**

Giemsa-stained thick and thin blood films should be examined whenever malaria is suspected. In the thick film, erythrocytes are lysed, releasing all blood stages of the parasite. This, as well as the fact that more blood is used in thick films, facilitates the diagnosis of low-level parasitaemia. A thin film is essential to confirm the diagnosis, to identify the species of parasite and, in *P. falciparum* infections, to quantify the parasite load (by counting the percentage of infected erythrocytes). *P. falciparum* parasites may be very scanty, especially in patients who have been partially treated. With *P. falciparum*, only ring forms are normally seen in the early stages (see Fig. 13.33); with the other species, all stages of the erythrocytic cycle may be found. Gametocytes appear after about 2 weeks, persist after treatment and are harmless, except that they are the source by which more mosquitoes become infected.

Immunochromatographic tests for malaria antigens, such as OptiMal® (which detects the *Plasmodium* lactate dehydrogenase of several species) and ParasightF® (which detects the *P. falciparum* histidine-rich protein 2), are extremely sensitive and specific for *falciparum* malaria but less so for other species. They should be used in parallel with blood film examination but are especially useful where the microscopist is less experienced in examining blood films (e.g. in the UK). The QBC Malaria Test is a fluorescence microscopy-based malaria diagnostic test which is also widely used.

DNA detection (PCR) is used mainly in research and is useful for determining whether a patient has a recrudescence of the same malaria parasite or a re-infection with a new parasite.
**Management**

**Mild P. falciparum malaria**

Since *P. falciparum* is now resistant to chloroquine and sulfadoxine-pyrimethamine (Fansidar) almost worldwide, an artemisinin-based treatment is recommended. Co-artemether (CoArtem® or Riamet®) contains artemether and lumefantrine and is given as 4 tablets at 0, 8, 24, 36, 48 and 60 hours. Alternatives are quinine by mouth (600 mg of quinine salt 3 times daily for 5–7 days), together with or followed by either doxycycline (200 mg once daily for 7 days) or clindamycin (450 mg 3 times daily for 7 days) or atovaquone-proguanil (Malarone®, 4 tablets once daily for 3 days). Doxycycline should not be used in pregnancy and artesunate should be avoided in early pregnancy.

WHO policy in Africa is moving towards always using artemisinin-based combination therapy (ACT), e.g. co-artemether or artesunate-amodiaquine. In India and other areas, artesunate (200 mg orally daily for 3 days) and mefloquine (1 g orally on day 2 and 500 mg orally on day 3) may be used. Unfortunately, artemisinin resistance has now been reported in Cambodia.

**Complicated P. falciparum malaria**

Severe malaria should be considered in any non-immune patient with a parasite count greater than 2% and is a medical emergency (see Box 13.59). Management includes early and appropriate antimalarial chemotherapy, active treatment of complications, correction of fluid, electrolyte and acid–base balance, and avoidance of harmful ancillary treatments.

The treatment of choice is intravenous artesunate given as 2.4 mg/kg IV at 0, 12 and 24 hours and then once daily for 7 days. However, as soon as the patient has recovered sufficiently to swallow tablets, oral artesunate 2 mg/kg once daily is given instead of intravenous therapy, to complete a total cumulative dose of 17–18 mg/kg. Rectal administration of artesunate is also being developed to allow administration in remote rural areas.

Quinine salt can also be used and is started with a loading dose infusion of 20 mg/kg over 4 hours, up to a maximum of 1.4 g. This is followed by maintenance doses of 10 mg/kg quinine salt given as 4-hour infusions 2–3 times daily, up to a maximum of 700 mg per dose,
13.59 Severe manifestations/complications of *falciparum* malaria and their immediate management

<table>
<thead>
<tr>
<th>Coma (cerebral malaria)</th>
<th>Spontaneous bleeding and coagulopathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Maintain airway</td>
<td>• Transfuse screened fresh whole blood (cryoprecipitate/fresh frozen plasma and platelets if available)</td>
</tr>
<tr>
<td>• Nurse on side</td>
<td>• Vitamin K injection</td>
</tr>
<tr>
<td>• Exclude other treatable causes of coma (e.g. hypoglycaemia, bacterial meningitis)</td>
<td>• Exclude or treat hypoglycaemia, hypovolaemia and Gram-negative septicaemia</td>
</tr>
<tr>
<td>• Avoid harmful ancillary treatments such as corticosteroids, heparin and adrenaline (epinephrine)</td>
<td>• Fluid resuscitation</td>
</tr>
<tr>
<td>• Intubate if necessary</td>
<td>• Give oxygen</td>
</tr>
</tbody>
</table>

**Hyperpyrexia**

- Tepid sponging, fanning, cooling blanket
- Antipyretic drug (paracetamol)

**Convulsions**

- Maintain airway
- Treat promptly with diazepam or paraldehyde injection

**Hypoglycaemia**

- Measure blood glucose
- Give 50% dextrose injection followed by 10% dextrose infusion (glucagon may be ineffective)

**Severe anaemia (packed cell volume < 15%)**

- Transfuse fresh whole blood or packed cells if pathogen screening of donor blood is available

**Acute pulmonary oedema**

- Nurse at 45°, give oxygen, venesect 250 mL of blood, give diuretic, stop intravenous fluids
- Intubate and add PEEP/CPAP (p. 193) in life-threatening hypoxaemia
- Haemofilter

**Acute renal failure**

- Exclude pre-renal causes
- Fluid resuscitation if appropriate
- Peritoneal dialysis (haemofiltration or haemodialysis if available)


until the patient can take drugs orally. The loading dose should not be given if the patient has received quinine, quinidine or mefloquine during the previous 24 hours. Patients should be monitored by ECG, with special attention to QRS duration and QT interval. Mefloquine should not be used for severe malaria since no parenteral form is available.

Exchange transfusion has not been tested in randomised controlled trials but may be beneficial for non-immune patients with persisting parasitaemias (>10% circulating erythrocytes).

**Management of non-falciparum malaria**

*P. vivax*, *P. ovale* and *P. malariae* infections should be treated with oral chloroquine: 600 mg chloroquine base, followed by 300 mg base in 6 hours, then 150 mg base twice daily for 2 more days. Some chloroquine resistance has been reported from Indonesia.

Late relapses can be prevented by prescribing antimarial drugs in suppressive doses. However, ‘radical cure’ is now achieved in most patients with *P. vivax* or *P. ovale* malaria using a course of primaquine (15 mg daily for 14 days), which destroys the hypnozoite phase in the liver. Haemolysis may develop in those who are G6PD-deficient. Cyanosis due to the formation of methaemoglobin in the red cells is more common but not dangerous.

**Prevention**

Clinical attacks of malaria may be preventable with chemoprophylaxis using chloroquine, atovaquone plus proguanil (Malarone), doxycycline or mefloquine. Box 13.60 gives the recommended doses for protection of the non-immune. The risk of malaria in the area to be visited and the degree of chloroquine resistance guide the recommendations for prophylaxis. Updated recommendations are summarised at www.fitfortravel.nhs.uk. Fansidar should not be used for chemoprophylaxis, as deaths have occurred from agranulocytosis or Stevens-Johnson syndrome (p. 1302). Mefloquine is useful in areas of multiple drug resistance, such as East and Central Africa and Papua New Guinea. Experience shows it to be safe for at least 2 years, but there are several contraindications to its use (see Box 13.60).

Expert advice is required for individuals unable to tolerate the first-line agents listed or in whom they are contraindicated. Mefloquine should be started 2–3 weeks before travel to give time for assessment of
13.60 Chemoprophylaxis of malaria

<table>
<thead>
<tr>
<th>Antimalarial tablets</th>
<th>Adult prophylactic dose</th>
<th>Regimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloroquine resistance high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mefloquine</td>
<td>250 mg weekly</td>
<td>Started 2–3 wks before travel and continued until 4 wks after travel</td>
</tr>
<tr>
<td>or Doxycycline</td>
<td>100 mg daily</td>
<td>Started 1 wk before and continued until 4 wks after travel</td>
</tr>
<tr>
<td>or Malarone</td>
<td>1 tablet daily</td>
<td>From 1–2 days before travel until 1 wk after return</td>
</tr>
<tr>
<td>Chloroquine resistance absent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloroquine and proguanil</td>
<td>300 mg base weekly</td>
<td>Started 1 wk before and continued until 4 wks after travel</td>
</tr>
<tr>
<td></td>
<td>100–200 mg daily</td>
<td></td>
</tr>
</tbody>
</table>

1Choice of regimen is determined by area to be visited, length of stay, level of malaria transmission, level of drug resistance, presence of underlying disease in the traveller and concomitant medication taken.
2Contraindicated in the first trimester of pregnancy, lactation, cardiac conduction disorders, epilepsy, psychiatric disorders; may cause neuropsychiatric disorders.
3Avoid in pregnancy.
4British preparations of chloroquine usually contain 150 mg base, French preparations 100 mg base and American preparations 300 mg base.

EBM 13.61 Prevention of malaria

Insecticide-treated bed nets (ITNs)

‘Five randomised controlled clinical trials provided strong evidence that widespread use of ITNs reduces overall mortality by about one-fifth in Africa. For every 1000 children aged 1–5 years protected, approximately 5.5 lives can be saved every year. In Africa, full ITN coverage could prevent 370000 child deaths per year.’

Electronic mosquito repellents (EMRs)

‘EMRs are not effective.’

Intermittent preventive treatment in pregnancy

‘Antimalarial drugs reduce antenatal parasitaemia and fever in pregnant women living in areas with endemic malaria. For women in their first or second pregnancy, this reduces the instances of severe antenatal anaemia, antenatal parasitaemia and perinatal deaths, and increases birth weight.’

Intermittent preventive treatment in children

‘Antimalarial drugs reduce clinical malaria, severe anaemia and hospital admissions. Effects on mortality and on health, if prophylaxis is stopped, are unknown.’


For further information: www.cochrane.org/cochrane-reviews

side-effects. Chloroquine should not be taken continuously as a prophylactic for more than 5 years without regular ophthalmic examination, as it may cause irreversible retinopathy. Pregnant and lactating women may take proguanil or chloroquine safely.

Prevention also involves advice about the use of high-percentage diethylenetriamine (DEET), covering up extremities when out after dark, and sleeping under permethrin-impregnated mosquito nets (Box 13.61).

Malaria control in endemic areas

There are major initiatives under way to reduce malaria in endemic areas and it is estimated that these would be cost-effective, even at a cost of $3 billion per year. Successful programmes have involved a combination of vector control, including indoor residual spraying, use of long-lasting insecticide-treated bed nets (ITNs) and intermittent preventative therapy (IPT; repeated dose of prophylactic drugs in high-risk groups, such as children and pregnant women) (see Box 13.61).

Development of a fully protective malaria vaccine is still some way off, which is not surprising, considering that natural immunity is incomplete and not long-lived. There is, however, some evidence that vaccination can reduce the incidence of severe malaria in populations. Trial vaccines are being evaluated in Africa.

Babesiosis

This is caused by a tick-borne intra-erythrocytic protozoan parasite. There are more than 100 species of Babesia, all of which have an animal reservoir, typically either rodents or cattle, and are transmitted to humans via the tick vector Ixodes scapularis. Most cases of babesiosis in the USA are due to B. microti and most in Europe to B. divergens. Patients present with fever and malaise 1–4 weeks after a tick bite. Illness may be complicated by haemolytic anaemia. Severe illness is seen in splenectomised patients. The diagnosis is made by blood-film examination. Treatment is with quinine and clindamycin.

African trypanosomiasis (sleeping sickness)

African sleeping sickness is caused by trypanosomes (Fig. 13.34) conveyed to humans by the bites of infected tsetse flies, and is unique to sub-Saharan Africa (Fig. 13.35). There has been a more than 60% decline in the incidence of sleeping sickness across Africa since 1990 due to better control measures. Trypanosoma brucei gambiense trypanosomiasis has a wide distribution in West and Central Africa and accounts for 90% of reported cases. T. brucei rhodesiense trypanosomiasis is found in parts of East and Central Africa. In West Africa, transmission is mainly at the riverside, where the fly rests in the shade of trees; no animal reservoir has been identified for T. gambiense. T. rhodesiense has a large reservoir in numerous wild animals and transmission takes place in the shade of woods bordering grasslands. Rural populations earning their livelihood from agriculture, fishing and animal husbandry are susceptible. Local
people and tourists visiting forests infested with tsetse flies and animal reservoirs may become infected.

**Clinical features**

A bite by a tsetse fly is painful and commonly becomes inflamed, but if trypanosomes are introduced, the site may again become painful and swollen about 10 days later (‘trypanosomal chancre’) and the regional lymph nodes enlarge (‘Winterbottom’s sign’). Within 2–3 weeks of infection, the trypanosomes invade the blood stream. The disease is characterised by an early haemolympathic stage and a late encephalitic stage, in which the parasite crosses the blood–brain barrier and chronic encephalopathy develops.

**Rhodesiense infections**

In these infections, the disease is more acute and severe than in gambiense infections, so that, within days or a few weeks, the patient is usually severely ill and may have developed pleural effusions and signs of myocarditis or hepatitis. There may be a petechial rash. The patient may die before there are signs of involvement of the CNS. If the illness is less acute, drowsiness, tremors and coma develop.

**Gambiense infections**

The distinction between early and late stages may not be apparent in gambiense infections. The disease usually runs a slow course over months or years, with irregular bouts of fever and enlargement of lymph nodes. These are characteristically firm, discrete, rubbery and painless, and are particularly prominent in the posterior triangle of the neck. The spleen and liver may become palpable. After some months without treatment, the CNS is invaded. This is shown clinically by headache and changed behaviour, blunting of higher mental functions, insomnia by night and sleepiness by day, mental confusion and eventually tremors, pareses, wasting, coma and death.

**Investigations**

Trypanosomiasis should be considered in any febrile patient from an endemic area. In rhodesiense infections, thick and thin blood films, stained as for the detection of malaria, will reveal trypanosomes. The trypanosomes may be seen in the blood or from puncture of the primary lesion in the earliest stages of gambiense infections, but it is usually easier to demonstrate them by aspiration of a lymph node. Concentration methods include buffy coat microscopy and miniature anion exchange chromatography.

Due to the cyclical nature of parasitaemia, the diagnosis is often made by demonstration of antibodies using a simple, rapid screening card agglutination trypanosomiasis test (CATT), followed by parasitological confirmation. If the CNS is affected, the cell count (> 20 x 10^9 leucocytes per litre) and protein content of the CSF are increased and the glucose is diminished. A very high level of serum IgM or the presence of IgM in the CSF is suggestive of trypanosomiasis. Recognition of CNS involvement is critical, as failure to treat it might be fatal.

**Management**

Unfortunately, therapeutic options for African trypanosomiasis are limited and most of the antitrypanosomal drugs are toxic and expensive. The prognosis is good if treatment is begun early, before the brain has been invaded. At this stage, intravenous suramin, after a test dose of 100–200 mg, should be given for rhodesiense infections (1 g on days 1, 3, 7, 14 and 21). For gambiense infections, intramuscular or intravenous pentamidine 4 mg/kg for 10 days is given (Box 13.62).
Once the nervous system is affected, treatment with melarsoprol (an arsenical) is effective for both East and West African diseases. It is used in a dose of 2–3.6 mg/kg/day IV for the first course and 3.6 mg/kg/day thereafter. Three 3-day treatment courses are given, separated by 7 days and by 10–21 days. Melarsoprol should be given with prednisolone 1 mg/kg up to 40 mg started 1–2 days before, continued during and tapered after treatment to reduce side-effects. Treatment-related mortality with melarsoprol is 4–12% due to reactive encephalopathy. For CNS infections due to gambiense, eflornithine (DFMO), an irreversible inhibitor of ornithine decarboxylase (100 and 150 mg/kg IV 4 times daily for 14 days for adults and children, respectively), is considered to be a safer and cost-effective option. Combinations of eflornithine (400 mg daily for 7 days) with oral nifurtimox (15 mg/kg daily for 15 days) have been shown to decrease relapses, deaths and drug toxicity.

**Prevention**

In endemic gambiense areas, various measures may be taken against tsetse flies, and field teams help to detect and treat early human infection. In rhodesiense areas, control is difficult.

**American trypanosomiasis (Chagas’ disease)**

Chagas’ disease occurs widely in South and Central America. The cause is *Trypanosoma cruzi*, transmitted to humans from the faeces of a reduviid (triatomine) bug, in which the trypanosomes have a cycle of development before becoming infective to humans. These bugs live in wild forests in crevices, burrows and palm trees. The *Triatoma infestans* bug has become domesticated in the Southern Cone countries (Argentina, Brazil, Chile, Paraguay and Uruguay). It lives in the mud and wattle walls and thatched roofs of simple rural houses, and emerges at night to feed and defecate on the sleeping occupants. Infected faeces are rubbed in through the conjunctiva, mucosa of mouth or nose, or abrasions of the skin. Over one hundred species of mammal, domestic, peridomestic and wild, may serve as reservoirs of infection. In some areas, blood transfusion accounts for about 5% of cases. Congenital transmission occasionally occurs.

**Pathology**

The trypanosomes migrate via the blood stream, develop into amastigote forms in the tissues and multiply intra-cellularly by binary fission. In the acute phase (primarily cell-mediated), inflammation of parasitised as well as non-parasitised cardiac muscles and capillaries occurs, resulting in acute myocarditis. In the chronic phase, focal myocardial atrophy, signs of chronic passive congestion and thromboembolic phenomena, cardiomegaly and apical cardiac aneurysm are salient findings. In the digestive form of disease, focal myositis and discontinuous lesions of the intramural myenteric plexus, predominantly in the oesophagus and colon, are seen.

**Clinical features**

**Acute phase**

Clinical manifestations of the acute phase are seen in only 1–2% of individuals who are infected before the age of 15 years. Young children (1–5 years) are most commonly affected. The entrance of *T. cruzi* through an abrasion produces a dusky-red firm swelling and enlargement of regional lymph nodes. A conjunctival lesion, although less common, is characteristic; the unilateral, reddish swelling of the lids may close the eye and constitutes ‘Romana’s sign’. In a few patients, an acute generalised infection soon appears, with a transient morbilliform or urticarial rash, fever, lymphadenopathy and enlargement of the spleen and liver. In a small minority of patients, acute myocarditis and heart failure or neurological features, including personality changes and signs of meningoencephalitis, may be seen. The acute infection may be fatal to infants.

**Chronic phase**

About 50–70% of infected patients become seropositive and develop an indeterminate form when no parasitaemia is detectable. They have a normal lifespan with no symptoms, but are a natural reservoir for the disease and maintain the life cycle of parasites. After a latent period of several years, 10–30% of chronic cases develop low-grade myocarditis, and damage to conducting fibres causes a cardiomyopathy characterised by cardiac dilatation, arrhythmias, partial or complete heart block and sudden death. In nearly 10% of patients, damage to Auerbach’s plexus results in dilatation of various parts of the alimentary canal, especially the colon and oesophagus, so-called ‘mega’ disease. Dilatation of the bile ducts and bronchi is also a recognised sequela. Autoimmune processes may be responsible for much of the damage. There are geographical variations of the basic pattern of disease. Reactivation of Chagas’ disease can occur in patients with HIV if the CD4 count falls lower than 200 cells/mm² (p. 393); this can cause space-occupying lesions with a presentation similar to Toxoplasma encephalitis, encephalitis, meningoencephalitis or myocarditis.

**Investigations**

*T. cruzi* is easily detectable in a blood film in the acute illness. In chronic disease, it may be recovered in up to 50% of cases by xenodiagnosis, in which infection-free, laboratory-bred reduviid bugs are allowed to feed on the patient; subsequently, the hind gut or faeces of the bug are examined for parasites. Parasite DNA detection by PCR in the patient’s blood is a highly sensitive method for documentation of infection and, in addition, can be employed in faeces of bugs used in xenodiagnosis tests to improve sensitivity. Antibody detection is also highly sensitive.

**Management and prevention**

Parasiticidal agents are used to treat the acute phase, congenital disease and early chronic phase (within 10 years of infection). Nifurtimox is given orally. The dose, which has to be carefully supervised to minimise toxicity while preserving parasiticidal activity, is 10 mg/kg divided into three equal doses, daily orally for 60–90 days. The paediatric dose is 15 mg/kg daily. Cure rates of 80% in acute disease are obtained. Benznidazole is an alternative, given at a dose of 5–10 mg/kg daily orally, in two divided doses for 60 days; children receive 10 mg/kg daily. Both nifurtimox and benznidazole are toxic, with adverse reaction rates of 30–55%. Specific drug treatment of the chronic form is now increasingly
favoured, but, in the cardiac or digestive ‘mega’ diseases, it does not reverse established tissue damage. Surgery may be needed.

Preventative measures include improving housing and destruction of reduviid bugs by spraying of houses with insecticides. Blood donors should be screened.

Toxoplasmosis

*Toxoplasma gondii* is an intracellular parasite. The sexual phase of the parasite’s life cycle (Fig. 13.36) occurs in the small intestinal epithelium of the domestic cat. Oöcysts are shed in cat faeces and are spread to intermediate hosts (pigs, sheep and also humans) through widespread contamination of soil. Oöcysts may survive in moist conditions for weeks or months. Once they are ingested, the parasite transforms into rapidly dividing tachyzoites through cycles of asexual multiplication. This leads to the formation of microscopic tissue cysts containing bradyzoites, which persist for the lifetime of the host. Cats become infected or re-infected by ingesting tissue cysts in prey such as rodents and birds.

Human acquisition of infection occurs via oöcyst-contaminated soil, salads and vegetables, or by the ingestion or tasting of raw or undercooked meats containing tissue cysts. Sheep, pigs and rabbits are the most common meat sources. Outbreaks of toxoplasmosis have been linked to the consumption of unfiltered water. In developed countries, toxoplasmosis is the most common protozoal infection; around 22% of adults in the UK are seropositive. Most primary infections are subclinical; however, toxoplasmosis is thought to account for about 15% of heterophile antibody-negative glandular fever (p. 320). In India or Brazil, approximately 40–60% of pregnant females are seropositive for *T. gondii*. In HIV-1 infection (p. 402), toxoplasmosis is an important opportunistic infection with considerable morbidity and mortality. Generalised toxoplasmosis has been described after accidental laboratory infection with highly virulent strains.

**Clinical features**

In most immunocompetent individuals, including children and pregnant women, the infection goes unnoticed. In approximately 10% of patients, it causes a self-limiting illness, most common in adults aged 25–35 years. The most common presenting feature is painless lymphadenopathy, either local or generalised. In particular, the cervical nodes are involved, but mediastinal, mesenteric or retroperitoneal groups may be affected. The spleen is seldom palpable. Most patients have no systemic symptoms, but some complain of malaise, fever, fatigue, muscle pain, sore throat and headache. Complete resolution usually occurs within a few months, although symptoms and lymphadenopathy tend to fluctuate unpredictably and some patients do not recover completely for a year or more. Very infrequently, patients may develop encephalitis, myocarditis, polymyositis, pneumonitis or hepatitis. Retinochoroiditis (Fig. 13.37) is nearly always the result of congenital infection but has also been reported in acquired disease.

**Congenital toxoplasmosis**

Acute toxoplasmosis, mostly subclinical, affects 0.3–1% of pregnant women, with an approximately 60% transmission rate to the fetus, which rises with increasing gestation. Seropositive females infected 6 months before conception have no risk of fetal transmission. Congenital disease affects approximately 40% of infected fetuses, and is more likely and more severe with infection early in gestation (see Box 13.28, p. 314). Many fetal infections are subclinical at birth but long-term sequelae include retinochoroiditis, microcephaly and hydrocephalus.

**Investigations**

In contrast to immunocompromised patients, in whom the diagnosis often requires direct detection of parasites, serology is often used in immunocompetent individuals.
**Infectious Disease**

The Sabin–Feldman dye test (indirect fluorescent antibody test), which detects IgG antibody, is most commonly used. Recent infection is indicated by a fourfold or greater increase in titre when paired sera are tested in parallel. Peak titres of 1/1000 or more are reached within 1–2 months of the onset of infection, and the dye test then becomes an unreliable indicator of recent infection. The detection of significant levels of *Toxoplasma*-specific IgM antibody may be useful in confirming acute infection. A false-positive result or persistence of IgM antibodies for years after infection makes interpretation difficult; however, negative IgM antibodies virtually rule out acute infection.

During pregnancy, it is critical to differentiate between recent and past infection; the presence of high-avidity IgG antibodies excludes infection acquired in the preceding 3–4 months.

If necessary, the presence of *Toxoplasma* organisms in a lymph node biopsy can be sought by staining sections histochemically with *T. gondii* antiserum, or by the use of PCR to detect *Toxoplasma*-specific DNA.

**Management**

In immunocompetent subjects, uncomplicated toxoplasmosis is self-limiting and responds poorly to antimicrobial therapy. Treatment with pyrimethamine, sulfadiazine and folinic acid is therefore usually reserved for rare cases of severe or progressive disease, and for infection in immunocompromised patients.

In a pregnant woman with an established recent infection, spiramycin (3 g daily in divided doses) should be given until term. Once fetal infection is established, treatment with sulfadiazine and pyrimethamine plus calcium folinate is recommended (spiramycin does not cross the placental barrier). The cost/benefit of routine *Toxoplasma* screening and treatment in pregnancy is being debated in many countries. There is insufficient evidence to determine the effects on mother or baby of current antiparasitic treatment for women who seroconvert in pregnancy.

**Leishmaniasis**

Leishmaniasis is caused by unicellular, flagellate, intracellular protozoa belonging to the genus *Leishmania* (order Kinetoplastidae). There are 21 leishmanial species that cause several diverse clinical syndromes, which can be placed into three broad groups:

- visceral leishmaniasis (VL, kala-azar)
- cutaneous leishmaniasis (CL)
- mucosal leishmaniasis (ML).

**Epidemiology and transmission**

Although most clinical syndromes are caused by zoonotic transmission of parasites from animals (chiefly canine and rodent reservoirs) to humans through phlebotomine sandfly vectors (Fig. 13.38A), humans are the only known reservoir (anthroponotic) in major VL foci in the Indian subcontinent and for transmission of leishmaniasis between injection drug-users (Fig. 13.38B and C). Leishmaniasis occurs in approximately 100 countries around the world, with an estimated annual incidence of 2 million new cases (500 000 for VL and 1.5 million for CL).

**Visceral leishmaniasis** (kala-azar)

VL is caused by the protozoon *Leishmania donovani* complex (comprising *L. donovani*, *L. infantum* and *L. chagasi*). India, Sudan, Bangladesh and Brazil account for 90% of cases of VL, while other affected regions include...
Pancytopenia affected; improved the and strip can VL weeks transmission, the visceral by with the main features. Clinical may a infants, quickly massive developed ELISA pedal except of to and oedema, (gross DNA HIV dominant retina, buffy test, k39 illness, patients. and relapse test bleeding develops disease in liver, the children is is immuno the generalised gas coat often Indian Sensitivity been in the spleen, progress. In manifest and diseases, as aspirate have features. PCR Spleno is common a lesser small VL cardiac few by and the Fig. 13.39 months transplant. The great by nose. countries. bone and Sandfly and associated in eastern hemisphere, Phlebotomus in western hemisphere. Lutzomyia and Psychodopygus

**Fig. 13.39 Life cycle of Leishmania.** From Knight 1982 – see p. 386.

![Leishmania life cycle diagram](image)

- **Dermis only** L. tropica etc.
- **Dermis and mucosae** (sometimes) L. brasilensis
- **Viscera and dermis** (sometimes) L. donovani

**Fig. 13.40 World distribution of visceral leishmaniasis.**

![World distribution map](image)

- L. chagasi
- L. infantum
- L. donovani

the Mediterranean, East Africa, China, Arabia, Israel and other South American countries (Fig. 13.40). In addition to sandfly transmission, VL has also been reported to follow blood transfusion, and disease can present unexpectedly in immunosuppressed patients – for example, after renal transplantation and in HIV infection.

The great majority of people infected remain asymptomatic. In visceral diseases, the spleen, liver, bone marrow and lymph nodes are primarily involved.

**Clinical features**

In the Indian subcontinent, adults and children are equally affected; elsewhere, VL is mainly a disease of small children and infants, except in adults with HIV co-infection. The incubation period ranges from weeks to months (occasionally several years).

The first sign of infection is high fever, usually accompanied by rigor and chills. Fever intensity decreases over time, and patients may become afebrile for intervening periods ranging from weeks to months. This is followed by a relapse of fever, often of lesser intensity. Splenomegaly develops quickly in the first few weeks and becomes massive as the disease progresses. Moderate hepatomegaly occurs later. Lymphadenopathy is seen in the majority of cases in Africa, the Mediterranean and South America, but is rare in the Indian subcontinent. Blackish discoloration of the skin, from which the disease derived its name, kala-azar (the Hindi word for ‘black fever’), is a feature of advanced illness and is now rarely seen. Pancytopenia is a common feature. Moderate to severe anaemia develops rapidly, and can result in congestive cardiac failure and associated clinical features. Thrombocytopenia, often compounded by hepatic dysfunction, may result in bleeding from the retina, gastrointestinal tract and nose. In advanced illness, hypoalbuminaemia may manifest as pedal oedema, ascites and anasarca (gross generalised oedema and swelling).

As the disease advances, there is profound immunosuppression and secondary infections are very common. These include tuberculosis, pneumonia, severe amoebic or bacillary dysentery, gastroenteritis, herpes zoster and chickenpox. Skin infections, boils, cellulitis and scabies are common. Without adequate treatment, most patients with clinical VL die.

**Investigations**

Pancytopenia is the most dominant feature, with granulocytopenia and monocytosis. Polyclonal hypergammaglobulinaemia, chiefly IgG followed by IgM, and hypoalbuminaemia are seen later.

Demonstration of amastigotes (Leishman–Donovan bodies) in splenic smears is the most efficient means of diagnosis, with 98% sensitivity (Fig. 13.41); however, it carries a risk of serious haemorrhage in inexperienced hands. Safer methods, such as bone marrow or lymph node smears, are not as sensitive. Parasites may be demonstrated in buffy coat smears, especially in immunosuppressed patients. Sensitivity can be improved by culturing the aspirate material or by PCR for DNA detection and species identification, but these tests can only be performed in specialised laboratories.

Serodiagnosis, by ELISA or immunofluorescence antibody test, is employed in developed countries. In endemic regions, a highly sensitive direct agglutination test using stained promastigotes and an equally efficient rapid immunochromatographic k39 strip test have become popular. These tests remain positive for several months after cure has been achieved, so do not predict...
response to treatment or relapse. The vast majority of people exposed to the parasite do not develop clinical illness but may have positive serological tests thereafter. Formal gel (aldehyde) or other similar tests based on the detection of raised globulin have limited value and should not be employed for the diagnosis of VL.

**Differential diagnosis**

This includes malaria, typhoid, tuberculosis, schistosomiasis and many other infectious and neoplastic conditions, some of which may coexist with VL. Fever, splenomegaly, pancytopenia and non-response to antimalarial therapy may provide clues before specific laboratory diagnosis is made.

**Management**

**Pentavalent antimonials**

Antimony (Sb) compounds were the first drugs to be used for the treatment of leishmaniasis and remain the mainstay of treatment in most parts of the world. The exception is the Indian subcontinent, especially Bihar state, where almost two-thirds of cases are refractory to Sb treatment. Traditionally, pentavalent antimony is available as sodium stibogluconate (100 mg/mL) in English-speaking countries and meglumine antimonate (85 mg/mL) in French-speaking ones. The daily dose is 20 mg/kg body weight, given either intravenously or intramuscularly for 28–30 days. Side-effects are common and include arthralgias, myalgias, raised hepatic transaminases, pancreatitis (especially in patients co-infected with HIV) and ECG changes (T wave inversion and reduced amplitude). Severe cardiotoxicity, manifest by concave ST segment elevation, prolongation of QT, greater than 0.5 msec, and ventricular dysrhythmias, is not uncommon. The incidence of cardiotoxicity and death can be very high with improperly manufactured Sb.

**Amphotericin B**

Amphotericin B deoxycholate, given once daily or on alternate days at a dose of 0.75–1.00 mg/kg for 15–20 doses, is used as the first-line drug in many regions where there is a significant level of Sb unresponsiveness. It has a cure rate of nearly 100%. Infusion-related side-effects, e.g. high fever with rigor, thrombophlebitis, diarrhea and vomiting, are extremely common. Serious adverse events, such as renal or hepatic toxicity, hypokalaemia and thrombocytopenia, are not uncommon.

Lipid formulations of amphotericin B (p. 160) are less toxic. AmBisome is approved by the US Food and Drug Administration and is first-line therapy in Europe for VL. Drug doses vary according to geographical location. In the Indian subcontinent, a total dose of 10 or 15 mg/kg, administered in a single dose or as multiple doses over several days, respectively, is considered adequate, whereas in Africa 14–18 mg, and in South America and Europe 21–24 mg, in divided doses, typically spread over 10 days, is needed for immunocompetent patients. High daily doses of the lipid formulations are well tolerated, and in one study a single dose of 10 mg/kg of AmBisome cured 96% of Indian patients. AmBisome has been made available at a preferential low price for developing countries, and greater use of this drug for treatment of VL is expected.

**EBM 13.63 Combination therapy for visceral leishmaniasis**

‘In India, combinations of a single dose of liposomal amphotericin B (AmBisome, 5 mg/kg) with either 7 days of miltefosine or 10 days of paromomycin, or 10 days each of miltefosine and paromomycin, cured at least 97% of patients.’

‘In Sudan, a combination of 17 days of antimony (Sb) with paromomycin produces a similar rate of cure (> 90%) to conventional treatment with 30 days of antimony.’


**Other drugs**

The oral drug miltefosine, an alkyl phospholipid, has been approved in several countries for the treatment of VL. A daily dose of 50 mg (patient’s body weight < 25 kg) to 100 mg (≥ 25 kg), or 2.5 mg/kg body weight for children, for 28 days cures over 90% of patients. Side-effects include mild to moderate vomiting and diarrhoea, and rarely skin allergy or renal or liver toxicity. Since it is a teratogenic drug, it cannot be used in pregnancy; female patients are advised not to become pregnant for the duration of treatment and 3 months thereafter, because of its half-life of nearly 1 week.

Paromomycin is an aminoglycoside that has undergone trials in India and Africa, and is highly effective if given intramuscularly at 11 mg/kg body weight of paromomycin base, daily for 3 weeks. No significant auditory or renal toxicity is seen. The drug has been approved in India for the treatment of VL.

Pentamidine isethionate was used to treat Sb-refractory patients with VL. However, declining efficacy and serious side-effects, such as type 1 diabetes mellitus, hypoglycaemia and hypotension, have led to it being abandoned.

Multidrug therapy of VL is likely to be used increasingly to prevent emergence of drug resistance (Box 13.63).

**Response to treatment**

A good response results in abatement of fever, a feeling of well-being, gradual decrease in spleen size, weight gain and recovery of blood counts. Patients should be followed regularly for a period of 6–12 months, as a small minority may experience a relapse of the disease during this period, irrespective of the treatment regimen.

**HIV–visceral leishmaniasis co-infection**

HIV-induced immunosuppression (Ch. 14) increases the risk of contracting VL 100–1000 times. Most cases of HIV–VL co-infection have been reported from Spain, France, Italy and Portugal. Antiretroviral therapy (ART) has led to a remarkable decline in the incidence of VL co-infection in Europe. However, numbers are increasing in Africa (mainly Ethiopia), Brazil and in the Indian subcontinent.

Although the clinical triad of fever, splenomegaly and hepatomegaly is found in the majority of co-infected patients, those with low CD4 count may have atypical clinical presentations, posing a diagnostic challenge. VL may present with gastrointestinal involvement (stomach, duodenum or colon), ascites, pleural or pericardial
In effusion, or involvement of lungs, tonsil, oral mucosa or skin. Diagnostic principles remain the same as those in non-HIV patients. Parasites are numerous and easily demonstrable, even in buffy coat preparations. Sometimes amastigotes are found in unusual sites, such as bronchoalveolar lavage fluid, pleural fluid or biopsies of the gastrointestinal tract. Immunofluorescence, Western blot, ELISA and other serological tests used singly have low sensitivity. DNA detection by PCR of the blood or its buffy coat are at least 95% sensitive, and accurately track recovery and relapse.

Treatment of VL with HIV co-infection is essentially the same as in immunocompetent patients but there are some differences in outcome. Conventional amphotericin B (0.7 mg/kg/day for 28 days) may be more effective in achieving initial cure than Sb (20 mg/kg/day for 28 days). Using high-dose liposomal amphotericin B (4 mg/kg on days 1–5, 10, 17, 24, 31 and 38), a high cure rate is possible. However, these co-infected patients have a tendency to relapse within 1 year. For prevention of relapse, maintenance chemotherapy with monthly liposomal amphotericin B is useful.

Post-kala-azar dermal leishmaniasis

After treatment and apparent recovery from the visceral disease in India and Sudan, some patients develop dermatological manifestations due to local parasitic infection.

Clinical features

In India, dermatological changes occur in a small minority of patients 6 months to at least 3 years after the initial infection. They are seen as macules, papules, nodules (most frequently) and plaques, which have a predilection for the face, especially the area around the chin. The face often appears erythematous (Fig. 13.42A). Hypopigmented macules can occur over all parts of the body and are highly variable in extent and location. There are no systemic symptoms and no spontaneous healing.

In Sudan, approximately 50% of patients with VL develop post-kala-azar dermal leishmaniasis (PKDL), experiencing skin manifestations concurrently with VL or within the following 6 months. In addition to the dermatological features described above, a measles-like micropapular rash (Fig. 13.42B) may be seen all over the body. In Sudan, children are more frequently affected than in India. Spontaneous healing occurs in about three-quarters of cases within 1 year.

Investigations and management

The diagnosis is clinical, supported by demonstration of scanty parasites in lesions by slit-skin smear and culture. Immunofluorescence and immunohistochemistry may demonstrate the parasite in skin tissues. In the majority of patients, serological tests (direct agglutination test or k39 strip tests) are positive.

Treatment of PKDL is difficult. In India, Sb for 120 days, several courses of amphotericin B infusions, or miltefosine for 12 weeks is required. In Sudan, Sb for 2 months is considered adequate. In the absence of a physical handicap, most patients are reluctant to complete the treatment. PKDL patients are a human reservoir, and focal outbreaks have been linked to patients with PKDL in areas previously free of VL.

Prevention and control

Sandflies are extremely sensitive to insecticides, and vector control through insecticide spray is very important. Mosquito nets or curtains treated with insecticides will keep out the tiny sandflies. In endemic areas with zoonotic transmission, infected or stray dogs should be destroyed.

In areas with anthropogenic transmission, early diagnosis and treatment of human infections, to reduce the reservoir and control epidemics of VL, is extremely important. Serology is useful in diagnosis of suspected cases in the field. No vaccine is currently available.

Cutaneous and mucosal leishmaniasis

Cutaneous leishmaniasis

CL (oriental sore) occurs in both the Old World and the New World (the Americas). Transmission is described on page 362.

Fig. 13.42 Post-kala-azar dermal leishmaniasis. A In India, with macules, papules, nodules and plaques. From Sundar S, et al. 2006 – see p. 386. B In Sudan, with micronodular rash.
In the Old World, CL is mild. It is found around the Mediterranean basin, throughout the Middle East and Central Asia as far as Pakistan, and in sub-Saharan West Africa and Sudan (Fig. 13.43). The causative organisms for Old World zoonotic CL are L. major, L. tropica and L. aethiopica (Box 13.64). Anthroponotic CL is caused by L. tropica, and is confined to urban or suburban areas of the Old World. Afghanistan is currently the biggest focus, but infection is endemic in Pakistan, the western deserts of India, Iran, Iraq, Syria and other areas of the Middle East. In recent years, there has been an increase in the incidence of zoonotic CL in both the Old and the New World due to urbanisation and deforestation which led to peridomestic transmission (in and around human dwellings).

New World CL is a more significant disease, which may disfigure the nose, ears and mouth, and is caused by the L. mexicana complex (comprising L. mexicana, L. amazonensis and L. venezuelensis) and by the Viannia subgenus L. (V.) aethiopica complex (comprising L. (V.) guyanensis, L. (V.) panamensis, L. (V.) brasiliensis and L. (V.) peruviana).

<table>
<thead>
<tr>
<th>Leishmania spp.</th>
<th>Host</th>
<th>Clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. tropica</td>
<td>Dogs</td>
<td>Slow evolution, less severe</td>
</tr>
<tr>
<td>L. major</td>
<td>Gerbils, desert rodents</td>
<td>Rapid necrosis, wet sores</td>
</tr>
<tr>
<td>L. aethiopica</td>
<td>Hyraxes</td>
<td>Solitary facial lesions with satellites</td>
</tr>
</tbody>
</table>

CL is commonly imported and should be considered in the differential diagnosis of an ulcerating skin lesion, especially in travellers who have visited endemic areas of the Old World or forests in Central and South America.

**Pathogenesis**

Inoculated parasites are taken up by dermal macrophages, in which they multiply and form a focus for lymphocytes, epithelioid cells and plasma cells. Self-healing may occur with necrosis of infected macrophages, or the lesion may become chronic with ulceration of the overlying epidermis, depending upon the aetiological pathogen.

**Clinical features**

The incubation period is typically 2–3 months (range 2 weeks to 5 years). In all types of CL, the common feature is development of a papule followed by ulceration of the skin with raised borders, usually at the site of the bite of the vector. Lesions, single or multiple, start as small red papules that increase gradually in size, reaching 2–10 cm in diameter. A crust forms, overlying an ulcer with a granular base (Fig. 13.44). These ulcers develop a few weeks or months after the bite. There can be satellite lesions, especially in L. major and occasionally in L. tropica infections. Regional lymphadenopathy, pain, pruritus and secondary bacterial infections may occur.

Clinically, lesions of L. mexicana and L. peruviana closely resemble those seen in the Old World, but lesions on the pinna of the ear are common, and are chronic and destructive. L. mexicana is responsible for chiclero ulcers, the self-healing sores of Mexico.

If immunity is good, there is usually spontaneous healing in L. tropica, L. major and L. mexicana lesions. In some patients with anergy to Leishmania, the skin lesions of L. aethiopica, L. mexicana and L. amazonensis infections progress to the development of diffuse CL; this is characterised by spread of the infection from the initial ulcer, usually on the face, to involve the whole body in the form of non-ulcerative nodules. Occasionally, in L. tropica infections, sores that have apparently healed relapse persistently (recidivans or lupoid leishmaniasis).

**Mucosal leishmaniasis**

The Viannia subgenus extends widely from the Amazon basin as far as Paraguay and Costa Rica, and is
responsible for deep sores and ML. In L. (V.) brasilensis complex infections, cutaneous lesions may be followed by mucosal spread of the disease simultaneously or even years later. Young men with chronic lesions are particularly at risk, and between 2% and 40% of infected persons develop ‘espundia’, metastatic lesions in the mucosa of the nose or mouth. This is characterised by thickening and erythema of the nasal mucosa, typically starting at the junction of the nose and upper lip. Later, ulceration develops. The lips, soft palate, fauces and larynx may also be invaded and destroyed, leading to considerable suffering and deformity. There is no spontaneous healing, and death may result from severe respiratory tract infections due to massive destruction of the pharynx.

Investigations in CL and ML

CL is often diagnosed on the basis of the lesions’ clinical characteristics. However, parasitological confirmation is important because clinical manifestations may be mimicked by other infections. Amastigotes can be demonstrated on a slit-skin smear with Giemsa staining; alternatively, they can be cultured from the sores early during the infection. Parasites seem to be particularly difficult to isolate from sores caused by L. brasilensis, responsible for the vast majority of cases in Brazil. Touch preparations from biopsies and histopathology usually have a low sensitivity. Culture of fine needle aspiration material has been reported to be the most sensitive method.

ML is more difficult to diagnose parasitologically. The leishmanin skin test measures delayed-type hypersensitivity to killed Leishmania organisms. A positive test is defined as induration of more than 5 mm 48 hours after intradermal injection. The test is positive, except in diffuse CL and during active VL. PCR is used increasingly for diagnosis and speciation, which is useful in selecting therapy.

Management of CL and ML

Small lesions may self-heal or are treated by freezing with liquid nitrogen or cryettage. There is no ideal antimicrobial therapy. Treatment should be individualised on the basis of the causative organism, severity of the lesions, availability of drugs, tolerance of the patient for toxicity, and local resistance patterns.

In CL, topical application of paromomycin 15% plus methylbenzethionium chloride 12% is beneficial. Intralesional antimony (Sb: 0.2–0.8 mL/lesion) up to 2 g seems to be rapidly effective in suitable cases; it is well tolerated and economic, and is safe in patients with cardiac, liver or renal diseases.

In ML, and in CL when the lesions are multiple or in a disfiguring site, it is better to treat with parenteral Sb in a dose of 20 mg/kg/day (usually given for 20 days for CL and 28 days for ML), or with conventional or liposomal amphotericin B (see treatment of VL above). Sb is also indicated to prevent the development of mucosal disease, if there is any chance that a lesion acquired in South America is due to an L. brasilensis strain. Refractory CL or ML should be treated with an amphotericin B preparation.

Other regimens may be effective. Two to four doses of pentamidine (2–4 mg/kg), administered on alternate days, are effective in New World CL caused by L. guyanensis. In ML, 8 injections of pentamidine (4 mg/kg on alternate days) cure the majority of patients. Ketoconazole (600 mg daily for 4 weeks) has shown some potential against L. mexicana infection. In Saudi Arabia, fluconazole (200 mg daily for 6 weeks) reduced healing times and cured 79% of patients with CL caused by L. major. In India, itraconazole (200 mg daily for 6 weeks) produced good results in CL.

Prevention of CL and ML

Personal protection against sandflies is important. No effective vaccine is yet available.

Gastrointestinal protozoal infections

Amoebiasis

Amoebiasis is caused by Entamoeba histolytica, which is spread between humans by its cysts. It is one of the leading parasitic causes of morbidity and mortality in the tropics and is occasionally acquired in other countries, such as the UK. Two non-pathogenic Entamoeba species (E. dispar and E. moshkovskii) are morphologically identical to E. histolytica, and are distinguishable only by molecular techniques, isoenzyme studies or monoclonal antibody typing. However, only E. histolytica causes amoebic dysentery or liver abscess. The life cycle of the amoeba is shown in Figure 13.45A.

Pathology

Cysts of E. histolytica are ingested in water or uncooked foods contaminated by human faeces. Infection may also be acquired through anal/oral sexual practices. In the colon, trophozoite forms emerge from the cysts. The parasite may invade the mucous membrane of the large bowel, producing lesions that are maximal in the caecum but found as far down as the anal canal. These are flask-shaped ulcers, varying greatly in size and surrounded by healthy mucosa. A localised granuloma (amoeboma), presenting as a palpable mass in the rectum or a filling defect in the colon on radiography, is a rare complication which should be differentiated from colonic carcinoma. Amoebic ulcers may cause severe haemorrhage but rarely perforate the bowel wall.

Amoebic trophozoites can emerge from the vegetative cyst from the bowel and be carried to the liver in a portal venule. They can multiply rapidly and destroy the liver parenchyma, causing an abscess (see also p. 956). The liquid contents at first have a characteristic pinkish colour, which may later change to chocolate-brown (like anchovy sauce).

Cutaneous amoebiasis, though rare, causes progressive genital, perianal or peri-abdominal surgical wound ulceration.

Clinical features

Intestinal amoebiasis – amoebic dysentery

Most amoebic infections are asymptomatic. The incubation period of amoebiasis ranges from 2 weeks to many years, followed by a chronic course with abdominal pains and two or more unformed stools a day. Offensive diarrhoea alternating with constipation, and blood or mucus in the stool, are common. There may be abdominal pain, especially right lower quadrant (which may simulate acute appendicitis). A dysenteric presentation
with passage of blood, simulating bacillary dysentery or ulcerative colitis, occurs particularly in older people, in the puerperium and with superadded pyogenic infection of the ulcers.

**Amoebic liver abscess**

The abscess is usually found in the right hepatic lobe. There may not be associated diarrhoea. Early symptoms may be local discomfort only and malaise; later, a swinging temperature and sweating may develop, usually without marked systemic symptoms or associated cardiovascular signs. An enlarged, tender liver, cough and pain in the right shoulder are characteristic, but symptoms may remain vague and signs minimal. A large abscess may penetrate the diaphragm and rupture into the lung, from where its contents may be coughed up through a hepaticobronchial fistula. Rupture into the pleural or peritoneal cavity, or rupture of a left lobe abscess in the pericardial sac, is less common but more serious.

**Investigations**

The stool and any exudate should be examined at once under the microscope for motile trophozoites containing red blood cells. Movements cease rapidly as the stool preparation cools. Several stools may need to be examined in chronic amoebiasis before cysts are found. Sigmoidoscopy may reveal typical flask-shaped ulcers, which should be scraped and examined immediately for *E. histolytica*. In endemic areas, one-third of the population are symptomless passers of amoebic cysts.

An amoebic abscess of the liver is suspected on clinical grounds; there is often a neutrophil leucocytosis and a raised right hemidiaphragm on chest X-ray. Confirmation is by ultrasonic scanning. Aspirated pus from an amoebic abscess has the characteristic anchovy sauce or chocolate-brown appearance but only rarely contains free amoebae (Fig. 13.45B).

Serum antibodies are detectable by immunofluorescence in over 95% of patients with hepatic amoebiasis and intestinal amoeboma, but in only about 60% of dysenteric amoebiasis. DNA detection by PCR has been shown to be useful in diagnosis of *E. histolytica* infections but is not generally available.

**Management**

Intestinal and early hepatic amoebiasis responds quickly to oral metronidazole (800 mg 3 times daily for 5–10 days) or other long-acting nitroimidazoles like tinidazole or ornidazole (both in doses of 2 g daily for 3 days). Nitafoxanide (500 mg twice daily for 3 days) is an alternative drug. Either diloxanide furoate or paromomycin, in doses of 500 mg orally 3 times daily for 10 days after treatment, should be given to eliminate luminal cysts.

If a liver abscess is large or threatens to burst, or if the response to chemotherapy is not prompt, aspiration is required and is repeated if necessary. Rupture of an abscess into the pleural cavity, pericardial sac or peritoneal cavity necessitates immediate aspiration or surgical drainage. Small serous effusions resolve without drainage.

**Prevention**

Personal precautions against contracting amoebiasis consist of not eating fresh, uncooked vegetables or drinking unclean water.

**Giardiasis**

Infection with *Giardia lamblia* is found worldwide and is common in the tropics. It particularly affects children, tourists and immunosuppressed individuals, and is the parasite most commonly imported into the UK. In cystic form, it remains viable in water for up to 3 months and infection usually occurs by ingesting contaminated water. Its flagellar trophozoite form attaches to the duodenal and jejunal mucosa, causing inflammation.
Clinical features and investigations

After an incubation period of 1–3 weeks, there is diarrhoea, abdominal pain, weakness, anorexia, nausea and vomiting. On examination, there may be abdominal distension and tenderness.

Stools obtained at 2–3-day intervals should be examined for cysts. Duodenal or jejunal aspiration by endoscopy gives a higher diagnostic yield. The ‘string test’ may be used, in which one end of a piece of string is passed into the duodenum by swallowing and retrieved after an overnight fast; expressed fluid is then examined for the presence of G. lamblia trophozoites. A number of stool antigen detection tests are available. Jejunal biopsy specimens may show G. lamblia on the epithelial surface.

Management

Treatment is with a single dose of tinidazole 2 g, metronidazole 400 mg 3 times daily for 10 days, or nitazoxanide 500 mg orally twice daily for 3 days.

Cryptosporidiosis

Cryptosporidium spp. are coccidian protozoal parasites of humans and domestic animals. Infection is acquired by the faecal–oral route through contaminated water supplies. The incubation period is approximately 7–10 days and is followed by watery diarrhoea and abdominal cramps. The illness is usually self-limiting, but in immunocompromised patients, especially those with HIV, the illness can be devastating, with persistent severe diarrhoea and substantial weight loss (p. 399).

Cyclosporiasis

Cyclospora cayetanensis is a globally distributed coccidian protozoal parasite of humans. Infection is acquired by ingestion of contaminated water. The incubation period is approximately 2–11 days and is followed by acute onset of diarrhoea with abdominal cramps, which may remit and relapse. Although usually self-limiting, the illness may last as long as 6 weeks, with significant associated weight loss and malabsorption, and is more severe in immunocompromised individuals. Diagnosis is by detection of oocysts on faecal microscopy. Treatment may be necessary in a few cases, and the agent of choice is co-trimoxazole 960 mg twice daily for 7 days.

INFECTIONS CAUSED BY HELMINTHS

Helminths (from the Greek helmins, meaning worm) include three groups of parasitic worm (Box 13.65), large multicellular organisms with complex tissues and organs.

Intestinal human nematodes

Diseases are caused by adult nematodes living in the human gut. There are two types:

• the hookworms, which have a soil stage in which they develop into larvae that then penetrate the host

• a group of nematodes which survive in the soil merely as eggs that have to be ingested for their life cycle to continue.

The geographical distribution of hookworms is limited by the larval requirement for warmth and humidity. Soil-transmitted nematode infections can be prevented by avoidance of faecal soil contamination (adequate sewerage disposal) or skin contact (wearing shoes), and by strict personal hygiene.

Ancylostomiasis (hookworm)

Ancylostomiasis is caused by parasitisation with Ancylostoma duodenale or Necator americanus. The complex life cycle is shown in Figure 13.46. The adult hookworm is 1 cm long and lives in the duodenum and upper jejunum. Eggs are passed in the faeces. In warm, moist, shady soil, the larvae develop into rhabditiform and then the infective filariform stages; they then penetrate human skin and are carried to the lungs. After entering the alveoli, they ascend the bronchi, are swallowed and mature in the small intestine, reaching maturity 4–7 weeks after infection. The worms attach themselves to the mucosa of the small intestine by their buccal capsule (Fig. 13.47) and withdraw blood. The mean daily loss of blood from one A. duodenale is 0.15 mL and from N. americanus 0.03 mL.

Hookworm infection is one of the main causes of anaemia in the tropics and subtropics. A. duodenale is endemic in the Far East and Mediterranean coastal regions, and is also present in Africa, while N. americanus is endemic in West, East and Central Africa, and Central and South America, as well as in the Far East.

Clinical features

An allergic dermatitis, usually on the feet (ground itch), may be experienced at the time of infection. The passage of the larvae through the lungs in a heavy infection causes a paroxysmal cough with blood-stained sputum, associated with patchy pulmonary consolidation and eosinophilia. When the worms have reached the small intestine, vomiting and epigastric pain resembling peptic ulcer disease may occur. Sometimes, frequent loose stools are passed. The degree of iron and protein

<table>
<thead>
<tr>
<th>13.65 Classes of helminth that parasitise humans</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nematodes or roundworms</strong></td>
</tr>
<tr>
<td>• Intestinal human nematodes: Ancylostoma duodenale, Necator americanus, Strongyloides stercoralis, Ascaris lumbricoides, Enterobius vermicularis, Trichuris trichuria</td>
</tr>
<tr>
<td>• Tissue-dwelling human nematodes: Wuchereria bancrofti, Brugia malayi, Loa loa, Onchocerca volvulus, Dracunculus medinensis, Mansonella perstans, Dirofilaria immitis</td>
</tr>
<tr>
<td>• Zoonotic nematodes: Trichinella spiralis</td>
</tr>
<tr>
<td><strong>Trematodes or flukes</strong></td>
</tr>
<tr>
<td>• Blood flukes: Schistosoma haematobium, S. mansoni, S. japonicum, S. mekongi, S. intercalatum</td>
</tr>
<tr>
<td>• Lung flukes: Paragonimus spp.</td>
</tr>
<tr>
<td>• Hepatobiliary flukes: Clonorchis sinensis, Fasciola hepatica, Opisthorchis felineus</td>
</tr>
<tr>
<td>• Intestinal flukes: Fasciolopsis buski</td>
</tr>
<tr>
<td><strong>Cestodes or tapeworms</strong></td>
</tr>
<tr>
<td>• Intestinal tapeworms: Taenia saginata, T. solium, Diphyllobothrium latum, Hymenolepis nana</td>
</tr>
<tr>
<td>• Tissue-dwelling cysts or worms: Taenia solium, Echinococcus granulosus</td>
</tr>
</tbody>
</table>
Strongyloidiasis

Strongyloides stercoralis is a very small nematode (2 mm × 0.4 mm) which parasitises the mucosa of the upper part of the small intestine, often in large numbers, causing persistent eosinophilia. The eggs hatch in the bowel but only larvae are passed in the faeces. In moist soil, they moult and become the infective filariform larvae. After penetrating human skin, they undergo a development cycle similar to that of hookworms, except that the female worms burrow into the intestinal mucosa and submucosa. Some larvae in the intestine may develop into filariform larvae, which may then penetrate the mucosa or the perianal skin and lead to autoinfection and persistent infection. Patients with Strongyloides infection persisting for more than 35 years have been described. Strongyloidiasis occurs in the tropics and subtropics, and is especially prevalent in the Far East.

Clinical features

These are shown in Box 13.66. The classic triad of symptoms consists of abdominal pain, diarrhoea and urticaria. Cutaneous manifestations, either urticaria or larva currens (a highly characteristic pruritic, elevated, erythematous lesion advancing along the course of larval migration), are characteristic and occur in 66% of patients.

Systemic strongyloidiasis (the Strongyloides hyper-infection syndrome), with dissemination of larvae throughout the body, occurs in association with immune suppression (intercurrent disease, HIV and HTLV-1 infection, corticosteroid treatment). Patients present with severe, generalised abdominal pain, abdominal distension and shock. Massive larval invasion of the lungs causes cough, wheeze and dyspnoea; cerebral involvement has manifestations ranging from subtle neurological signs to coma. Gram-negative sepsis frequently complicates the picture.

Investigations

There is eosinophilia. Serology (ELISA) is helpful but definitive diagnosis depends upon finding the larvae. The faeces should be examined microscopically for motile larvae; excretion is intermittent and so repeated examinations may be necessary. Larvae can also be found in jejunal aspirate or detected using the string test (p. 369). Larvae may also be cultured from faeces.

<table>
<thead>
<tr>
<th>Box 13.66 Clinical features of strongyloidiasis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Penetration of skin by infective larvae</strong></td>
</tr>
<tr>
<td>• Itchy rash</td>
</tr>
<tr>
<td><strong>Presence of worms in gut</strong></td>
</tr>
<tr>
<td>• Abdominal pain, diarrhoea, steatorrhoea, weight loss</td>
</tr>
<tr>
<td><strong>Allergic phenomena</strong></td>
</tr>
<tr>
<td>• Urticarial plaques and papules, wheezing, arthralgia</td>
</tr>
<tr>
<td><strong>Autoinfection</strong></td>
</tr>
<tr>
<td>• Transient itchy, linear, urticarial weals across abdomen and buttocks (larva currens)</td>
</tr>
<tr>
<td><strong>Systemic (super)infection</strong></td>
</tr>
<tr>
<td>• Diarrhoea, pneumonia, meningoencephalitis, death</td>
</tr>
</tbody>
</table>

deficiency which develops depends not only on the load of worms but also on the nutrition of the patient and especially on the iron stores. Anaemia with high-output cardiac failure may result. The mental and physical development of children may be retarded in severe infection.

Investigations

There is eosinophilia. The characteristic ovum can be recognised in the stool. If hookworms are present in numbers sufficient to cause anaemia, faecal occult blood testing will be positive and many ova will be present.

Management

A single dose of albendazole (400 mg) is the treatment of choice. Alternatively, mebendazole 100 mg twice daily for 3 days may be used. Anaemia and heart failure associated with hookworm infection respond well to oral iron, even when severe; blood transfusion is rarely required.
Management
A course of two doses of ivermectin (200 µg/kg), administered on successive days, is effective. Alternatively, albendazole is given orally in a dose of 15 mg/kg body weight twice daily for 3 days. A second course may be required. For the Strongyloides hyperinfection syndrome, ivermectin is given at 200 µg/kg for 5–7 days.

Ascariis lumbricoides (roundworm)
This pale yellow nematode is 20–35 cm long. Humans are infected by eating food contaminated with mature ova. Ascariis larvae hatch in the duodenum, migrate through the lungs, ascend the bronchial tree, are swallowed and mature in the small intestine. This tissue migration can provoke both local and general hypersensitivity reactions, with pneumonitis, eosinophilic granulomas, bronchial asthma and urticaria.

Clinical features
Intestinal ascariasis causes symptoms ranging from occasional vague abdominal pain through to malnutrition. The large size of the adult worm and its tendency to aggregate and migrate can result in obstructive complications. Tropical and subtropical areas are endemic for ascariasis, and in these areas it causes up to 35% of all intestinal obstructions, most commonly in the terminal ileum. Obstruction can be complicated further by intussusception, volvulus, haemorrhagic infarction and perforation. Other complications include blockage of the bile or pancreatic duct and obstruction of the appendix by adult worms.

Investigations
The diagnosis is made microscopically by finding ova in the faeces. Adult worms are frequently expelled rectally or orally. Occasionally, the worms are demonstrated radiographically by a barium examination. There is eosinophilia.

Management
A single dose of albendazole (400 mg), pyrantel pamoate (11 mg/kg; maximum 1 g), ivermectin (150–200 µg/kg) or mebendazole (100 mg twice daily for 3 days) is effective for intestinal ascariasis. Patients should be warned that they might expel numerous whole, large worms. Obstruction due to ascariasis should be treated with nasogastric suction, piperazine and intravenous fluids.

Prevention
Community chemotherapy programmes have been used to reduce Ascariis infection. The whole community can be treated every 3 months for several years. Alternatively, schoolchildren can be targeted; treating them lowers the prevalence of ascariasis in the community.

Enterobius vermicularis (threadworm)
This helminth is common throughout the world and affects mainly children. After the ova are swallowed, development takes place in the small intestine, but the adult worms are found chiefly in the colon.

Clinical features
The gravid female worm lays ova around the anus, causing intense itching, especially at night. The ova are often carried to the mouth on the fingers and so re-infection or human-to-human infection takes place (Fig. 13.48). In females, the genitalia may be involved. The adult worms may be seen moving on the buttocks or in the stool.

Investigations
Ova are detected by applying the adhesive surface of cellophane tape to the perianal skin in the morning. This is then examined on a glass slide under the microscope. A perianal swab, moistened with saline, is an alternative sampling method.

Management
A single dose of mebendazole (100 mg), albendazole (400 mg), pyrantel pamoate (11 mg/kg) or piperazine (4 g) is given and may be repeated after 2 weeks to control auto-reinfection. If infection recurs in a family, each member should be treated as above. During this period all nightclothes and bed linen are laundered. Fingernails must be kept short and hands washed carefully before meals. Subsequent therapy is reserved for those family members who develop recurrent infection.

Trichuris trichiura (whipworm)
Infections with whipworm are common all over the world under unhygienic conditions. Infection is contracted by the ingestion of earth or food contaminated with ova which have become infective after lying for 3 weeks or more in moist soil. The adult worm is 3–5 cm long and has a coiled anterior end resembling a whip. Whipworms inhabit the caecum, lower ileum, appendix, colon and anal canal. There are usually no symptoms, but intense infections in children may cause persistent diarrhoea or rectal prolapse, and growth retardation. The diagnosis is readily made by identifying ova in faeces. Treatment is with mebendazole in doses of
Tissue-dwelling human nematodes

Filarial worms are tissue-dwelling nematodes. The larval stages are inoculated by biting mosquitoes or flies, each specific to a particular filarial species. The larvae develop into adult worms (2–50 cm long) which, after mating, produce millions of microfilariae (170–320 μm long) that migrate in blood or skin. The life cycle is completed when the vector takes up microfilariae while feeding on humans. In the insect, ingested microfilariae develop into infective larvae for inoculation in humans, normally the only host.

Disease is due to the host’s immune response to the worms (both adult and microfilariae), particularly dying worms, and its pattern and severity vary with the site and stage of each species (Box 13.67). The worms are long-lived; microfilariae survive 2–3 years and adult worms 10–15 years. The infections are chronic and worst in individuals constantly exposed to re-infection.

Lymphatic filariasis

Infection with the filarial worms Wuchereria bancrofti and Brugia malayi is associated with clinical outcomes ranging from subclinical infection to hydrocele and elephantiasis.

W. bancrofti is transmitted by night-biting culicine or anopheline mosquitoes in most areas (Fig. 13.49). The adult worms, 4–10 cm in length, live in the lymphatics, and the females produce microfilariae which circulate in large numbers in the peripheral blood, usually at night. The infection is widespread in tropical Africa, on the North African coast, in coastal areas of Asia, Indonesia and northern Australia, the South Pacific islands, the West Indies and also in North and South America.

B. malayi usually causes less severe disease than W. bancrofti and is transmitted by Mansonia or Anopheles mosquitoes in Indonesia, Borneo, Malaysia, Vietnam, South China, South India and Sri Lanka.

Pathology

Several factors contribute to the pathogenesis of lymphatic filariasis. Toxins released by the adult worm cause lymphangiectasia; this dilatation of the lymphatic vessels leads to lymphatic dysfunction and the chronic clinical manifestations of lymphatic filariasis, lymphoedema and hydrocele. Death of the adult worm results in acute filarial lymphangitis. The filaricaria are symbiotically infected with rickettsia-like bacteria (Wolbachia spp.), and release of lipopolysaccharide from these bacteria contributes to inflammation. Lymphatic obstruction persists after death of the adult worm. Secondary bacterial infections cause tissue destruction. The host response to microfilariae is central to the pathogenesis of tropical pulmonary eosinophilia.

Clinical features

Acute filarial lymphangitis presents with fever, pain, tenderness and erythema along the course of inflamed lymphatic vessels. Inflammation of the spermatic cord, epididymis and testis is common. The whole episode lasts a few days but may recur several times a year. Temporary oedema becomes more persistent and regional lymph nodes enlarge. Progressive enlargement, coarsening, corrugation, fissuring and bacterial infection of the skin and subcutaneous tissue develop gradually, causing irreversible ‘elephantiasis’. The scrotum may reach an enormous size. Chyluria and chylous effusions are milky and opalescent; on standing, fat globules rise to the top.

The acute lymphatic manifestations of filariasis must be differentiated from thrombophlebitis and infection. The oedema and lymphatic obstructive changes must be distinguished from congestive cardiac failure, malignancy, trauma and idiopathic abnormalities of the lymphatic system. Silicates absorbed from volcanic soil can also cause non-filarial elephantiasis.

Tropical pulmonary eosinophilia is a complication seen mainly in India and is likely to be due to microfilariae trapped in the pulmonary capillaries and destroyed.
by allergic inflammation. Patients present with paroxysmal cough, wheeze and fever. If untreated, this may progress to debilitating chronic interstitial lung disease.

**Investigations**

In the earliest stages of lymphangitis, the diagnosis is made on clinical grounds, supported by eosinophilia and sometimes by positive filarial serology. Filarial infections cause the highest eosinophil counts of all helminthic infections.

Microfilariae can be found in the peripheral blood at night, and either are seen moving in a wet blood film or are detected by microfiltration of a sample of lysed blood. They are usually present in hydrocele fluid, which may occasionally yield an adult filaria. By the time elephantiasis develops, microfilariae become difficult to find. Calculi filariae may sometimes be demonstrable by radiography. Movement of adult worms can be seen on scrotal ultrasound. PCR-based tests for detection of *W. bancrofti* and *B. malayi* DNA from blood have been developed.

Indirect fluorescence and ELISA detect antibodies in over 95% of active cases and 70% of established elephantiasis. The test becomes negative 1–2 years after cure. Serological tests cannot distinguish the different filarial infections. Highly sensitive and specific immunochromatographic card tests for detection of circulating *W. bancrofti* antigen are now commercially available; fingerprick blood taken at any time of the day can be used for these.

In tropical pulmonary eosinophilia, serology is strongly positive and IgE levels are massively elevated, but circulating microfilariae are not found. The chest X-ray shows miliary changes or motled opacities. Pulmonary function tests show a restrictive picture.

**Management**

Treatment of the individual is aimed at reversing and halting disease progression. Diethylcarbamazine (DEC, 2 mg/kg orally 3 times daily for 12 days, or as a single dose) kills microfilariae and adult worms. Most adverse effects seen with DEC treatment are due to the host response to dying microfilariae, and the reaction intensity is directly proportional to the microfilarial load. The main symptoms are fever, headache, nausea, vomiting, arthralgia and prostration. These usually occur within 24–36 hours of the first dose of DEC. Antibiotics or corticosteroids may be required to control these allergic phenomena. Combining albendazole (400 mg) with ivermectin (200 µg/kg) in a single dose, with or without DEC (300 mg), is also highly effective in clearing the parasites. Treatment of *Wolbachia* with doxycycline (200 mg/day) for 4–8 weeks provides additional benefit by eliminating the bacteria; this leads to interruption of parasite embryogenesis. For tropical pulmonary eosinophilia, DEC for 14 days is the treatment of choice.

**Chronic lymphatic pathology**

Experience in India and Brazil shows that active management of chronic lymphatic pathology can alleviate symptoms. Patients should be taught meticulous skin care of their lymphoedematous limbs to prevent secondary bacterial and fungal infections. Tight bandaging, massage and bed rest with elevation of the affected limb may help to control the lymphoedema. Prompt diagnosis and antibiotic therapy of bacterial cellulitis are important in preventing further lymphatic damage and worsening of existing elephantiasis. Plastic surgery may be indicated in established elephantiasis. Great relief can be obtained by removal of excess tissue but recurrences are probable unless new lymphatic drainage is established. Hydroceles and chyluria can be repaired surgically.

**Prevention**

Treatment of the whole population in endemic areas with annual single-dose DEC (6 mg/kg), either alone or in combination with albendazole or ivermectin, can reduce filarial transmission. This mass treatment should be combined with mosquito control programmes.

**Loiasis**

Loiasis is caused by infection with the filaria *Loa loa*. The disease is endemic in forested and swampy parts of Western and Central Africa. The adult worms, 3–7 cm × 4 mm, chiefly parasitise the subcutaneous tissue of humans, releasing larval microfilariae into the peripheral blood in the daytime. The vector is *Chrysops*, a forest-dwelling, day-biting fly.

The host response to *Loa loa* is usually absent or mild, so that the infection may be harmless. From time to time a short-lived, inflammatory, oedematous swelling (a Calabar swelling) is produced around an adult worm. Heavy infections, especially when treated, may cause encephalitis.

**Clinical features**

The infection is often symptomless. The incubation period is commonly over a year but may be just 3 months. The first sign is usually a Calabar swelling, an irritating, tense, localised swelling that may be painful, especially if it is near a joint. The swelling is generally on a limb; it measures a few centimetres in diameter but sometimes is more diffuse and extensive. It usually disappears after a few days but may persist for 2 or 3 weeks. A succession of such swellings may appear at irregular intervals, often in adjacent sites. Sometimes, there is urticaria and pruritus elsewhere. Occasionally, a worm may be seen wriggling under the skin, especially that of an eyelid, and may cross the eye under the conjunctiva, taking many minutes to do so.

**Investigations**

Diagnosis is by demonstrating microfilariae in blood taken during the day, but they may not always be found in patients with Calabar swellings. Antifilarial antibodies are positive in 95% of patients and there is massive eosinophilia. Occasionally, a calcified worm may be seen on X-ray.

**Management**

DEC (see above) is curative, in a dose of 9–12 mg/kg daily, continued for 21 days. Treatment may precipitate a severe reaction in patients with a heavy microfilaraemia characterised by fever, joint and muscle pain, and encephalitis; microfilaraemic patients should be given corticosteroid cover.

**Prevention**

Protection is afforded by building houses away from trees and by having dwellings wire-screened. Protective
categorising and insect repellents are also useful. DEC in a dose of 5 mg/kg daily for 3 days each month is partially protective.

**Onchocerciasis (river blindness)**

Onchocerciasis is the result of infection by the filarial *Onchocerca volvulus*. The infection is conveyed by flies of the genus *Simulium*, which breed in rapidly flowing, well-aerated water. Adult flies inflict painful bites during the day, both inside and outside houses. While feeding, they pick up the microfilariae, which mature into the infective larva and are transmitted to a new host in subsequent bites. Humans are the only known hosts (Fig. 13.50).

Onchocerciasis is endemic in sub-Saharan Africa, Yemen and a few foci in Central and South America. It is estimated that 17.7 million people are infected, of whom 500,000 are visually impaired and 270,000 blind. Due to onchocerciasis, huge tracts of fertile land lie virtually untilled, and individuals and communities are impoverished.

**Pathology**

After inoculation of larvae by a bite from an infected fly, the worms mature in 2–4 months and live for up to 17 years in subcutaneous and connective tissues. At sites of trauma, over bony prominences and around joints, fibrosis may form nodules around adult worms, which otherwise cause no direct damage. Innumerable microfilariae, discharged by the female *O. volvulus*, move actively in these nodules and in the adjacent tissues, are widely distributed in the skin, and may invade the eye. Live microfilariae elicit little tissue reaction, but dead ones may cause severe allergic inflammation, leading to hyaline necrosis and loss of collagen and elastin. Death of microfilariae in the eye causes inflammation and may lead to blindness.

**Clinical features**

The infection may remain symptomless for months or years. The first symptom is usually itching, localised to one quadrant of the body and later becoming generalised and involving the eyes. Transient oedema of part or all of a limb is an early sign, followed by papular urticaria spreading gradually from the site of infection. This is difficult to see on dark skins, in which the most common signs are papules excoriated by scratching, spotty hyperpigmentation from resolving inflammation, and more chronic changes of a rough, thickened or inelastic, wrinkled skin. Both infected and uninfected superficial lymph nodes enlarge and may hang down in folds of loose skin in the groin. Hydrocele, femoral hernias and scrotal elephantiasis can occur. Firm subcutaneous nodules of more than 1 cm in diameter (onchocercomas) occur in chronic infection.

Eye disease is most common in highly endemic areas and is associated with chronic heavy infections and nodules on the head. Early manifestations include itching, lacrimation and conjunctival injection. These cause conjunctivitis, sclerosing keratitis with pannus formation, uveitis which may lead to glaucoma and cataract, and, less commonly, choroiditis and optic neuritis. Classically, ‘snowflake’ deposits are seen in the edges of the cornea.

**Investigations**

The finding of nodules or characteristic lesions of the skin or eyes in a patient from an endemic area, associated with eosinophilia, is suggestive. Skin snips or shavings, taken with a corneoscleral punch or scalpel blade from calf, buttock and shoulder, are placed in saline under a cover slip on a microscope slide and examined after 4 hours. Microfilariae are seen wriggling free in all but the lightest infections. Slit-lamp examination may reveal microfilariae moving in the anterior chamber of the eye or trapped in the cornea. A nodule may be removed and incised, showing the coiled, thread-like adult worm.

Filarial antibodies may be detected in up to 95% of patients. Several promising rapid strip tests based on antibody or antigen detection are under clinical evaluation. When there is a strong suspicion of onchocerciasis but tests are negative, a provocative Mazzotti test, in which administration of 0.5–1.0 mg/kg of DEC exacerbates pruritus or dermatitis, strongly suggests onchocerciasis.

**Management**

Ivermectin, in a single dose of 100–200 µg/kg, repeated several times at 3-monthly intervals to prevent relapses, is recommended. It kills microfilariae and has minimal toxicity. In the rare event of a severe reaction causing oedema or postural hypotension, prednisolone 20–30 mg may be given daily for 2 or 3 days. Ivermectin has little macrofilaricidal effect so that, 1 year after ivermectin treatment, skin microfilarial densities regain at least 20% of pre-treatment levels; repeated treatments are required for the lifespan of the adult worm. Eradication of *Wolbachia* with doxycycline (100 mg daily for 6 weeks) prevents worm reproduction.

**Prevention**

Mass treatment with ivermectin is practised. It reduces morbidity in the community and slows the progression
of eye disease, but it does not clear worm infection. *Simulium* can be destroyed in its larval stage by the application of insecticide to streams. Long trousers, skirts and sleeves discourage the fly from biting.

**Dracunculiasis (Guinea worm)**

Infestation with the Guinea worm *Dracunculus medinensis* manifests itself when the female worm, over a metre long, emerges from the skin. Humans are infected by ingesting a small crustacean, *Cyclops*, which inhabits wells and ponds, and contains the infective larval stage of the worm. The worm was widely distributed across Africa and the Middle East, but after a successful eradication programme is now seen only in sub-Saharan Africa.

**Management and prevention**

Traditionally, the protruding worm is extracted by winding it out gently over several days on a matchstick. The worm must never be broken. Antibiotics for secondary infection and prophylaxis of tetanus are also required.

A global elimination campaign is based on the provision of clean drinking water and eradication of water fleas from drinking water. The latter is being achieved by simple filtration of water through a plastic mesh filter and chemical treatment of water supplies.

**Other filariases**

*Mansonella perstans*

This filarial worm is transmitted by the midges *Culicoides austeni* and *C. grahami*. It is common throughout equatorial Africa as far south as Zambia, and also in Trinidad and parts of northern and eastern South America.

*M. perstans* has never been proven to cause disease but it may be responsible for a persistent eosinophilia and occasional allergic manifestations. *M. perstans* is resistant to ivermectin and DEC, and the infection may persist for many years.

*Mansonella immittis*

This dog heart worm infects humans, causing skin and lung lesions. It is not uncommon in the USA, Japan and Australia.

**Zoonotic nematodes**

**Trichinosis (trichinellosis)**

*Trichinella spiralis* is a nematode that parasitises rats and pigs, and is only transmitted to humans if they eat partially cooked infected pork, usually as sausage or ham. Bear meat is another source. Symptoms result from invasion of intestinal submucosa by ingested larvae, which develop into adult worms, and the secondary invasion of striated muscle by fresh larvae produced by these adult worms. Outbreaks have occurred in the UK, as well as in other countries where pork is eaten.

**Clinical features**

The clinical features of trichinosis are determined by the larval numbers. A light infection with a few worms may be asymptomatic; a heavy infection causes nausea and diarrhoea 24–48 hours after the infected meal. A few days later, the symptoms associated with larval invasion predominate: there is fever and oedema of the face, eyelids and conjunctivae; invasion of the diaphragm may cause pain, cough and dyspnoea; and involvement of the muscles of the limbs, chest and mouth causes stiffness, pain and tenderness in affected muscles. Larval migration may cause acute myocarditis and encephalitis. An eosinophilia is usually found after the second week. An intense infection may prove fatal but those who survive recover completely.

**Investigations**

Commonly, a group of people who have eaten infected pork from a common source develop symptoms at about the same time. Biopsy from the deltoid or gastrocnemius muscle after the third week of symptoms may reveal encysted larvae. Serological tests are also helpful.

**Management**

Treatment is with albendazole (400 mg twice daily for 8–14 days) or mebendazole (200–400 mg three times daily for 3 days, followed by 400–500 mg three times daily for 10 days). Given early in the infection, this may kill newly formed adult worms in the submucosa and thus reduce the number of larvae reaching the muscles. Corticosteroids are necessary to control the serious effects of acute inflammation.

**Cutaneous larva migrans**

Cutaneous larva migrans (CLM) is the most common linear lesion seen in travellers (Fig. 13.51). Intensely pruritic, linear, serpiginous lesions result from the larval migration of the dog hookworm (*Ancylostoma caninum*). The track moves across the skin at a rate of 2–3 cm/day. This contrasts with the rash of *Strongyloides* (p. 370), which is fast-moving and transient. Although the larvae of dog hookworms frequently infect humans, they do not usually develop into the adult form. The most common site for CLM is the foot but elbows, breasts and buttocks may be affected. Most patients with CLM have recently visited a beach where the affected part was exposed. The diagnosis is clinical. Treatment may be

---

**Fig. 13.51** Cutaneous larva migrans.
local with 12-hourly application of 15% thiabendazole cream, or systemic with a single dose of albendazole (400 mg) or ivermectin (150-200 µg/kg).

**Trematodes (flukes)**

These leaf-shaped worms are parasitic to humans and animals. Their complex life cycles may involve one or more intermediate hosts, often freshwater molluscs.

**Schistosomiasis**

Schistosomiasis is one of the most important causes of morbidity in the tropics. There are five species of the genus *Schistosoma* which commonly cause disease in humans: *S. haematobium*, *S. mansoni*, *S. japonicum*, *S. mekongi* and *S. intercalatum*. *S. haematobium* was discovered by Theodor Bilharz in Cairo in 1861 and the disease is sometimes called bilharziasis or bilharziasis. Schistosome eggs have been found in Egyptian mummies dated 1250 BC.

The life cycle is shown in Figure 13.52A. The ovum is passed in the urine or faeces of infected individuals and gains access to fresh water, where the ciliated miracidium inside it is liberated; it enters its intermediate host, a species of freshwater snail, in which it multiplies. Large numbers of fork-tailed cercariae are then liberated into the water, where they may survive for 2-3 days. Cercariae can penetrate the skin or the mucous membrane of the mouth of humans. They transform into schistosomulae and moult as they pass through the lungs; thence they are carried by the blood stream to the liver, and so to the portal vein, where they mature. The male worm is up to 20 mm in length and the more slender cylindrical female, usually enfolded longitudinally by the male, is rather longer (Fig. 13.52B). Within 4-6 weeks of infection, they migrate to the venules draining the pelvic viscera, where the females deposit ova.

**Pathology**

This depends on the species and the stage of infection (Box 13.68). Most disease is due to the passage of eggs through mucosa and to the granulomatous reaction to eggs deposited in tissues. The eggs of *S. haematobium* pass mainly through the wall of the bladder, but may also involve rectum, seminal vesicles, vagina, cervix and uterine tubes. *S. mansoni* and *S. japonicum* eggs pass mainly through the wall of the lower bowel or are carried to the liver. The most serious, although rare, site of ectopic deposition of eggs is in the CNS. Granulomas are composed of macrophages, eosinophils, and epithelioid and giant cells around an ovum. Later, there is fibrosis and eggs calcify, often in sufficient numbers to become radiologically visible. Eggs of *S. haematobium* may leave the vesical plexus and be carried directly to the lung. Those of *S. mansoni* and *S. japonicum* may also reach the lungs after the development of portal hypertension and consequent portosystemic collateral circulation. In both circumstances, egg deposition in the pulmonary vasculature, and the resultant host response, can lead to the development of pulmonary hypertension.

**Clinical features**

Recent travellers, especially those overlanding through Africa, may present with allergic manifestations and eosinophilia; residents of schistosomiasis-endemic areas are more likely to present with chronic urinary tract pathology or portal hypertension.

During the early stages of infection, there may be itching lasting 1-2 days at the site of cercarial penetration

![Fig. 13.52 Schistosoma](image-url) **A** Life cycle. **B** Scanning electron micrograph of adult schistosome worms, showing the larger male worm embracing the thinner female.
### Pathogenesis of schistosomiasis

<table>
<thead>
<tr>
<th>Time</th>
<th>(S. \text{haematobium})</th>
<th>(S. \text{mansoni}) and (S. \text{japonicum})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cercarial penetration</td>
<td>Days</td>
<td>----</td>
</tr>
<tr>
<td>Pneumonitis, myositis, hepatitis, fever, ‘serum sickness’, eosinophilia, seroconversion</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Early egg deposition</td>
<td>Months</td>
<td>Cystitis, haematuria</td>
</tr>
<tr>
<td>Ectopic granulomatous lesions: skin, CNS etc.</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Immune complex glomerulonephritis</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Late egg deposition</td>
<td>Years</td>
<td>Colonic polyposis and strictures, perportal fibrosis, portal hypertension</td>
</tr>
<tr>
<td>Fibrosis and calcification of ureters, bladder: bacterial infection, calculi, hydrenephrosis, carcinoma</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Pulmonary granulomas and pulmonary hypertension</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

The severity of \(S. \text{haematobium}\) infection varies greatly, and many with a light infection are asymptomatic. However, as adult worms can live for 20 years or more and lesions may progress, these patients should always be treated.

**Schistosoma mansoni**

\(S. \text{mansoni}\) is endemic throughout Africa, the Middle East, Venezuela, Brazil and the Caribbean (see Fig. 13.53).

Characteristic symptoms begin 2 months or more after infection. They may be slight, no more than malaise, or consist of abdominal pain and frequent stools which contain blood-stained mucus. With severe advanced disease, increased discomfort from rectal polyps may be experienced. The early hepatomegaly is reversible, but portal hypertension may cause massive splenomegaly, fatal haematemesis from oesophageal varices, or progressive ascites (p. 938). Liver function is initially preserved because the pathology is fibrotic rather than cirrhotic. \(S. \text{mansoni}\) and other schistosome infections predispose to the carriage of \(Salmonella\), in part because \(Salmonella\) may attach to the schistosomes and in part because shared antigens on schistosomes may induce immunological tolerance to \(Salmonella\).

**Schistosoma japonicum**, \(S. \text{mekongi}\) and \(S. \text{intercalatum}\)

In addition to humans, the adult worm of \(S. \text{japonicum}\) infects the dog, rat, field mouse, water buffalo, ox, cat, pig, horse and sheep. Although other \(Schistosoma\) spp. can infect species other than humans, the non-human reservoir seems to be particularly important in transmission for \(S. \text{japonicum}\) but not for \(S. \text{haematobium}\) or \(S. \text{mansoni}\). \(S. \text{japonicum}\) is prevalent in the Yellow River and Yangtze-Jiang basins in China, where the infection is a major public health problem. It also has a focal distribution in the Philippines, Indonesia and Thailand (see Fig. 13.53). The related \(S. \text{mekongi}\) occurs in Laos, Thailand and Myanmar, and \(S. \text{intercalatum}\) in West and Central Africa.

The pathology of \(S. \text{japonicum}\) is similar to that of \(S. \text{mansoni}\), but as this worm produces more eggs, the
lesions tend to be more extensive and widespread. The clinical features resemble those of severe infection with *S. mansoni*, with added neurological features. The small and large bowel may be affected, and hepatic fibrosis with splenic enlargement is usual. Deposition of eggs or worms in the CNS, especially in the brain or spinal cord, causes symptoms in about 5% of infections, notably epilepsy, blindness, hemiplegia or paraplegia.

**Investigations**

There is marked eosinophilia. Serological tests (ELISA) are useful as screening tests but remain positive after chemotherapeutic cure.

In *S. haematobium* infection, dipstick urine testing shows blood and albumin. The eggs can be found by microscopic examination of the centrifuged deposit of terminal stream urine (Fig. 13.54). Ultrasound is useful for assessing the urinary tract; bladder wall thickening, hydrenephrosis and bladder calcification can be detected. Cystoscopy reveals ‘sandy’ patches, bleeding mucosa and later distortion.

In a heavy infection with *S. mansoni* or *S. japonicum*, the characteristic egg with its lateral spine can usually be found in the stool. When the infection is light or of long duration, a rectal biopsy can be examined. Sigmodioscopy may show inflammation or bleeding. Biopsies should be examined for ova.

**Management**

The object of specific treatment is to kill the adult schistosomes and so stop egg-laying. Praziquantel (20 mg/kg orally twice daily for 1 day) is the drug of choice for all forms of schistosomiasis. The drug produces parasitological cure in 80% of treated individuals and over 90% reduction in egg counts in the remainder. Side-effects are uncommon but include nausea and abdominal pain. Praziquantel therapy in early infection reverses pathologies such as hepatomegaly and bladder wall thickening and granulomas.

Surgery may be required to deal with residual lesions such as ureteric stricture, small fibrotic urinary bladders, or granulomatous masses in the brain or spinal cord. Removal of rectal papillomas by diathermy or by other means may provide symptomatic relief.

**Prevention**

So far, no satisfactory single means of controlling schistosomiasis has been established. The life cycle is terminated if the ova in urine or faeces are not allowed to contaminate fresh water containing the snail host. The provision of latrines and of a safe water supply, however, remains a major problem in rural areas throughout the tropics. Furthermore, *S. japonicum* has so many hosts besides humans that latrines would be of little avail.

Annual mass treatment of the population helps prevent *S. haematobium* and *S. mansoni* infection, but this method has so far had little success with *S. japonicum*. Targeting the intermediate host, the snail, presents many difficulties and has not, on its own, proved successful on any scale. For personal protection, contact with infected water must be avoided.

**Liver flukes**

Liver flukes infect at least 20 million people and remain an important public health problem in many endemic areas. They are associated with abdominal pain, hepatomegaly and relapsing cholangitis. *Clonorchis sinensis* is a major aetiological agent of bile duct cancer. The three major liver flukes have similar life cycles and pathologies, as outlined in Box 13.69.

Other flukes of medical importance include lung and intestinal flukes (see Box 13.65, p. 369).

**Cestodes (tapeworms)**

Cestodes are ribbon-shaped worms which inhabit the intestinal tract. They have no alimentary system and absorb nutrients through the tegumental surface. The anterior end, or scolex, has suckers for attaching to the host. From the scolex, a series of progressively developing segments arise, the proglottides, which may continue to show active movements when shed. Cross-fertilisation takes place between segments. Ova, present in large numbers in mature proglottides, remain viable for weeks, and during this period, they may be consumed by the intermediate host. Larvae liberated from the ingested ova pass into the tissues, forming larval cysticerci.

Tapeworms cause two distinct patterns of disease, either intestinal infection or systemic cysticercosis (Fig. 13.55). *Taenia saginata* (beef tapeworm), *Taenia asiatica* and *Diphyllobothrium latum* (fish tapeworm) cause only intestinal infection, following human ingestion of intermediate hosts that contain cysticerci (the larval stage of the tapeworm). *Taenia solium* causes intestinal infection if a cysticerci-containing intermediate host is ingested, and cysticercosis (systemic infection from larval migration) if ova are ingested. *Echinococcus granulosus* (dog tapeworm) does not cause human intestinal infection, but causes hydatid disease (which is analogous to cysticercosis) following ingestion of ova and subsequent larval migration.

**Intestinal tapeworm**

Humans acquire tapeworm by eating undercooked beef infected with the larval stage of *T. saginata*, undercooked pork containing the larval stage of *T. solium* or *T. asiatica*, or undercooked freshwater fish containing larvae of *D. latum*. Usually, only one adult tapeworm is present in the gut but up to ten have been reported. The ova of all

---

**Fig. 13.54** Ova of *Schistosoma haematobium* in urine. Note the terminal spike.
13.69 Diseases caused by flukes in the bile duct

<table>
<thead>
<tr>
<th>Parasite</th>
<th>Clonorchis sinensis</th>
<th>Opisthorchis felineus</th>
<th>Fasciola hepatica</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other mammalian hosts</td>
<td>Dogs, cats, pigs</td>
<td>Dogs, cats, foxes, pigs</td>
<td>Sheep, cattle</td>
</tr>
<tr>
<td>Mode of spread</td>
<td>Ova in faeces, water</td>
<td>As for C. sinensis</td>
<td>Ova in faeces on to wet pasture</td>
</tr>
<tr>
<td>1st intermediate host</td>
<td>Snails</td>
<td>Snails</td>
<td>Snails</td>
</tr>
<tr>
<td>2nd intermediate host</td>
<td>Freshwater fish</td>
<td>Freshwater fish</td>
<td>Encysts on vegetation</td>
</tr>
<tr>
<td>Geographical distribution</td>
<td>Far East, especially S. China</td>
<td>Far East, especially NE Thailand</td>
<td>Cosmopolitan, including UK</td>
</tr>
<tr>
<td>Pathology</td>
<td>E. coli cholangitis, abscesses, biliary carcinoma</td>
<td>As for C. sinensis</td>
<td>Toxaemia, cholangitis, eosinophilia</td>
</tr>
<tr>
<td>Symptoms</td>
<td>Often symptom-free, recurrent jaundice</td>
<td>As for C. sinensis</td>
<td>Unexplained fever, tender liver, may be ectopic, e.g. subcutaneous fluke</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Ova in in stool or duodenal aspirate</td>
<td>As for C. sinensis</td>
<td>As for C. sinensis, also serology</td>
</tr>
<tr>
<td>Prevention</td>
<td>Cook fish</td>
<td>Cook fish</td>
<td>Avoid contaminated watercress</td>
</tr>
<tr>
<td>Treatment</td>
<td>Praziquantel 25 mg/kg 3 times daily for 2 days</td>
<td>As for C. sinensis but for 1 day only</td>
<td>Triclabendazole 10 mg/kg single dose; repeat treatment may be required*</td>
</tr>
</tbody>
</table>

*In the UK, available from the Hospital for Tropical Diseases, London.

\[ Fig. 13.55 \text{ Cysticercosis. Life cycle of Taenia solium.} \]

If eggs are swallowed by humans they develop to cysticerci in various sites, e.g. brain, muscle

If cysticerci are swallowed they develop to adult tapeworms in the human intestine

Eggs ingested by pig become cysticerci in muscles

Human pork tapeworm infection results from eating undercooked pork containing cysticerci

Cysticerci

Adult worms in gut

Eggs passed in human faeces

Human cysticercosis results from ingestion of the tapeworm eggs as a result of faecal contamination of food

Eggs in the faeces

Ingestion of meat

Pig

Faecal-oral route

If eggs are swallowed by humans they develop to cysticerci in various sites, e.g. brain, muscle

The three Taenia are indistinguishable microscopically. However, examination of scolex and proglottides can differentiate: T. solium has a rostellum and two rows of hooklets on the scolex, and discharges multiple proglottides (3–5) attached together with lower degrees of uterine branching (approximately 10); T. saginata has only four suckers in its scolex, and discharges single proglottids with greater uterine branching (up to 30); T. asiatica has a rostellum without hooks on its scolex, and is difficult to differentiate from T. saginata, except that there are fewer uterine branches (16–21).

**Taenia saginata**

Infection with T. saginata occurs in all parts of the world. The adult worm may be several metres long and produces little or no intestinal upset in human beings, but knowledge of its presence, by noting segments in the faeces or on underclothing, may distress the patient. Ova may be found in the stool. Praziquantel is the drug of choice; niclosamide or nitazoxanide is an alternative. Prevention depends on efficient meat inspection and the thorough cooking of beef.

**Taenia solium**

*T. solium*, the pork tapeworm, is common in central Europe, South Africa, South America and parts of Asia. It is not as large as *T. saginata*. The adult worm is found only in humans following the eating of undercooked pork containing cysticerci. Intestinal infection is treated with praziquantel (5–10 mg/kg) or niclosamide (2 g), both as a single dose, or alternatively with nitazoxanide (500 mg twice daily for 3 days). These are followed by a mild laxative (after 1–2 hours) to prevent retrograde intestinal autoinfection. Cooking pork well prevents intestinal infection. Great care must be taken while attending a patient harbouring an adult worm to avoid ingestion of ova or segments.

**Taenia asiatica**

*T. asiatica* is a newly recognised species of *Taenia*, restricted to Asia. It is acquired by eating uncooked meat or viscera of pigs. Clinical features and treatment are similar to those of *T. saginata*.
**INFECTIONOUS DISEASE**

**Cysticercosis**

Human cysticercosis is acquired by ingesting *T. solium* tapeworm ova, from either contaminated fingers or food (see Fig. 13.55). The larvae are liberated from eggs in the stomach, penetrate the intestinal mucosa and are carried to many parts of the body, where they develop and form cysticerci, 0.5–1 cm cysts that contain the head of a young worm. They do not grow further or migrate. Common locations are the subcutaneous tissue, skeletal muscles and brain.

**Clinical features**

When superficially placed, cysts can be palpated under the skin or mucosa as pea-like ovoid bodies. Here they cause few or no symptoms, and will eventually die and become calcified.

Heavy brain infections, especially in children, may cause features of encephalitis. More commonly, however, cerebral signs do not occur until the larvae die, 5–20 years later. Epilepsy, personality changes, staggering gait or signs of hydrocephalus are the most common features.

**Investigations**

Calcified cysts in muscles can be recognised radiologically. In the brain, however, less calcification takes place and larvae are only occasionally visible by plain X-ray; usually CT or MRI will show them. Epileptic fits starting in adult life suggest the possibility of cysticercosis if the patient has lived in or travelled to an endemic area. The subcutaneous tissue should be palpated and any nodule excised for histology. Radiological examination of the skeletal muscles may be helpful. Antibody detection is available for serodiagnosis.

**Management and prevention**

Albendazole, 15 mg/kg daily for a minimum of 8 days, has now become the drug of choice for parenchymal neurocysticercosis. Praziquantel is another option, 50 mg/kg in three divided doses daily for 10 days. Prednisolone, 10 mg 3 times daily, is also given for 14 days, starting 1 day before the albendazole or praziquantel. In addition, anti-epileptic drugs should be given until the reaction in the brain has subsided. Operative intervention is indicated for hydrocephalus. Studies from India and Peru suggest that most small, solitary cerebral cysts will resolve without treatment.

**Echinococcus granulosus (Taenia echinococcus) and hydatid disease**

Dogs are the definitive hosts of the tiny tapeworm *E. granulosus*. The larval stage, a hydatid cyst, normally occurs in sheep, cattle, camels and other animals that are infected from contaminated pastures or water. By handling a dog or drinking contaminated water, humans may ingest eggs (Fig. 13.56). The embryo is liberated from the ovum in the small intestine and gains access to the blood stream and thus to the liver. The resultant cyst grows very slowly, sometimes intermittently. It is composed of an enveloping fibrous pericyst, laminated hyaline membrane (ectocyst) and inner germinal layers (endocyst) which gives rise to daughter cysts, or germinating cystic brood capsule in which larvae (protoscolices) develop. Over time, some cysts may calcify and become non-viable. The disease is common in the Middle East, North and East Africa, Australia and Argentina. Foci of infection persist in the UK in rural Wales and Scotland. *E. multilocularis*, which has a cycle between foxes and voles, causes a similar but more severe

---

**Fig. 13.56** Hydatid disease. A Life cycle of *Echinococcus granulosus*. B Daughter cysts removed at surgery. C Within the daughter cysts are the protoscolices.
infection, ‘alveolar hydatid disease’, which invades the liver like cancer.

**Clinical features**

A hydatid cyst is typically acquired in childhood and may, after growing for some years, cause pressure symptoms. These vary, depending on the organ or tissue involved. In nearly 75% of patients with hydatid disease, the right lobe of the liver is invaded and contains a single cyst (p. 956). In others, a cyst may be found in lung, bone, brain or elsewhere.

**Investigations**

The diagnosis depends on the clinical, radiological and ultrasound findings in a patient who has lived in close contact with dogs in an endemic area. Complement fixation and ELISA are positive in 70–90% of patients.

**Management and prevention**

Hydatid cysts should be excised wherever possible. Great care is taken to avoid spillage and cavities are sterilised with 0.5% silver nitrate or 2.7% sodium chloride. Albendazole (400 mg twice daily for 3 months) should also be used. The drug is now often combined with PAIR (percutaneous puncture, aspiration, injection of scleroidal agent and re-aspiration) to good effect. Praziquantel (20 mg/kg twice daily for 14 days) also kills protoscolices perioperatively.

Prevention is difficult in situations where there is a close association with dogs and sheep. Personal hygiene, satisfactory disposal of carcasses, meat inspection and deworming of dogs can greatly reduce the prevalence of disease.

**Other tapeworms**

There are many other cestodes whose adult or larval stages may infect humans. Sparganosis is a condition in which an immature worm develops in humans, usually subcutaneously, as a result of eating or applying to the skin the secondary or tertiary intermediate host.

---

**ECTOPARASITES**

Ectoparasites only interact with the outermost surfaces of the host; see also page 1280.

**Jiggers (tunga)**

This is widespread in tropical America and Africa, and is caused by the sand flea *Tunga penetrans*. The pregnant flea burrows into the skin around toes and produces large numbers of eggs. The burrows are intensely irritating and the whole inflammatory nodule should be removed with a sterile needle. Secondary infection of tunga lesions is common.

**Myiasis**

Myiasis is due to skin infestation with larvae of the South American botfly, *Dermatobia hominis*, or the African Tumbu fly, *Cordylobia anthropophaga*. The larvae develop in a subcutaneous space with a central sinus. This orifice is the air source for the larva, and periodically the larval respiratory spiracles protrude through the sinus. Patients with myiasis feel movement within the larval burrow and can experience intermittent sharp, lancinating pains. Myiasis is diagnosed clinically and should be suspected with any furuncular lesion accompanied by pain and a crawling sensation in the skin. The larva may be suffocated by blocking the respiratory orifice with petroleum jelly and gently removing it with tweezers. Secondary infection of myiasis is remarkably infrequent and rapid healing follows removal of intact larva.

---

**FUNGAL INFECTIONS**

Fungal infections, or mycoses, are classified as superficial, subcutaneous or systemic (deep), depending on the degree of invasion of the host. They are also classified by the kind of fungus that causes the infection, which may be a filamentous fungus (mould) or a yeast, or may vary between these two forms, depending on the environmental conditions (dimorphic fungi; Fig. 13.57).

**Superficial mycoses**

Superficial cutaneous fungal infections caused by dermatophyte fungi are described in Chapter 28.

**Candidiasis (thrush)**

Superficial candidiasis is caused by *Candida* spp., mainly *C. albicans*. Manifestations include oropharyngeal (pp. 864 and 399) and vaginal candidiasis (‘thrush’), intertrigo and chronic paronychia. Superficial candidiasis often follows antibiotic therapy. Intertrigo is characterised by inflammation in skin folds with surrounding ‘satellite lesions’. Chronic paronychia is associated with frequent wetting of the hands. Superficial candidiasis is treated mainly with topical azoles (p. 159), oral azoles being reserved for refractory or recurrent disease. Severe oropharyngeal and oesophageal candidiasis is a consequence of CD4+ T lymphocyte depletion/dysfunction, as in HIV infection (p. 399). Recurrent vaginal or penile candidiasis may be a manifestation of diabetes mellitus. Rarely, mutations in the autoimmune regulator gene, *AIRE*, cause a syndrome of chronic mucocutaneous candidiasis (p. 796).

**Subcutaneous mycoses**

**Chromoblastomycosis**

Chromoblastomycosis is a predominantly tropical or subtropical fungal disease caused by environmental dematiaceous (dark-pigmented) fungi, most commonly *Fonsecaea pedrosii*. Other causes include *F. compacta*, *Cladophialophora carrionii* and *Phialophora verrucosa*. The disease is a cutaneous/subcutaneous mycosis acquired by traumatic inoculation. Consequently, the most commonly affected areas are the foot, ankle and lower leg. Lesions may start several months after the initial injury, and medical attention is often sought several years later. The initial lesion is a papule. Further papules develop, and coalesce to form irregular plaques. Nodular lesions may produce a characteristic ‘cauliflower’ appearance.

Diagnosis is by histopathological examination of infected material, which shows dematiaceous, rounded, thick-walled ‘sclerotic bodies’ with septa at right angles
Filamentous fungi (moulds)

Characterised by the production of elongated, cylindrical, often septate cells (hyphae) and conidia (spores)

Examples:
- Aspergillus spp. (A. fumigatus shown here)
- Fusarium spp.
- Dermatophyte fungi (Tricophyton spp., Microsporum spp. etc.)
- Mucorales

Dimorphic fungi

Exist in filamentous (top) or yeast (bottom) form, depending on environmental conditions

Examples:
- Histoplasma capsulatum,
- Coccioides immitis, Paracoccidioides brasiliensis (shown here), Blastomyces dermatitidis
- Sporothrix schenckii
- Penicillium marneffei
- Malassezia spp.

Yeasts

Characterised by the production of oval or round cells, which reproduce by binary fission (budding)

Examples:
- Candida spp.*
- Cryptococcus spp. (C. neoformans shown here)

**Fig. 13.57 Classification of medically important fungi.** Fungal classification is based on simple morphological characteristics. *Pneumocystis jirovecii (carinii) is morphologically distinct from other fungi and does not fit into this classification. Although Candida albicans exists in a number of forms, including filamentous (hyphae and pseudohyphae), it is generally encountered in its yeast form, so is classified in this category.

Infectious Disease

to each other. The aetiological agent is confirmed by culture. Many therapeutic approaches have been explored, including antifungal agents, cryosurgery and surgical excision, alone or in combination, but the optimal therapy is unknown. Of the antifungal agents, itraconazole and terbinafine are considered to be the most effective. However, posaconazole has also been used with a good outcome.

**Mycetoma**

Mycetoma is a chronic suppurative infection of the deep soft tissues and bones, most commonly of the limbs but also of the abdominal or chest wall or head. It is caused by either aerobic or anaerobic branching Gram-positive bacilli, *Actinomyctes* (actinomyctoma – 60%, p. 135), or by true fungi, *Eumycetes* (eumycetoma – 40%). Many fungi cause eumycotomas, the most common being Madurella mycetomatis, *M. grisea*, Leptosphaeria senegalensis and Scedosporium apiospermum. Actinomyctomas are caused by Actinomadura, Nocardia and Streptomyces spp. Both groups produce characteristically coloured grains, the colour depending on the organism (black grains – eumycetoma, red and yellow grains – actinomyctoma, white grains – either). The disease occurs mostly in the tropics and subtropics.

**Clinical features**

The disease is acquired by inoculation (e.g. from a thorn) and most commonly affects the foot (Madura foot). Mycetoma begins as a painless swelling at the implantation site, which becomes chronic and progressive, grows and spreads steadily within the soft tissues, eventually extending into bone. Nodules develop under the epidermis and these rupture, revealing sinuses through which grains (fungal colonies) may be discharged. Sinuses heal with scarring, while fresh sinuses appear elsewhere. Deeper tissue invasion and bone involvement are less rapid and extensive in eumycetoma than actinomyctoma. There is little pain and usually no fever or lymphadenopathy, but there is progressive disability.

**Investigations**

Diagnosis is confirmed by demonstration of fungal grains in pus, and/or histopathological examination of tissue. Culture is necessary for species identification and susceptibility testing. Serological tests are not available.

**Management**

Eumycetoma is usually treated with a combination of surgery and antifungal therapy. Antifungal susceptibility testing, if available, is recommended, although clinical outcome does not necessarily correspond to *in vitro* test results. Itraconazole and ketoconazole (both 200–400 mg/day) are used most commonly. Success has also been reported with terbinafine monotherapy, and refractory cases have responded to both voriconazole and posaconazole. Amphotericin B is not generally considered effective. Therapy is continued for 6–12 months or longer. In extreme cases, amputation may be required. Management of actinomyctoma is described on page 341.

**Phaeohyphomycosis**

Phaeohyphomycoses are a heterogenous group of fungal diseases caused by a large number (more than 70) of dematiaceous fungi. In phaeohyphomycosis, the tissue form of the fungus is predominantly mycelial (filamentous), as opposed to eumycetoma (graín) or chromoblastomycosis (sclerotic body). Disease may be superficial,
subcutaneous or deep. The most serious manifestation is cerebral phaeohyphomycosis, which presents with a ring-enhancing, space-occupying cerebral lesion. Optimal therapy for this condition has not been established, but treatment usually consists of neurosurgical intervention and antifungal (usually triazole) therapy. Causative agents are Cladophialophora bantiana, Fonsecaea spp. and Rhinocladiella mackenzii, which occurs in the Middle East and is usually fatal.

**Sporotrichosis**

Sporotrichosis is caused by *Sporothrix schenckii*, a dimorphic fungal saprophyte of plants in tropical and subtropical regions. Disease is caused by dermal inoculation of the fungus, usually from a thorn (occasionally from a cat scratch). In fixed cutaneous sporotrichosis, a subcutaneous nodule develops at the site of infection and subsequently ulcerates, with a purulent discharge. The disease may then spread along the cutaneous lymphatic channels, resulting in multiple cutaneous nodules along their route, which ulcerate and discharge (lymphocutaneous sporotrichosis). Rarer forms of disease are seen: for example, in patients with cutaneous disease presenting with arthritis. Later, draining sinuses may form. Pulmonary sporotrichosis occurs as a result of inhalation of the conidia, and manifests itself as chronic cavitary fibronodular disease with haemoptysis and constitutional symptoms. Disseminated disease may occur, especially in patients with HIV.

**Investigations**

Typical yeast forms detected on histology of the biopsy confirm the diagnosis but are rarely seen; the fungus can be grown from the specimen in culture. A latex agglutination test is available to detect *S. schenckii* antibodies in serum.

**Management**

Cutaneous and lymphocutaneous disease is treated with oral itraconazole (200–400 mg daily, prescribed as the oral solution, which has better bio-availability than the capsule formulation) for 3–6 months. Alternative agents include a saturated solution of potassium iodide (SSKI, given orally), initiated with 5 drops and increased to 40–50 drops 3 times daily, or terbinafine (500 mg twice daily). Localised hyperthermia may be used in pregnancy (to avoid azole use). Osteoarticular disease requires a longer course of therapy (at least 12 months). Severe or life-threatening disease is treated with amphoterica B (lipid formulation preferred).

**Systemic mycoses**

**Aspergillosis**

Aspergillosis is an opportunistic systemic mycosis, which affects predominantly the respiratory tract. It is described on page 697.

**Candidiasis**

Systemic candidiasis is an opportunistic mycosis caused by *Candida* spp. The most common cause is *C. albicans*. Other agents include *C. dubliniensis*, *C. glabrata*, *C. krusei*, *C. parapsilosis* and *C. tropicalis*. Species distribution varies geographically. *Candida* species identification often enables prediction of susceptibility to fluconazole: *C. krusei* is universally resistant, many *C. glabrata* isolates have reduced susceptibility or are resistant, and other species are mostly susceptible. Candidiasis is usually an endogenous disease that originates from oropharyngeal, genitourinary or skin colonisation, although nosocomial spread has been reported.

**Syndromes of systemic candidiasis**

Acute disseminated candidiasis

This usually presents as candidaemia (isolation of *Candida* spp. from the blood). The main predisposing factor is the presence of a central venous catheter. Other major factors include recent abdominal surgery, total parenteral nutrition (TPN), recent antibiotic therapy and localised *Candida* colonisation. Up to 40% of cases will have ophthalmic involvement, with characteristic retinal ‘cotton wool’ exudates. As this is a sight-threatening condition, candidaemic patients should be assessed by detailed ophthalmoscopy. Skin lesions (non-tender pink/red nodules) may be seen. Although predominantly a disease of intensive care and surgical patients, acute disseminated candidiasis and/or *Candida* endophthalmitis is seen occasionally in injection drug-users, thought to be due to candidal contamination of citric acid or lemon juice used to dissolve heroin.

Chronic disseminated candidiasis (hepatosplenic candidiasis)

In this condition, a neutropenic patient has a persistent fever, despite antibacterial therapy. The fever persists, even though there is neutrophil recovery, and is associated with the development of abdominal pain, raised alkaline phosphatase and multiple lesions in abdominal organs (e.g. liver, spleen and/or kidneys) on radiological imaging. Chronic disseminated candidiasis is a form of immune reconstitution syndrome (p. 138) in patients recovering from neutropenia and usually lasts for several months, despite appropriate therapy.

**Other manifestations**

Renal tract candidiasis, osteomyelitis, septic arthritis, peritonitis, meningitis and endocarditis are all well recognised, and are usually sequelae of acute disseminated disease. Diagnosis and treatment of these conditions require specialist mycological advice.

**Management**

Blood cultures positive for *Candida* spp. must never be ignored. Acute disseminated candidiasis is treated with antifungal therapy, removal of any in-dwelling central venous catheter (whether known to be the source of infection or not) and removal of any known source. Current evidence suggests that candidaemia should be treated initially with an echinocandin (p. 160), with subsequent adjustment (usually to IV or oral fluconazole) guided by clinical response, species identification and susceptibility testing. Treatment should continue for a minimum of 14 days. Other appropriate therapies include voriconazole and amphoterica B formulations.

Chronic disseminated candidiasis requires prolonged treatment over several months with fluconazole or other agents, depending on species and clinical response. The duration of the condition may be reduced by adjuvant therapy with systemic corticosteroids.
Cryptococcosis

Cryptococcosis is a systemic mycosis caused by two environmental yeast species, *Cr. neoformans* and *Cr. gattii*. *Cr. neoformans* is distributed worldwide and is primarily an opportunistic pathogen, most commonly associated with HIV infection. *Cr. gattii* is a primary pathogen with a widespread distribution that includes Australasia, Africa, Canada (Vancouver Island) and the north-western USA.

Cryptococcosis is acquired by inhalation of yeasts. These may disseminate to any organ, most commonly the CNS and skin. The manifestations of *Cr. neoformans* are most severe in immunocompromised individuals. Conversely, *Cr. gattii* causes severe disease most often in immunocompetent hosts. Disseminated cryptococcosis (sepsis with cryptococci present in the blood stream or at multiple sites) is largely restricted to immunocompromised patients. CNS manifestations of cryptococcosis include meningitis (p. 403) and cryptococcoma (Fig. 13.58), the latter more likely with *Cr. gattii* infection. Manifestations of pulmonary cryptococcosis range from severe pneumonia (in immunocompromised patients) to asymptomatic disease with single or multiple pulmonary nodules, sometimes exhibiting cavitation (in immunocompetent patients). Cryptococcal nodules may mimic other causes of lung pathology, such as tuberculosis or malignancy, and diagnosis is often made by histopathology and/or culture.

Treatment of severe cryptococcosis is the same as for cryptococcal meningitis (p. 403). Mild pulmonary disease is usually treated with fluconazole, although, for asymptomatic nodules, resection of the lesions is likely to be sufficient. Relevant guidelines (e.g. www.idsociety.org) should be consulted.

**Fusariosis**

*Fusarium* spp. cause disseminated disease in patients with profound or prolonged neutropenia. The disease presents with antibiotic-resistant fever and evidence of dissemination (e.g. skin nodules, endophthalmitis, septic arthritis, pulmonary disease; Fig. 13.59). In contrast to *Aspergillus* spp., *Fusarium* spp. is often recovered from blood cultures. Treatment is challenging because of resistance to several antifungal agents. Voriconazole, posaconazole and lipid-formulated amphotericin B are the most commonly used antifungal agents.

**Mucormycosis**

Mucormycosis is a severe but uncommon opportunistic systemic mycosis caused by a number of ‘mucoraceous’ moulds, most commonly *Lichtheimia* (formerly *Absidia*) spp., *Rhizomucor* spp., *Mucor* spp. and *Rhizopus* spp. Disease patterns include rhinocerebral/craniofacial, pulmonary, cutaneous and systemic disease. All are characterised by the rapid development of severe tissue necrosis, which is almost always fatal if left untreated. The most common predisposing factors are profound immunosuppression from neutropenia and/or haematopoietic stem cell transplantation, uncontrolled diabetes mellitus, iron chelation therapy with desferrioxamine and severe burns.

Definitive diagnosis is by culture, but histopathological confirmation is required as the fungi may be environmental contaminants. Treatment requires a combination of antifungal therapy and surgical débridement, with correction of predisposing factor(s) if
possible. High-dose lipid-formulated amphotericin B is used most commonly. Posaconazole is active against many mucaceous moulds in vitro and may be used as a second-line agent or as oral ‘step-down’ therapy.

**Penicillium marneffei infection**

*P. marneffei* is a thermally dimorphic pathogen (filamentous in environmental conditions and yeast at body temperature), which causes disease in South-east Asia, mainly in association with HIV infection (although immunocompetent patients may also be infected). Acquisition is most likely to be by inhaling of environmental spores, with primary lung infection followed by haematogenous dissemination. A generalised papular rash, which progresses to widespread necrosis and ulceration, is a characteristic feature. Skin lesions may resemble those of molluscum contagiosum. Diagnosis is by histopathology and/or culture of respiratory secretions, blood or any infected clinical material (e.g. skin lesions, bone marrow, biopsies). Recommended treatment is with an amphotericin B formulation followed by itraconazole (in severe infection), or with itraconazole alone.

**Histoplasmosis**

Histoplasmosis is a primary systemic mycosis caused by the dimorphic fungus *Histoplasma capsulatum*. *H. capsulatum var. capsulatum* is endemic to east-central USA (especially the Mississippi and Ohio river valleys), parts of Canada, Latin America, the Caribbean, East and South-east Asia, and Africa. It occurs sporadically in Australia and India, and is very rare in Europe. *H. capsulatum var. duboisii* is found in West Africa and Madagascar. Genetic analysis suggests that *H. capsulatum* may, in fact, be made up of several different species, and its taxonomy is under review.

**Habitat**

The primary reservoir of *H. capsulatum* is soil enriched by bird and bat droppings, in which the fungus remains viable for many years. Infection is by inhalation of dust from such soil. Natural infections are found in bats, which represent a secondary reservoir of infection (via bat faeces). Histoplasmosis is a specific hazard for explorers of caves and people who clear out bird (including chicken) roosts.

**Pathology**

The organism is inhaled in the form of conidia (spores) or hyphal fragments and transforms to the yeast phase during infection. Conidia or yeasts are phagocytosed by alveolar macrophages and neutrophils, and this may be followed by haematogenous dissemination to any organ. Subsequent development of a T-lymphocyte response brings the infection under control, resulting in a latent state in most exposed individuals.

**Clinical features**

Disease severity depends on the quantity of spores inhaled and the immune status of the host. In most cases, infection is asymptomatic. Pulmonary symptoms are the most common disease presentation, with fever, non-productive cough and an influenza-like illness. Erythema nodosum, myalgia and joint pain are common, and chest radiography may reveal a pneumonitis with hilar or mediastinal lymphadenopathy.

Patients with pre-existing lung disease, such as chronic obstructive pulmonary disease (COPD) or emphysema, may develop chronic pulmonary histoplasmosis. The predominant features of this condition, which may easily be mistaken for tuberculosis, are fever, cough, dyspnoea, weight loss and night sweats. Radiological findings include fibrosis, nodules, caviation and hilar/mediastinal lymphadenopathy.

Disease caused by *H. capsulatum var. duboisii* presents more commonly with papulonodular and ulcerating lesions of the skin and underlying subcutaneous tissue and bone (sometimes referred to as ‘African histoplasmosis’). Multiple lesions of the ribs are common and the bones of the limbs may be affected. Lung involvement is relatively rare. Radiological examination may show rounded foci of bone destruction, sometimes associated with abscess formation. Other disease patterns include a visceral form with liver and splenic invasion, and disseminated disease.

Acute disseminated histoplasmosis is seen in association with immunocompromise, including HIV infection. Features include fever, pancytopenia, hepatosplenomegaly, lymphadenopathy and often a papular skin eruption. Chronic disseminated disease presents with fever, anorexia and weight loss. Cutaneous and mucosal lesions, lymphadenopathy, hepatosplenomegaly and meningitis may also develop.

**Investigations**

In areas where the disease occurs, histoplasmosis should be suspected in every undiagnosed infection in which there are pulmonary signs, enlarged lymph nodes, hepatosplenomegaly or characteristic cutaneous/bony lesions. Radiological examination in long-standing cases may show calcified lesions in the lungs, spleen or other organs. In the more acute phases of the disease, single or multiple soft pulmonary shadows with enlarged tracheobronchial nodes are seen on chest X-ray.

Laboratory diagnosis is by direct detection (histopathology or antigen detection), culture and serology; although antigen detection is the most effective method, it is not widely available. Antibody is detected by complement fixation testing or immunodiffusion; the pattern of antibody production is complex and the results require specialist interpretation. *Histoplasma* antigen may be detectable in blood or urine. Culture is definitive but slow (up to 12 weeks). Histopathology may show characteristic intracellular yeasts. Diagnosis of subcutaneous or bony infection is mainly by histopathological examination and/or culture.

**Management**

Mild pulmonary disease does not require treatment. However, if prolonged, it may be treated with itraconazole. More severe pulmonary disease is treated with an amphotericin B formulation for 2 weeks, followed by itraconazole for 12 weeks, with methylprednisolone added for the first 2 weeks of therapy if there is hypoxia or ARDS. Chronic pulmonary histoplasmosis is treated with itraconazole (prescribed as the oral solution) for 12–24 months, and disseminated histoplasmosis with an amphotericin B formulation followed by itraconazole. Lipid formulations of amphotericin B are preferred, but their use is subject to availability. Treatment should be guided by current evidence-based guidelines (e.g.
INFECTIONOUS DISEASE

Infectious Diseases Society of America practice guidelines: www.idsociety.org. In subcutaneous and bone infection, patterns of remission and relapse are more common than cure. A solitary bony lesion may require local surgical treatment only.

Coccidioidomycosis

This is a primary systemic mycosis caused by the dimorphic fungi *Coccidioides immitis* and *C. posadasi*, found in the south-western USA, and Central and South America. The disease is acquired by inhalation of conidia (arthrospores). In 60% of cases it is asymptomatic, but in the remainder it affects the lungs, lymph nodes and skin. Rarely (in approximately 0.5%), it may spread haemato-genously to bones, adrenal glands, meninges and other organs. Pulmonary coccidioidomycosis has two forms: primary and progressive. If symptomatic, primary coccidioidomycosis presents with cough, fever, chest pain, dyspnoea and (commonly) arthritis and a rash (erythema multiforme). Progressive disease presents with systemic upset (e.g. fever, weight loss, anorexia) and features of lobar pneumonia, and may resemble tuberculosis.

*Coccidioides* meningitis (which may be associated with CSF eosinophils) is the most severe disease manifestation; it is fatal if untreated, and requires life-long suppressive therapy with antifungal azoles.

**Investigations and management**

Diagnosis is by direct detection (histopathological examination of infected tissue), culture of infected tissue or fluids, or antibody detection. IgM may be detected after 1–3 weeks of disease by precipitin tests. IgG appears later and is detected with the complement fixation test. Change in IgG titre may be used to monitor clinical progress.

Treatment depends on specific disease manifestations, and ranges from regular clinical re-assessment without antifungal therapy (in mild pulmonary, asymptomatic cavitary or single nodular disease) to high-dose treatment with an antifungalazole, which may be continued indefinitely (e.g. in meningitis). Amphotericin B is used in diffuse pneumonia, disseminated disease and, intrathecally, in meningitis. Posaconazole has been used successfully in refractory disease.

Paracoccidioidomycosis

This is a primary systemic mycosis caused by inhalation of the dimorphic fungus *Paracoccidioides brasiliensis*, which is restricted to South America. The disease affects the lungs, mucous membranes (painful destructive ulceration in 50% of cases), skin, lymph nodes and adrenal glands (hypoadrenalism). Diagnosis is by microscopy and culture of lesions, and antibody detection. Oral itraconazole solution (200 mg/day) has demonstrated 98% efficacy and is currently the treatment of choice (mean duration 6 months). Ketoconazole, fluconazole and voriconazole have also been used, as have long (2–3-year) courses of sulphonamides. Amphotericin B may be used in severe or refractory disease, followed by an azole or sulphonamide.

Blastomycosis

*Blastomyces dermatitidis* is a dimorphic fungus endemic to restricted parts of North America, mainly around the Mississippi and Ohio rivers. Very occasionally, it is reported from Africa. The disease usually presents as a chronic pneumonia similar to pulmonary tuberculosis. Bones, skin and the genitourinary tract may also be affected. Diagnosis is by culture of the organism or identification of the characteristic yeast form in a clinical specimen. Antibody detection is rarely helpful. Treatment is with amphotericin B (severe disease) or itraconazole.

Further information and acknowledgements

**Websites**


www.cdc.gov Centers for Disease Control, USA. Source of general information about infectious diseases.

www.fitfortravel.nhs.uk Scottish site with valuable information for travellers.

www.hpa.org.uk Health Protection Agency. Information on infectious diseases in the UK.


www.who.int World Health Organization. Invaluable links on travel medicine with updates on outbreaks of infections, changing resistance patterns and vaccination requirements.

**Figure acknowledgements**

Fig. 13.12 Reproduced from Halstead SB. Dengue. Medicine 1997; 25:1 and Month MP Yellow fever. Medicine 1997; 25:1. Copyright Elsevier and Dr TP Month.

Fig. 13.23 Reproduced by permission of WHO.

Fig. 13.27 Based on Bryceson ADM, Pfalzgraf RE. Leprosy. 3rd edn. Edinburgh: Churchill Livingstone; 1990; copyright Elsevier.


Fig. 13.47 Gibbons LM. SEM guide to the morphology of nematode parasites of vertebrates. Farnham Royal, Slough: Commonwealth Agricultural Bureau International; 1986.

Fig. 13.53 Cook GC, ed. Manson’s tropical diseases. 20th edn. London: WB Saunders; 1995; copyright Elsevier.
HIV infection and AIDS

Clinical examination in HIV disease 388

Epidemiology 390
- Global epidemic and regional patterns 390
- Modes of transmission 390

Virology and immunology 391

Diagnosis and investigations 392
- Diagnosing HIV infection 392
- Viral load and CD4 counts 393

Natural history and staging of HIV 394

Presenting problems in HIV infection 395
- Lymphadenopathy 395
- Weight loss 396
- Fever 396
- Mucocutaneous disease 396
- Gastrointestinal disease 399
- Hepatobiliary disease 400
- Respiratory disease 400
- Nervous system and eye disease 402
- Rheumatological disease 404
- Haematological abnormalities 404
- Renal disease 405
- Cardiac disease 405
- HIV-related cancers 405

Prevention of opportunistic infections 405
- Preventing exposure 405
- Chemoprophylaxis 406
- Immunisation 407

Antiretroviral therapy 407
- ART complications 408
- ART in special situations 409
- Prevention of HIV 410
Clinical Examination in HIV Disease

- **Neck**
  - Lymph node enlargement
  - Tuberculosis
  - Lymphoma
  - Kaposi’s sarcoma
  - Persistent generalised lymphadenopathy
  - Parotidomegaly

- **Eyes**
  - Retina
  - Toxoplasmosis
  - HIV retinopathy
  - Progressive outer retinal necrosis

- **Central nervous system**
  - Higher mental function
  - HIV dementia
  - Progressive multifocal leuкоencephalopathy
  - Focal signs
  - Toxoplasmosis
  - Primary CNS lymphoma
  - Neck stiffness
  - Cryptococcal meningitis
  - Tuberculous meningitis

- **Chest**
  - Lungs
  - Pleural effusion
  - Tuberculosis
  - Kaposi’s sarcoma
  - Parapneumonic

- **Abdomen**
  - Hepatosplenomegaly

- **Anogenital region**
  - Rashes

- **Lungs**
  - Pleural effusion
  - Tuberculosis
  - Kaposi’s sarcoma
  - Parapneumonic

- **Skin**
  - Papular pruritic eruption
  - Molluscum contagiosum
  - Herpes zoster
  - Seborrhoeic dermatitis

- **Observation**
  - Weight loss
  - Tachypnoea
  - Fevers and sweats

- **Legs**
  - Peripheral nerve examination
  - Spastic paraparesis
  - Peripheral neuropathy
# HIV clinical staging classifications

<table>
<thead>
<tr>
<th>World Health Organization (WHO) clinical stage (used in low- and middle-income countries)</th>
<th>Centers for Disease Control (CDC) clinical categories (used in high-income countries)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1</strong></td>
<td><strong>Category A</strong></td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>Primary HIV infection</td>
</tr>
<tr>
<td>Persistent generalised lymphadenopathy</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td><strong>Stage 2</strong></td>
<td><strong>Category B</strong></td>
</tr>
<tr>
<td>Unexplained moderate weight loss (&lt; 10% of body weight)</td>
<td>Bacillary angiomatosis</td>
</tr>
<tr>
<td>Recurrent upper respiratory tract infections</td>
<td>Candidiasis, oropharyngeal (thrush)</td>
</tr>
<tr>
<td>Herpes zoster</td>
<td>Candidiasis, vulvovaginal; persistent, frequent or poorly responsive to therapy</td>
</tr>
<tr>
<td>Angular cheilitis</td>
<td>Cervical dysplasia (moderate or severe)/cervical carcinoma in situ</td>
</tr>
<tr>
<td>Recurrent oral ulceration</td>
<td>Constitutional symptoms, such as fever (38.5°C) or diarrhoea lasting &gt; 1 mth</td>
</tr>
<tr>
<td>Papular pruritic eruptions</td>
<td>Oral hairy leucoplakia</td>
</tr>
<tr>
<td>Seborrhoeic dermatitis</td>
<td>Herpes zoster, involving two distinct episodes or more than one dermatome</td>
</tr>
<tr>
<td>Fungal nail infections</td>
<td>Idiopathic thrombocytopenic purpura</td>
</tr>
<tr>
<td><strong>Stage 3</strong></td>
<td><strong>Category C</strong></td>
</tr>
<tr>
<td>Unexplained severe weight loss (&gt; 10% of body weight)</td>
<td>Candidiasis of oesophagus, trachea, bronchi or lungs</td>
</tr>
<tr>
<td>Unexplained chronic diarrhoea for &gt; 1 mth</td>
<td>Cervical carcinoma – invasive</td>
</tr>
<tr>
<td>Unexplained persistent fever (&gt; 37.5°C for &gt; 1 mth)</td>
<td>Cryptococcosis – extrapulmonary</td>
</tr>
<tr>
<td>Persistent oral candidiasis</td>
<td>Cryptosporidiosis, chronic (&gt; 1 mth)</td>
</tr>
<tr>
<td>Oral hairy leucoplakia</td>
<td>Cytomegalovirus disease (outside liver, spleen and nodes)</td>
</tr>
<tr>
<td>Pulmonary tuberculosis</td>
<td>Herpes simplex chronic (&gt; 1 mth) ulcers or visceral</td>
</tr>
<tr>
<td>Severe bacterial infections</td>
<td>HIV encephalopathy</td>
</tr>
<tr>
<td>Acute necrotising ulcerative stomatitis, gingivitis or periodontitis</td>
<td>HIV wasting syndrome</td>
</tr>
<tr>
<td>Unexplained anaemia (&lt; 80 g/L (8 g/dL)), neutropenia (&lt; 0.5 × 10^9/L) and/or chronic thrombocytopenia (&lt; 50 × 10^9/L)</td>
<td>Isosporiasis, chronic (&gt; 1 mth)</td>
</tr>
<tr>
<td><strong>Stage 4</strong></td>
<td><strong>Symptomatic HIV-associated nephropathy</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Symptomatic HIV-associated cardiomyopathy</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Leishmaniasis, atypical disseminated</strong></td>
</tr>
</tbody>
</table>

*These conditions are in WHO stage 4 but not in CDC category C.*
HIV INFECTION AND AIDS

EPIDEMIOLOGY

The acquired immunodeficiency syndrome (AIDS) was first recognised in 1981, although the earliest documented case of HIV infection has been traced to a blood sample from the Democratic Republic of Congo in 1959. AIDS is caused by the human immunodeficiency virus (HIV), which progressively impairs cellular immunity. The origin of HIV is a zoonotic infection with simian immunodeficiency viruses (SIV) from African primates, probably first infecting local hunters. SIVs do not cause disease in their natural primate hosts. HIV-1 was transmitted from chimpanzees and HIV-2 from sooty mangabeys. HIV-1 is the cause of the global HIV pandemic, while HIV-2, which causes a similar illness to HIV-1 but progresses more slowly and is less transmissible, is restricted mainly to western Africa. It has been estimated from mutation rates of SIV and HIV that both HIV-1 and HIV-2 first infected humans about 100 years ago.

There are three groups of HIV-1 representing three separate transmission events from chimpanzees: M (‘major’, worldwide distribution), O (‘outlier’) and N (‘non-major and non-outlier’). Groups O and N are restricted to West Africa. Group M consists of nine subtypes: A-D, F-H, J and K (subtypes E and I were subsequently shown to be recombinants of other subtypes). Globally, subtype C (Africa and India) accounts for half of strains and appears to be more readily transmitted. Subtype B predominates in Western Europe, the Americas and Australia. In Europe, the prevalence of non-B subtypes is increasing because of migrants (predominantly from Africa). Subtypes A and D are associated with slower and faster disease progression respectively.

Global epidemic and regional patterns

In 2011 it was estimated that there were 34.2 million people living with HIV/AIDS, 2.5 million new infections and 1.7 million deaths (Fig. 14.1). Globally, new infections have declined by 20% over the last 10 years. Not all regions have experienced reductions in new infections and the dominant modes of transmission also vary regionally (Box 14.1). Expanding access to combination antiretroviral therapy (ART) has resulted in a 24% decline in global AIDS-related deaths since the peak in 2005. The improved life expectancy on ART has resulted in an increase in the number of people living with HIV. Despite these encouraging epidemiological data, HIV remains an important cause of death globally and has caused over 30 million deaths since the epidemic started. HIV has had a devastating effect in sub-Saharan Africa, particularly in southern African where average life expectancy of the general population fell to below 40 years.

Modes of transmission

HIV is transmitted by sexual contact, by exposure to blood (e.g. injection drug use, occupational exposure in
health-care workers) and blood products, or to infants of HIV-infected mothers (who may be infected in utero, perinatally or via breastfeeding). Worldwide, the major route of transmission is heterosexual. The risk of contracting HIV after exposure to infected body fluid is dependent on the integrity of the exposed site, the type and volume of fluid, and the level of viraemia in the source person. The approximate transmission risk after exposure is given in Box 14.2. Factors that increase the risk of transmission are listed in Box 14.3.

A high proportion of patients with haemophilia in high-income countries had been infected through contaminated blood products by the time HIV antibody screening was adopted in 1985. Routine screening of blood and blood products for HIV infection using antibody and antigen tests (or polymerase chain reaction, PCR) has virtually eliminated this as a mode of transmission. However, the World Health Organization (WHO) estimates that, because of the lack of adequate screening facilities in resource-poor countries, 5–10% of blood transfusions globally are with HIV-infected blood.

**Virology and immunology**

HIV is an enveloped ribonucleic acid (RNA) retrovirus from the lentivirus family. After mucosal exposure, HIV is transported to the lymph nodes via dendritic cells, where infection becomes established. This is followed by viraemia and dissemination to lymphoid organs, which are the main sites of viral replication.

Each mature virion has a lipid membrane lined by a matrix protein that is studded with glycoprotein (gp) 120 and gp41 spikes. The inner cone-shaped protein core (p24) houses two copies of the single-stranded RNA genome and viral enzymes. The HIV genome consists of three characteristic retroviral genes – _gag_ (encodes a polypeptide that is processed into structural proteins, including p24), _pol_ (codes for the enzymes reverse transcriptase, integrase and protease) and _env_ (codes for envelope proteins gp120 and gp41) – as well as six regulatory genes ( _vif, vpr, vpu, nef, tat_ and _rev_).

HIV can only infect cells bearing the CD4 receptor; these are T-helper lymphocytes, monocyte-macrophages, dendritic cells, and microglial cells in the central nervous system (CNS). Entry into the cell commences with binding of gp120 to the CD4 receptor (stage 1, Fig. 14.2), which results in a conformational change in gp120 that permits binding to one of two chemokine co-receptors (CXCR4 or CCR5: stage 2). The chemokine co-receptor CCR5 is utilised during initial infection, but later on the virus may adapt to use CXCR4. Individuals who are homozygous for the CCR5 delta 32 mutation do not express CCR5 on CD4 cells and are immune to HIV infection. Chemokine receptor binding is followed by membrane fusion and cellular entry involving gp41 (stage 3). After penetrating the cell and uncoating, a deoxyribonucleic acid (DNA) copy is transcribed from the RNA genome by the reverse transcriptase (RT) enzyme (stage 4) that is carried by the infecting virion. Reverse transcription is an error-prone process and multiple mutations arise with ongoing replication, which results in considerable viral genetic heterogeneity. Viral DNA is transported into the nucleus and integrated within the host cell genome by the integrase enzyme (stage 5). Integrated virus is known as proviral DNA and persists for the life of the cell. Cells infected with proviral HIV DNA produce new virions only if they undergo cellular activation, resulting in the transcription of viral messenger RNA (mRNA) copies (stage 6), which are then translated into viral peptide chains (stage 7). The precursor polypeptides are then cleaved by the viral protease enzyme to form new viral structural proteins and enzymes (stage 8). These then migrate to the cell surface and are assembled using the host cellular apparatus to produce infectious viral particles, which bud from the cell surface, incorporating the host cell membrane into the viral envelope (stage 9). The mature virion then infects other CD4 cells and the process is repeated. CD4 lymphocytes that are replicating HIV have a very short

---

### 14.2 Risk of HIV transmission after single exposure to an HIV-infected source

<table>
<thead>
<tr>
<th>HIV exposure</th>
<th>Approximate risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexual</td>
<td></td>
</tr>
<tr>
<td>Vaginal intercourse: female to male</td>
<td>0.05%</td>
</tr>
<tr>
<td>Vaginal intercourse: male to female</td>
<td>0.1%</td>
</tr>
<tr>
<td>Anal intercourse: receptive</td>
<td>0.05%</td>
</tr>
<tr>
<td>Anal intercourse: insertive</td>
<td>0.5%</td>
</tr>
<tr>
<td>Oral intercourse: insertive</td>
<td>0.005%</td>
</tr>
<tr>
<td>Oral intercourse: receptive</td>
<td>0.01%</td>
</tr>
<tr>
<td>Blood exposure</td>
<td>90%</td>
</tr>
<tr>
<td>Blood transfusion</td>
<td>15%</td>
</tr>
<tr>
<td>Intravenous drug users sharing needles</td>
<td>0.67%</td>
</tr>
<tr>
<td>Percutaneous needle stick injury</td>
<td>0.3%</td>
</tr>
<tr>
<td>Mucous membrane splash</td>
<td>0.09%</td>
</tr>
<tr>
<td>Mother to child</td>
<td></td>
</tr>
<tr>
<td>Vaginal delivery</td>
<td>0.5%</td>
</tr>
<tr>
<td>Breastfeeding (per month)</td>
<td>15%</td>
</tr>
</tbody>
</table>

### 14.3 Factors increasing the risk of transmission of HIV

**Common to all transmission categories**
- High viral load

**Sexual transmission**
- STIs, especially genital ulcers
- Cervical ectopy
- Rectal or vaginal lacerations
- Menstruation
- Uncircumcised male partner
- Depot intramuscular progesterone contraceptive use

**Injection drug use transmission**
- Sharing equipment
- Linked commercial sex
- Intravenous use
- Concomitant cocaine use
- Incarceration

**Occupational transmission**
- Deep injury
- Visible blood on device
- Needle was in a blood vessel

**Vertical transmission**
- Older gestational age
- Prolonged rupture of membranes

(STIs = sexually transmitted infections)
Infection and a

survival time of about 1 day. It has been estimated that in asymptomatic HIV-infected people, more than $10^{10}$ virions are produced and $10^9$ CD4 lymphocytes destroyed each day.

A small percentage of T-helper lymphocytes enter a post-integration latent phase. Latently infected cells are important as sanctuary sites from antiretroviral drugs, which only act on replicating virus. Current ART is unable to eradicate HIV infection due to the persistence of proviral DNA in long-lived latent CD4 cells. Novel HIV eradication strategies are being devised to target latently infected cells.

The host immune response to HIV infection is both humoral, with the development of antibodies to a wide range of antigens, and cellular, with a dramatic expansion of HIV-specific CD8 cytotoxic T lymphocytes, resulting in a CD8 lymphocytosis and reversal of the usual CD4:CD8 ratio. CD8 cytotoxic T lymphocytes kill activated CD4 cells that are replicating HIV, but not latently infected CD4 cells. HIV evades destruction despite this vigorous immune response in part because the highly conserved regions of gp120 and gp41 that are necessary for viral attachment and entry are covered by highly variable protein loops that change over time as a result of mutations selected for by the immune response. The initial peak of viraemia in primary infection settles to a plateau phase of persistent chronic viraemia. With time, there is gradual attrition of the T-helper lymphocyte population and, as these cells are pivotal in orchestrating the immune response, the patient becomes susceptible to opportunistic diseases. The predominant opportunist infections in HIV-infected people are the consequences of impaired cell-mediated rather than antibody-mediated immunity (e.g. mycobacteria, herpesviruses). However, there is also a B lymphocyte defect with impaired antibody production to new antigens and dysregulated antibody production with a polyclonal increase in gamma globulins, resulting in an increased risk of infection with encapsulated bacteria, notably *Streptococcus pneumoniae*.

**DIAGNOSIS AND INVESTIGATIONS**

**Diagnosing HIV infection**

Globally, the trend is towards universal HIV testing, rather than testing patients at high risk or those with manifestations of HIV infection only. However, in the UK, testing is still targeted (Box 14.4). HIV is diagnosed by detecting host antibodies either by using rapid point-of-care tests or in the laboratory, where enzyme-linked immunosorbent assay (ELISA) tests are usually done. Most tests detect antibodies to both HIV-1 and HIV-2. A positive antibody test from two different immunoassays is sufficient to confirm infection. Western blot assays can also be used to confirm infection, but they are expensive.
and sometimes yield indeterminate results. Screening tests often include a test for p24 antigen in addition to antibodies, in order to detect patients with primary infection before the antibody response occurs. Nucleic acid amplification tests (usually PCR) to detect HIV-RNA are used to diagnose infections in infants of HIV-infected mothers, who carry maternal antibodies to HIV for up to 15 months irrespective of whether they are infected, and to diagnose primary infection before antibodies have developed. (PCR is more sensitive than p24 antigen detection, but p24 is more widely available.)

The purpose of HIV testing is not simply to identify infected individuals, but also to educate people about prevention and transmission of the virus. Counselling is essential both before HIV testing and after the result is obtained (Boxes 14.5 and 14.6). There are major advantages to using rapid point-of-care HIV tests in that pre- and post-test counselling can be done at the same visit. Counselling should always be given in the client’s home language.

A number of baseline investigations should be done at the initial medical evaluation (Box 14.7). The extent of these investigations will depend on the resources available.

Viral load and CD4 counts

CD4 counts

CD4 lymphocyte counts are usually determined by flow cytometry, but cheaper methods have been developed for low-income countries. The CD4 count is the most clinically useful laboratory indicator of the degree of immune suppression and is used, together with clinical staging, in decisions to start ART and prophylaxis against opportunistic infections, and in the differential diagnosis of clinical problems.

The CD4 count varies by up to 20% from day to day and is also transiently reduced by intercurrent infections. Due to this variability, major therapeutic decisions should not be taken on the basis of a single count. This is particularly important when ART is being initiated in patients who do not fulfil the clinical criteria to start ART. The percentage of lymphocytes that are CD4+, rather than the absolute count, is routinely used in paediatrics, as the normal CD4 counts in infants and young children are much higher. In adults, the CD4 percentage is occasionally useful when evaluating significant reductions in an individual’s CD4 count, which may be associated with transient lymphopenia due to intercurrent infection or pregnancy. In this case, the CD4 percentage will be unchanged.

14.4 HIV testing in the UK

Settings where recommended to all

• Genitourinary medicine or sexual health clinics
• Antenatal services
• Termination of pregnancy services
• Drug dependency programmes
• Services for those with hepatitis B, hepatitis C, tuberculosis and lymphoma

Patients routinely offered and recommended testing

• All those presenting with a possible primary HIV infection or where HIV enters the differential diagnosis
• All with an STI
• All sexual partners of an HIV-positive individual
• All MSM and female sexual contacts of MSM
• All with a history of injecting drug use
• All from a country of high HIV prevalence
• All who report sexual contact abroad or in the UK with individuals from a country of high HIV prevalence (>1%)

Settings where testing should be considered, in areas where HIV prevalence in the local population is ≥2 in 1000 population

• All registering with a general practice
• All general medical admissions

Groups in whom testing is routinely performed

• Blood donors
• Dialysis patients
• Organ transplant donors and recipients

14.6 How to carry out post-test counselling

Test result negative

• Discuss transmission and need for behaviour modification
• Advise second test 3 mths after last exposure

Test result positive

• Explain meaning of result
• Organise medical follow-up
• Assess coping strategy
• Stress importance of disclosure
• Explain value of antiretroviral therapy
• Provide written information and useful Internet resources
• Discuss confidentiality issues
• Organise emotional and practical support (provide names/phone numbers)
• Facilitate notification of sexual partners

14.7 Baseline investigations

• CD4 count
• Viral load
• Hepatitis B surface antigen
• Hepatitis C antibody (injection drug users)
• Liver function tests
• Full blood count
• Urinalysis and serum creatinine
• Syphilis serology
• Cervical smear in women
• Serum cryptococcal antigen (if CD4 < 100)
• Tuberculin skin test
• STI screen

14.5 How to carry out pre-test counselling

• Discuss meaning of positive and negative test results
• Realise importance of maintaining confidentiality
• Identify person to whom positive result could be disclosed
• Explore knowledge and explain natural history of HIV
• Discuss transmission and risk reduction
• Assess coping strategy
• Explain test procedure
• Obtain informed consent
The normal CD4 count is > 500 cells/mm³. The rate of decline in CD4 count is highly variable. People with CD4 counts between 200 and 500 cells/mm³ have a low risk of developing major opportunistic infections. Morbidity due to inflammatory dermatoses, herpes zoster, oral candidiasis, tuberculosis, bacterial pneumonia and HIV-related immune disorders (e.g. immune thrombocytopenia) becomes increasingly common as CD4 counts decline. Once the count is below 200 cells/mm³, there is severe immune suppression and a high risk of AIDS-defining conditions. It is important to note that patients can be asymptomatic despite very low CD4 counts and that major opportunistic diseases occasionally present with high CD4 counts.

The CD4 count should be performed every 3–6 months in patients not yet eligible for ART and is usually done at similar intervals in patients on ART, together with measurement of the viral load.

Viral load
The level of viraemia is measured by quantitative PCR of HIV-RNA, known as the viral load. Determining the viral load is important for monitoring responses to ART (p. 407) and also has some prognostic value before starting ART. However, many low-income countries are unable to afford viral load measurements. People with high viral loads (e.g. > 100 000 copies/mL) experience more rapid declines in CD4 count, while those with low viral loads (< 1000 copies/mL) usually have slow or even no decline in CD4 counts. There is little point in repeated measurements of viral load before starting ART, as viral loads remain at a relatively stable plateau after primary infection (Fig. 14.3).

Transient increases in viral load occur with intercurrent infections and immunisations, so the test should be done at least 2 weeks afterwards. Viral load results vary because of assay variability and fluctuations within patients. Only changes in viral load of more than 0.5 log₁₀ copies/mL are considered clinically significant. The same laboratory and viral load test manufacturer should be used for follow-up tests in individual patients if possible.

Clinical staging of patients should be done at the initial medical examination, as it provides prognostic information and is a key criterion for initiating ART and prophylaxis against opportunistic infections. Two clinical staging systems are used internationally (p. 389). In both systems, patients are staged according to the most severe manifestation and do not improve their classification. For example, a patient who is asymptomatic following a major opportunistic disease (AIDS) remains at stage 4 or category C of the WHO and CDC systems respectively, and never reverts to earlier stages. Finally, patients do not always progress steadily through all stages and may present with AIDS, having previously been asymptomatic.

Primary infection
Primary infection is symptomatic in more than 50% of cases, but the diagnosis is often missed. The incubation period is usually 2–4 weeks after exposure. The duration of symptoms is variable, but is seldom longer than 2 weeks. The clinical manifestations (Box 14.8) resemble a glandular fever-type illness, but the presence of maculo-papular rash or mucosal ulceration strongly suggests primary HIV infection rather than the other viral causes of glandular fever (p. 320). In infectious mononucleosis due to other viruses, rashes generally only occur if aminopenicillins are given. Atypical lymphocytosis occurs less frequently than in Epstein–Barr virus (EBV) infection. Transient lymphopenia, including CD4 lymphocytes, is found in most cases (see Fig. 14.3), which may result in opportunistic infections, notably oropharyngeal candidiasis. Major opportunistic infections like Pneumocystis jirovecii pneumonia (PJP) may rarely occur. Thrombocytopenia and moderate elevation of liver enzymes are commonly present. The differential diagnosis of primary HIV includes acute EBV, primary cytomegalovirus (CMV) infection, rubella, primary toxoplasmosis and secondary syphilis.

![Fig. 14.3 Virological and immunological progression of untreated HIV infection.](image-url)
Early diagnosis is made by detecting HIV-RNA on PCR or p24 antigenaemia. The appearance of specific anti-HIV antibodies in serum (seroconversion) occurs 2–12 weeks after the development of symptoms. The window period during which antibody tests may be false-negative is prolonged when post-exposure prophylaxis has been used.

**Asymptomatic infection**

A prolonged period of clinical latency follows primary infection, during which infected individuals are asymptomatic. Persistent generalised lymphadenopathy with nodes typically < 2 cm diameter is a common finding. Eventually the lymph nodes regress, with destruction of node architecture as disease advances.

Viraemia peaks during primary infection and then drops as the immune response develops, to reach a plateau about 3 months later. The level of viraemia post-seroconversion is a predictor of the rate of decline in CD4 counts, which is highly variable and explained in part by genetic factors affecting the immune response. The median time from infection to the development of AIDS in adults is about 9 years (see Fig. 14.3). A small proportion of untreated HIV-infected people are long-term non-progressors with CD4 counts in the reference range for 10 years or more. Some long-term non-progressors have undetectable viral loads and are known as ‘elite controllers’.

**Minor HIV-associated disorders**

A wide range of disorders indicating some impairment of cellular immunity occur in most patients before they develop AIDS (CDC category B or WHO stages 2 and 3). Careful examination of the mouth is important when patients are being followed up, as oral candidiasis and oral hairy leucoplaquia are common and important conditions that require initiation of ART and prophylaxis against opportunistic infections, irrespective of the CD4 count.

**Acquired immunodeficiency syndrome**

AIDS is defined by the development of specified opportunistic infections, cancers and severe manifestations of HIV itself (p. 389). CDC category C is the most widely used definition of AIDS. WHO updated its classification more recently and added a few conditions of similar prognosis to its stage 4 disease. **Box 14.9** outlines the correlation between CD4 count and HIV-related diseases.

**PRESENTING PROBLEMS IN HIV INFECTION**

HIV itself is associated with a wide variety of clinical manifestations, and opportunistic diseases add many more. All body systems can be affected by HIV. The CD4 count is useful in differential diagnosis (see **Box 14.9**): opportunistic diseases that may present at higher CD4 counts become increasingly common as CD4 counts decline, so the CD4 count helps to rule out certain disorders. For example, in a patient with a pulmonary infiltrate and a CD4 count of 350 cells/mm$^3$, pulmonary tuberculosis is a likely diagnosis and *Pneumocystis* infection is very unlikely, but if the patient’s CD4 count is 50 cells/mm$^3$, both *Pneumocystis* and tuberculosis are likely.

Globally, tuberculosis is the most common cause of morbidity and mortality in HIV-infected patients. Tuberculosis should be considered in the differential diagnosis of most presenting problems in patients from communities where tuberculosis is common.

**Lymphadenopathy**

Persistent generalised lymphadenopathy due to HIV is described above under asymptomatic infection. Lymphadenopathy may also be due to malignancy (Kaposi’s sarcoma or lymphoma) or infections, especially tuberculosis, which is an extremely common cause in low- and middle-income countries. Rapid enlargement of a node, asymmetric enlargement or lymphadenopathy associated with constitutional symptoms (even if the nodes are symmetrical) warrants further investigation. Tuberculous lymph nodes often undergo extensive caseous necrosis, causing them to become fluctuant, and inexperienced clinicians often inappropriately perform incision and drainage. Lymphoma typically presents with large nodes that are not fluctuant. Lymph node needle aspiration (using a wide-bore needle such as 19G) should be undertaken for microscopy. One slide should be air-dried and sent for staining for acid-fast bacilli, which has about a 70% yield in tuberculosis. The other slide should be fixed and sent for cytology. If caseous liquid is
aspirated, this should be sent for mycobacterial culture. If needle aspiration is unhelpful, or if lymphoma or Kaposi’s sarcoma is suspected, excision biopsy should be performed.

**Weight loss**

Weight loss is a very common finding in advanced HIV infection. The HIV wasting syndrome is an AIDS-defining condition and is defined as weight loss of more than 10% of body weight, plus either unexplained chronic diarrhoea (lasting > 1 month) or chronic weakness and unexplained prolonged fever (lasting > 1 month). This is a diagnosis of exclusion. If the weight loss is rapid (more than 1 kg a month), then major opportunistic infections or cancers become more likely. Painful oral conditions and nausea from drugs contribute by limiting intake. Depression is very common and can cause significant weight loss. Measurement of C-reactive protein is helpful in the work-up of weight loss, as this is markedly raised with most opportunistic diseases but not with HIV itself. Erythrocyte sedimentation rate (ESR) is elevated by HIV infection and is therefore not useful. The presence of fever or diarrhoea is helpful in the differential diagnosis (Fig. 14.4).

**Fever**

Fever is a very common presenting feature. Common causes of prolonged fever with weight loss are listed in Figure 14.4. Non-typhoid *Salmonella* bacteraemia, which commonly presents with fever in low-income countries, presents without diarrhoea in about 50% of patients. Pyrexia of unknown origin (PUO) in HIV infection is defined as temperature over 38°C with no cause found after 4 weeks in outpatients or 3 days in inpatients, and initial investigations such as chest X-rays, urinalysis and blood cultures will have failed to identify the cause. HIV itself can present with prolonged fever, but this is a diagnosis of exclusion, as a treatable cause will be found in most patients. Abdominal imaging, preferably by computed tomography (CT), should be requested. Abdominal nodes (especially if they are hypodense in the centre) or splenic microabscesses strongly suggest tuberculosis. Mycobacterial (or ‘lytic’) blood cultures, which can also detect fungi, should be performed. Bone marrow aspirate and trephine biopsy are helpful if the full blood count shows cytopenias. Liver biopsy may be helpful if the liver enzymes are elevated, but is invasive and seldom necessary. Mycobacterial and fungal stains and cultures should be done on all biopsies. Chest X-rays should be repeated about once a week, as micronodular or interstitial infiltrates may have become apparent (see p. 401 for differential diagnosis).

Tuberculosis is by far the most common cause of PUO in low- and middle-income countries, and in these settings an early trial of empiric therapy is warranted after cultures have been sent. In high-income countries, disseminated *Mycobacterium avium* complex (MAC) infection is an important cause of PUO, often with diarrhoea and splenomegaly. Disseminated endemic mycoses (histoplasmosis, coccidiodomycosis and penicilliosis) present with PUO, often with papular skin eruptions. Skin biopsy for histology and fungal culture is often diagnostic.

**Mucocutaneous disease**

The skin and mouth must be carefully examined, as mucocutaneous manifestations are extremely common in HIV and many prognostically important conditions can be diagnosed by simple inspection. The differential diagnosis of dermatological conditions is simplified by categorising disorders according to the lesion type (Box 14.10). Some common dermatological diseases, notably psoriasis, are exacerbated by HIV. The risk of drug

---

**Fig. 14.4** Presentation and differential diagnosis of weight loss. (AZT = zidovudine; CMV = cytomegalovirus; d4T = stavudine; KS = Kaposi’s sarcoma; MAC = *Mycobacterium avium* complex; NHL = non-Hodgkin lymphoma; PI = protease inhibitor)
Presenting problems in HIV infection

Herpes zoster

This usually presents with a pathognomonic vesicular rash on an erythematous base in a dermatomal distribution (p. 318). The median CD4 count at the first episode of zoster is 350 cells/mm³. In patients with advanced disease, the rash may be multidermatomal and recurrent episodes may occur. Disseminated zoster is rare. In HIV-infected patients, zoster is generally more extensive, has a longer duration, and there is a higher risk of developing post-herpetic neuralgia. High doses of aciclovir or its congeners should be given for all cases with active disease, irrespective of the time since the onset of the rash. Post-herpetic neuralgia is difficult to manage. Analgesic adjuvants, e.g. amitriptyline and pregabalin, should be commenced in all patients with prolonged pain. Topical capsaicin has modest efficacy.

Kaposi’s sarcoma

Kaposi’s sarcoma (KS) is a spindle-cell tumour of lympho-endothelial origin. All forms of KS are due to sexually transmitted human herpesvirus 8, also known as KS-associated herpesvirus. KS occurs in four patterns:
- classic KS: rare, indolent and restricted largely to elderly Mediterranean or Jewish men
- endemic KS: occurs in sub-Saharan Africa, is more aggressive, presents at earlier ages than classic KS, and affects men more than women
- KS in patients on immunosuppressant drugs: usually transplant recipients, who experience disseminated disease
- AIDS-associated KS.

In Africa, the male:female ratio of AIDS-associated KS is much lower than is seen with endemic KS, but men are still more affected than women despite the fact that the seroprevalence of human herpesvirus 8 is the same in both sexes.

AIDS-associated KS is always a multicentric disease. Early mucocutaneous lesions are macular and may be difficult to diagnose. Subsequently, lesions become papular or nodular, and may ulcerate. KS lesions typically have a red-purple colour (Fig. 14.6 and p. 388), but may become hyperpigmented, especially in dark-skinned patients. As the disease progresses, the skin

Seborrhoeic dermatitis

Seborrhoeic dermatitis is very common in HIV. The severity increases as the CD4 count falls. It presents as scaly red patches, typically in the nasolabial folds and in hairy areas. Fungal infections are thought to play a role in the pathogenesis of this condition. It responds well to a combined topical antifungal and steroid. Selenium sulphide shampoo is helpful for scalp involvement.

Herpes simplex infections

Recurrences of herpes simplex infection are very common and primarily affect the nasolabial and anogenital areas (Fig. 14.5). As immune suppression worsens, the ulcers take longer to heal and become more extensive. Ulcers that persist for more than 4 weeks are AIDS-defining. The diagnosis is clinical, but PCR of vesicle fluid or from ulcer swabs may be diagnostic with unusual presentations. Response to a course of aciclovir or a related drug is good, but relapses are common. Frequent relapses that persist despite ART should be treated with aciclovir 400 mg twice daily for 6–12 months.

Scaly rashes
- Seborrhoeic dermatitis
- Psoriasis* (exacerbated by HIV)
- Tinea corporis*
- Dry skin/itchyosis
- Norwegian scabies*
- Drug rashes*

Pruritic papules
- Pruritic papular eruption (‘itchy red bump disease’)
- Eosinophilic folliculitis
- Scabies*

Papules and nodules (non-pruritic)
- Molluscum contagiosum*
- Secondary syphilis
- Kaposi’s sarcoma
- Bacillary angiomatosis
- Cryptococcosis
- Warts*
- Disseminated endemic mycoses (histoplasmosis, coccidioidomycosis and penicilliosis)

Blisters
- Herpes simplex
- Herpes zoster
- Fixed drug eruptions
- Drug rashes (especially toxic epidermal necrolysis)

Mucocutaneous ulcers
- Ecthyma
- Herpes simplex
- Aphthous ulcers (minor and major)
- Histoplasmosis
- Drug rashes (Stevens–Johnson syndrome)
- Zidovudine
- Emtricitabine (palms and soles)

Hyperpigmentation
- Post-inflammatory (especially pruritic papular eruption)
- Zidovudine
- Emtricitabine (palms and soles)

*See Chapter 28 for more information.
lesions become more numerous and larger. Lymphoedema is common, as lymphatic vessels are infiltrated. KS also commonly spreads to lymph nodes and viscerally, especially to the lungs and gastrointestinal tract. Visceral disease occasionally occurs in the absence of mucocutaneous involvement. B symptoms of fever, night sweats and weight loss may occur.

KS may respond to ART. Chemotherapy should be reserved for those patients who fail to remit on ART, or given together with ART if there are poor prognostic features such as visceral involvement, oedema, ulcerated lesions and B symptoms.

**Bacillary angiomatosis**

Bacillary angiomatosis is a bacterial infection caused by *Bartonella henselae* or *B. quintana*. Skin lesions range from solitary superficial red-purple lesions resembling Kaposi’s sarcoma or pyogenic granuloma, to multiple subcutaneous nodules or plaques. Lesions are painful and may bleed or ulcerate. The infection may become disseminated with fevers, lymphadenopathy and hepatosplenomegaly. Diagnosis is made by biopsy of a lesion and Warthin–Starry silver staining, which reveals aggregates of bacilli. Treatment with doxycycline or azithromycin is effective.

**Papular pruritic eruption**

Papular pruritic eruption (‘itchy red bump disease’) is an intensely itchy, symmetrical rash affecting the trunk and extremities. It thought to be due to an allergic reaction to insect bites. In sub-Saharan Africa, it is the most common skin manifestation of HIV. Post-inflammatory hyperpigmentation is common. Topical steroids, emollients and antihistamines are useful but response is variable. Measures to reduce insect bites are logical, but difficult to implement in low-income settings.

**Drug rashes**

Cutaneous hypersensitivity to drugs is said to occur 100 times more frequently in HIV infection. The most common type is an erythematous maculo-papular rash, which may be scaly. The drugs most commonly associated with rashes are sulphonamides and non-nucleoside reverse transcriptase inhibitors (NNRTIs – see below). Severe, life-threatening features of drug rashes include blistering (when this affects more than 30% of surface area it is known as toxic epidermal necrolysis), involvement of mucous membranes (Stevens-Johnson syndrome, pp. 1264 and 1302), or systemic involvement with fever or organ dysfunction (especially hepatitis, which is often delayed for a few days after the rash develops). Because sulphonamides are important in the treatment and prophylaxis of opportunistic infections, rechallenge or desensitisation is often attempted in patients who have previously experienced rashes, provided the reaction was not life-threatening. Details of rashes caused by ART are given below.

**Oral conditions**

Oropharyngeal candidiasis is very common. It is nearly always caused by *C. albicans* (p. 381), but other azole-resistant *Candida* species may be selected for if there have been repeated courses of azole drugs. Pseudomembranous candidiasis is the most common manifestation, with white patches on the buccal mucosa (p. 388) that can be scraped off to reveal a red raw surface. Erythematous candidiasis is more difficult to diagnose and presents with a reddened mucosa and a smooth shiny tongue. Angular cheilitis due to *Candida* is a common manifestation. Topical antifungals are usually effective. Antifungal lozenges and gentian violet are both more effective than antifungal solutions. Systemic azole therapy, usually fluconazole, should be given if topical therapy fails or if there are oesophageal symptoms.

Oral hairy leucoplaikia (p. 388) appears as corrugated white plaques running vertically on the side of the tongue, and is virtually pathognomonic of HIV disease. It is usually asymptomatic and is due to EBV. If high doses of aciclovir or a related drug are given for varicella zoster virus infections, the lesions may temporarily regress.

Oral ulcers are common. Herpetiform oral ulcers occur in primary infection. Herpes simplex typically affects the nasolabial area, but may cause oral ulcers. In early disease, minor aphthous ulcers are common. In advanced disease, giant aphthous ulcers occur. These destroy tissue, are painful and need to be differentiated from herpes simplex and CMV ulcers by biopsy. They respond to systemic steroids. Histoplasmosis (p. 385) is an uncommon cause of oral ulcers, usually associated with constitutional symptoms. Finally, superficial oral ulcers may occur as part of the Stevens-Johnson syndrome, usually caused by sulphonamides or NNRTIs.

KS often involves the mouth, especially the hard palate (see above and Fig. 14.6). Nodular oral lesions are associated with a worse prognosis.

Gingivitis is very common. Good oral hygiene and regular dental checkups are important. Acute necrotising ulcerative gingivitis and periostis (p. 388) can result in loss of teeth; they should be treated with a course of metronidazole and a dental referral should be made.

**Nail disorders**

Fungal infections (onychomycosis, p. 1280) are very common and often involve multiple nails. Blue discoloration of nails is common and may be due to HIV or to the antiretroviral drug zidovudine.
Gastrointestinal disease

Oesophageal diseases

Oesophageal candidiasis (Fig. 14.7) is the most common cause of pain on swallowing (odynophagia), dysphagia and regurgitation. Concomitant oral candidiasis is present in about 70% of patients. Systemic azole therapy, e.g. fluconazole 200 mg daily for 14 days, is usually curative, but relapses are common. Patients whose oesophageal symptoms fail to respond to azoles should be investigated with oesophagoscopy. Major aphthous ulceration and CMV ulcers are the most likely causes and need to be differentiated by biopsy. Occasionally, herpes simplex oesophagitis or obstructive KS is responsible.

Diarrhoea

Chronic diarrhoea is a very common presenting problem in patients with advanced HIV, especially in areas where there is no access to safe water. It is a major cause of wasting. The differential diagnosis of diarrhoea depends on whether the presentation is with large- or small-bowel symptoms (see Fig. 14.4, p. 396). The presentation and aetiology of acute diarrhoea are similar to those in HIV-uninfected patients.

Large-bowel diarrhoea

Acute diarrhoea caused by the bacterial enteric pathogens Campylobacter, Shigella and Salmonella occurs more frequently than in HIV-uninfected people and the illness is more severe. Bacteraemia is much more common, notably due to non-typhoid Salmonella. Diarrhoea caused by Clostridium difficile should be considered if there has been prior exposure to antibiotics, as is often the case in patients with symptomatic HIV.

CMV colitis presents with chronic large-bowel symptoms and fever in patients with CD4 counts below 100 cells/mm³. On colonoscopy, ulcers are seen, mostly involving the left side of the colon. Biopsy of ulcers shows typical ‘owl’s-eye’ inclusion bodies.

Small-bowel diarrhoea

Chronic small-bowel diarrhoea may be due to HIV enteropathy, but this is a diagnosis of exclusion. It typically presents with chronic watery diarrhoea and wasting without fever. Infection with one of three unicellular organisms is responsible for most cases: cryptosporidiosis, microsporidiosis and isosporiasis (Box 14.11). All three are intracellular parasites that invade enterocytes. If the diagnosis is not made by stool microscopy on at least two specimens, a duodenal biopsy should be performed (Fig. 14.8). Electron microscopy is essential for speciation of microsporidia.

### 14.11 Common causes of chronic watery diarrhoea

<table>
<thead>
<tr>
<th>Organism</th>
<th>Cryptosporidiosis</th>
<th>Microsporidiosis</th>
<th>Isosporiasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
<td>Cryptosporidium parvum</td>
<td>Encephalitozoon bieneusi</td>
<td>Isospora belli</td>
</tr>
<tr>
<td>C. hominis</td>
<td>E. intestinalis, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animal host</td>
<td>Multiple</td>
<td>Multiple</td>
<td>No</td>
</tr>
<tr>
<td>Distribution</td>
<td>Global</td>
<td>Global</td>
<td>Tropics</td>
</tr>
<tr>
<td>Stool examination</td>
<td>Acid-fast stain</td>
<td>Trichrome stain</td>
<td>Acid-fast stain</td>
</tr>
<tr>
<td></td>
<td>PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific treatment</td>
<td>No established therapy</td>
<td>Albendazole (some species)</td>
<td>Co-trimoxazole</td>
</tr>
</tbody>
</table>

**Fig. 14.7** Oesophageal candidiasis. Endoscopy showing typical pseudomembranous candidiasis.

**Fig. 14.8** Cryptosporidiosis. Duodenal biopsy may be necessary to confirm cryptosporidiosis or microsporidiosis. Arrow indicates an œocyst.
14 HIV INFECTION AND AIDS

About 40% of patients with disseminated MAC infections have watery diarrhoea. Unlike in cryptosporidiosis, microsporidiosis and isosporiasis, fever is a prominent feature of MAC infection. Intestinal tuberculosis typically involves the ileocaecal area and may present with fever and diarrhoea, but the diarrhoea is seldom profuse.

Hepatobiliary disease

Chronic viral hepatitis

Hepatitis B and/or C (HBV and HCV) co-infection is common in HIV-infected people due to shared risk factors for transmission. The natural history of both HBV and HCV is altered by HIV co-infection. In the ART era, chronic liver disease from viral hepatitis has emerged as a major cause of morbidity and mortality. HBV and HCV are further described on pages 950 and 954.

Hepatitis B

HBV infection is common in several groups of people at risk of HIV infection: residents of low- and middle-income countries, injection drug users, haemophiliacs and MSM. HIV co-infection increases HBV viraemia, associated with less elevation of transaminase (presumably due to immune suppression), and increases the risk of liver fibrosis and hepatoma. Several NRTIs (lamivudine, emtricitabine and tenofovir) are also effective against HBV. HBV status should be checked at baseline in all HIV-infected patients. Treatment with anti-HBV drugs should be considered for all patients who have active HBV replication (HBeAg-positive or HBV DNA > 2000 U/mL) and/or evidence of inflammation or fibrosis on liver biopsy. This is best achieved by starting an ART regimen that includes tenofovir with either lamivudine or emtricitabine. Interferon is seldom used, but may be considered in patients with CD4 counts above 500 cells/mm³. A flare of hepatitis may be associated with improved immune function after starting ART or discontinuing antiretrovirals that have anti-HBV activity. There is an increased risk of antiretroviral hepatotoxicity.

Hepatitis C

HCV infection is extremely common in injection drug users and haemophiliacs. HIV co-infection increases HCV viraemia and increases the risk of liver fibrosis and hepatoma. Treatment for HCV should preferably be deferred in patients with low CD4 counts until the CD4 count has risen to 350 cells/mm³ or more. As with HBV, a flare of hepatitis may be associated with improved immune function after starting ART, and there is an increased risk of antiretroviral hepatotoxicity. Response to anti-HCV therapy is similar to that seen in HIV-uninfected people, but there is more toxicity and there are important drug–drug interactions between several antiretrovirals and both ribavirin and the newer HCV protease inhibitors.

HIV cholangiopathy

HIV cholangiopathy, a form of secondary sclerosing cholangitis (p. 965), is seen in patients with severe immune suppression. In some patients, co-existing intestinal infection with CMV, cryptosporidiosis or microsporidiosis is present, but it is uncertain if these organisms play an aetiological role. Papillary stenosis is common and is amenable to cautery via endoscopic retrograde cholangiopancreatography (ERCP), which provides symptomatic relief. Acalculous cholecystitis is a common complication of cholangiopathy. ART may improve the condition.

Respiratory disease

Pulmonary disease is very common and is the major reason for hospital admission. More than 90% of patients who are admitted for respiratory diseases will have either bacterial pneumonia, pulmonary tuberculosis or Pneumocystis jirovecii pneumonia (PJP). PJP is more common in high-income countries, while tuberculosis is more common in low- and middle-income countries. An approach to the differential diagnosis of all three conditions is given in Box 14.12.

Pneumocystis jirovecii pneumonia

The key presenting feature of PJP is progressive dyspnoea. Dry cough and fever are common. The chest X-ray typically shows a bilateral interstitial infiltrate spreading out from the hilar regions (Fig. 14.9), but may be normal initially. High-resolution CT scan is more sensitive than chest X-ray, usually showing typical ‘ground-glass’ interstitial infiltrates. Pneumatocoeles may occur and may rupture, resulting in a pneumothorax. The diagnosis is made with silver stains, PCR or immunofluorescence of broncho-alveolar lavage or induced sputum. Treatment is with high-dose co-trimoxazole, together with adjunctive steroids if the patient is hypoxic.

Pulmonary tuberculosis

Tuberculosis is the most common cause of admission in countries with a high tuberculosis incidence. Pulmonary tuberculosis in patients with mild immune suppression typically presents as in HIV-uninfected patients, with a chronic illness and apical pulmonary cavities (p. 688). However, in patients with CD4 counts below 200 cells/mm³, there are four important differences in the clinical presentation of pulmonary tuberculosis.
Sputum smears, which are positive in most HIV-uninfected adults with pulmonary tuberculosis, are negative in more than half of patients. The main reason for this is the absence of pulmonary cavities.

Many patients have disseminated tuberculosis, sometimes with a classic miliary pattern on chest X-ray, but more commonly presenting with pulmonary and extrapulmonary tuberculosis. The most common sites of concomitant extrapulmonary tuberculosis are the pleura and lymph nodes. Acid-fast bacilli are more often found on wide-needle aspirate of nodes than on sputum (p. 395). Pleural aspirate showing a lymphocytic exudate suggests tuberculosis as a likely cause and pleural biopsy will usually confirm the diagnosis.

Tuberculosis in HIV-infected patients responds well to standard short-course therapy (p. 693).

**Bacterial pneumonia**
The incidence of bacterial pneumonia is increased about 100-fold by HIV infection. The severity, likelihood of bacteraemia, risk of recurrent pneumonia, and mortality are all increased compared with HIV-uninfected patients. The aetiology is similar to that of community-acquired pneumonia in HIV-uninfected patients with co-morbidity: *S. pneumoniae* is the commonest, followed by *Haemophilus influenzae*, enterobacteriaceae (e.g. *Klebsiella pneumoniae*) and *Staphylococcus aureus*. The prevalence of atypical bacteria in HIV-infected patients with pneumonia is probably similar to that in the general population, but the data are limited. Treatment is with a broad-spectrum β-lactam (e.g. ceftriaxone, amoxicillin-clavulanate), with the addition of a macrolide if the pneumonia is severe.

Uncommon bacteria causing pneumonia include *Pseudomonas aeruginosa*, *Nocardia* (which mimics tuberculosis) and *Rhodococcus equi* (which can cause pulmonary cavities).

**Miscellaneous causes of pulmonary infiltrates**
Pulmonary cryptococcosis may present as a component of disseminated disease or be limited to the lungs. The chest X-ray appearances are variable. Cryptococcomas occur less commonly than in HIV-uninfected people. The commonest radiographic pattern seen in HIV infection is patchy consolidation, often with small areas of cavitation resembling tuberculosis. Pleural involvement is rare. The endemic mycoses (histoplasmosis, coccidioidomycosis and penicilliosis) often also cause non-specific pulmonary infiltrates.

Lymphoid interstitial pneumonitis is a slowly progressive disorder with a diffuse reticulonodular infiltrate. It is caused by a benign polyclonal lymphocytic interstitial infiltrate and is part of the diffuse infiltrative lymphocytosis syndrome (DILS - see p. 404). Patients may have other features of DILS, notably parotidomegaly.

Kaposi’s sarcoma often spreads to the lungs. Typical chest X-ray appearances are large, irregular nodules, linear reticular pattern and pleural effusions. Bronchoscopy is diagnostic.
HIV INFECTION AND AIDS

Nervous system and eye disease

The central and peripheral nervous systems are commonly involved in HIV, either as a direct consequence of HIV infection or due to opportunistic diseases. Presentations are outlined in Figure 14.11.

Cognitive impairment

HIV-associated neurocognitive disorders

HIV is a neurotropic virus and invades the CNS early during infection. Meningo-encephalitis may occur at seroconversion. About 50% of HIV-infected people have abnormal neuropsychiatric testing, the proportion increasing with declining CD4 counts. The term HIV-associated neurocognitive disorders (HAND) describes a spectrum of disorders: asymptomatic neurocognitive impairment (which is the most common), minor neurocognitive disorder and HIV-associated dementia (also called HIV encephalopathy). Dementia occurs in late disease and is a subcortical dementia characterised by impairment of executive function, psychomotor retardation and impaired memory. There is no diagnostic test for HIV-associated dementia. CT or magnetic resonance imaging (MRI) shows diffuse cerebral atrophy out of keeping with age. It is important to exclude depression, cryptococcal meningitis and neurosyphilis. ART usually improves HIV-associated dementia, but milder forms of HAND often persist.

Progressive multifocal leucoencephalopathy

Progressive multifocal leucoencephalopathy (PML) is a progressive disease that presents with stroke-like episodes and cognitive impairment. Vision is often impaired due to involvement of the occipital cortex. PML is caused by the JC virus. A combination of characteristic appearances on MRI (Fig. 14.12) and detection of JC virus DNA in the cerebrospinal fluid (CSF) by PCR is diagnostic. No specific treatment exists and prognosis remains poor despite ART.

CMV encephalitis

This presents with behavioural disturbance, cognitive impairment and a reduced level of consciousness. Focal signs may also occur. Detection of CMV DNA in the CSF supports the diagnosis. Response to anti-CMV therapy is poor.

Space-occupying lesions

Space-occupying lesions in AIDS patients typically present over days to weeks. The most common cause is toxoplasmosis. As toxoplasmosis responds rapidly to therapy, a trial of anti-toxoplasmosis therapy should be given to all patients presenting with space-occupying lesions while the results of diagnostic tests are being awaited.

Cerebral toxoplasmosis

Cerebral toxoplasmosis is caused by reactivation of residual Toxoplasma gondii cysts from past infection, which results in the development of space-occupying lesions. The characteristic findings on imaging are multiple space-occupying lesions with ring enhancement on contrast and surrounding oedema (Fig. 14.13). Toxoplasma serology shows evidence of previous exposure (positive immunoglobulin (Ig)G antibodies). The standard therapy for toxoplasmosis is sulfadiazine with pyrimethamine, together with folic acid, to reduce the risk of bone marrow suppression. However, co-trimoxazole has been shown to be as effective and less toxic, and is also more widely available. Response to a trial of therapy is usually diagnostic, with clinical improvement in 1–2 weeks and shrinkage of lesions on imaging in 2–4 weeks. Definitive diagnosis is by brain biopsy but this is seldom necessary.

Primary CNS lymphoma (PCNSL)

Primary CNS lymphomas (PCNSLs) are high-grade B-cell lymphomas associated with EBV infection. Characteristically, imaging demonstrates a single,
Presenting problems in HIV infection

The CSF may show features consistent with tuberculous meningitis. Response to antituberculosis therapy is slow and paradoxical expansion of lesions despite therapy is not uncommon.

**Stroke**

There is a higher incidence of stroke in patients with HIV disease. Atherosclerosis is accelerated by HIV-associated inflammation and by some antiretroviral drugs. HIV vasculopathy with occlusion can cause a stroke. The aetiology is thought to be a vasculitis. It is important to exclude tuberculous meningitis and meningovascular syphilis in all patients who present with a stroke.

**Meningitis**

**Cryptococcal meningitis**

*Cryptococcus neoformans* is the most common cause of meningitis in AIDS patients. Patients usually present subacutely with headache, vomiting and mild confusion. Neck stiffness is present in less than half. CSF pleocytosis is often mild or even absent, and protein and glucose concentrations are variable. It is important to request CSF cryptococcal antigen tests in all HIV-infected patients undergoing lumbar puncture, as this test has a sensitivity and specificity of almost 100%. Treatment is with amphotericin B (plus flucytosine if available) for 2 weeks, followed by fluconazole. Raised intracranial pressure is common and should be treated with repeated therapeutic lumbar punctures, removing sufficient CSF to reduce pressure to less than 20 cmH₂O. (Most experts would be reluctant to withdraw more than 30 mL at a time.)

**Tuberculous meningitis**

The presentation and CSF findings of tuberculous meningitis are similar to those in HIV-uninfected patients (p. 1201), except that concomitant tuberculosis at other sites is more common in HIV infection.

**Peripheral nerve disease**

HIV infection causes axonal degeneration, resulting in a sensorimotor peripheral neuropathy in about one-third of AIDS patients. The incidence is increased with lower CD4 counts, older age and increased height. Sensory symptoms predominate. Treatment involves foot care, analgesia and analgesic adjuvants. ART has minimal effect on halting or reversing the process. The NRTIs stavudine and didanosine can cause drug-induced peripheral neuropathy, which is typically more painful and more rapidly progressive than HIV neuropathy. It may remit if the offending drug is withdrawn early.

**Acute inflammatory demyelinating polyneuropathy** is an uncommon manifestation, usually occurring in primary infection. It resembles Guillain–Barré syndrome (p. 1224), except that CSF pleocytosis is more prominent. Mononeuritis may also occur, commonly involving the facial nerve.

**Myelopathy and radiculopathy**

Globally, the most common cause of myelopathy in HIV infection is cord compression from tuberculous spondylitis. Vacuolar myelopathy is seen in advanced disease and is due to HIV. It presents with a slowly progressive
paraparesis with no sensory level. MRI of the spine is normal, but is an important investigation to exclude other causes. Most patients have concomitant HIV-associated dementia. CMV polyradiculitis presents with painful legs, progressive flaccid paraparesis, saddle anaesthesia, absent reflexes and sphincter dysfunction. CSF shows a neutrophil pleocytosis (which is unusual for a viral infection), and the detection of CMV DNA by PCR confirms the diagnosis. Despite treatment with ganciclovir or valganciclovir, functional recovery is poor.

**Psychiatric disease**
Significant psychiatric morbidity is very common and is a major risk factor for poor adherence. Reactive depression is the most common disorder. Diagnosis is often difficult, as many patients have concomitant HAND. Substance misuse is common in many groups of people at risk of HIV. Some forms of ART can cause psychiatric adverse effects and these are detailed on page 409.

**Retinopathy**
CMV retinitis presents with painless, progressive visual loss in patients with severe immune suppression. On fundoscopy the vitreous is clear. Haemorrhages and exudates are seen in the retina (p. 388), often with sheathing of vessels ('frosted branch angiitis'). The disease usually starts unilaterally, but progressive bilateral involvement occurs in most untreated patients. Diagnosis is usually clinical, but if there is doubt, demonstrating CMV DNA by PCR of vitreous fluid is diagnostic. Treatment with ganciclovir or valganciclovir stops progression of the disease, but vision does not recover. Some patients may develop immune recovery uveitis in response to ART, with intraocular inflammation, macular oedema and cataract formation that requires prompt treatment with oral and intraocular corticosteroids to prevent visual loss.

Three other conditions may mimic CMV retinitis. Like CMV, they all occur in patients with CD4 counts below 100 cells/mm³. Ocular toxoplasmosis typically presents with a vitritis and retinitis without retinal haemorrhages. HIV retinopathy is a microangiopathy that causes cotton wool spots, which are not sight-threatening. Varicella zoster virus can cause rapidly progressive outer retinal necrosis.

**Rheumatological disease**
The immune dysregulation associated with HIV infection may result in autoantibody formation, usually in low titres. Mild arthralgias and a fibromyalgia-like syndrome are common in HIV-infected people.

**Arthritis**
HIV can cause a seronegative arthritis, which resembles rheumatoid arthritis. A more benign oligoarthritis may also occur. Reactive arthritis and Reiter’s syndrome are more severe in HIV infection (Ch. 25).

**Diffuse infiltrative lymphocytosis syndrome**
Diffuse infiltrative lymphocytosis syndrome (DILS) is a benign disorder involving polyclonal CD8 lymphocytic infiltration of tissues, which has some features in common with Sjögren’s syndrome (p. 1114). It is linked to human leucocyte antigen (HLA)-DRB1. Most patients have a CD8 lymphocytosis. DILS usually presents in patients with mild immune suppression. The most common manifestation is bilateral parotid gland enlargement; the glands are often massive, with lymphoepithelial cysts on histology (Fig. 14.15). Other salivary glands may also be enlarged. Sicca symptoms are common but usually mild. Lymphocytic interstitial pneumonitis is the most common manifestation outside the salivary glands. Generalised lymphadenopathy with nodes, larger than those seen with persistent generalised lymphadenopathy of HIV, may occur. Hepatitis, mononeuritis, polyarthritis and polymyositis may also occur. The manifestations outside the salivary glands usually respond to steroids. The parotid glands are treated for cosmetic reasons but surgery is best avoided. Aspiration of parotid cysts and instillation of a sclerosant are of some benefit. Low-dose irradiation has also been used successfully. DILS may regress on ART but response is variable.

**Haematological abnormalities**
Disorders of all three major cell lines may occur in HIV. In advanced disease, haematopoiesis is impaired due to the direct effect of HIV and by cytokines. Pancytopenia may occur as a consequence of HIV but it is important to exclude a disorder infiltrating the bone marrow, such as mycobacterial or fungal infections, or lymphoma.

**Anaemia**
Normochromic, normocytic anaemia is very common in advanced HIV disease. Opportunistic diseases may cause anaemia of chronic disease (e.g. tuberculosis) or marrow infiltration (e.g. MAC, tuberculosis, lymphoma, fungi). Anaemia is a common adverse effect of zidovudine, which also causes a macrocytosis. Red cell aplasia
is rare and may be caused either by parvovirus B19 infection or by lamivudine.

**Neutropenia**
Isolated neutropenia is occasionally due to HIV but is nearly always caused by drug toxicity (e.g. zidovudine, co-trimoxazole, ganciclovir).

**Thrombocytopenia**
Mild thrombocytopenia is common in HIV-infected people. The most common disorder causing severe thrombocytopenia is immune-mediated platelet destruction resembling idiopathic thrombocytopenic purpura (p. 1050). This responds to steroids or intravenous immunoglobulin, together with ART. Splenectomy should be avoided if possible because it further increases the risk of infection with encapsulated bacteria. Severe thrombocytopenia with a microangiopathic anaemia also occurs in a thrombotic thrombocytopenic purpura-like illness (p. 1056), which seems to have a better prognosis than the classical disease. Transient thrombocytopenia is common in primary infection.

**Renal disease**
Acute renal failure is common, usually due to acute infection or nephrotoxicity of drugs (e.g. tenofovir, p. 481), amphotericin B). HIV-associated nephropathy (HIVAN) is the most important cause of chronic renal failure and is seen most frequently in patients of African descent and those with low CD4 counts. Progression to end-stage renal failure is more rapid than with most other causes of chronic renal failure, and renal size may be preserved even when it is severe. HIVAN usually presents with nephrotic syndrome, chronic renal failure or a combination of both. ART has some effect in slowing progression of HIVAN. Other important HIV-associated renal diseases include HIV immune complex kidney diseases and thrombotic microangiopathy. With the overall improvement in life expectancy from ART, conditions such as diabetes mellitus, hypertension and vascular disease add to the burden of chronic kidney disease. Outcomes of renal transplantation are good in patients on ART.

**Cardiac disease**
HIV-associated cardiomyopathy resembles idiopathic dilated cardiomyopathy (p. 636) but progresses more rapidly. ART may improve cardiac failure but does not reverse established cardiomyopathy. Pericardial disease due to opportunistic diseases is not uncommon. Globally, the most common cause is tuberculous pericardial effusions. Tuberculous constrictive pericarditis is less common than in HIV-uninfected people. Kaposi’s sarcoma and lymphoma may cause pericardial effusions. Septic pericarditis, usually due to *S. pneumoniae*, is uncommon.

HIV is associated with an increased risk of myocardial infarction due to accelerated atherogenesis caused by the inflammatory state. Certain protease inhibitors (p. 407) that cause dyslipidaemia have been associated with an increased risk of myocardial infarction.

**HIV-related cancers**
The AIDS-defining cancers are Kaposi’s sarcoma (see above), cervical cancer and non-Hodgkin lymphoma (NHL – p. 1043). NHL may occur at any CD4 count but is more commonly seen below 200 cells/mm³. Almost all NHL are B-cell tumours and most are stage 4 when the patient presents. Long-term remission rates of about 50% can be achieved with NHL in AIDS patients using ART and chemotherapy (including the anti-B-cell monoclonal antibody rituximab if it is a B-cell tumour).

The incidence of a number of other cancers induced by viruses is also increased in HIV-infected people (Box 14.13). Regular cytological examination of the cervix, and of the anus in people who practise anal sex, should be performed to detect pre-malignant lesions, which are easier to treat. In general, the incidence of cancers that are not induced by viruses is similar to that in the general population.

**PREVENTION OF OPPORTUNISTIC INFECTIONS**
The best way to prevent opportunistic infections is to improve the CD4 count with ART. However, infections continue to occur in the ART era; CD4 counts take time to improve if ART is initiated in patients with profound immune suppression, immune reconstitution on ART is often suboptimal, and CD4 counts may decline because antiretroviral resistance develops.

**Preventing exposure**
The best method for avoiding infection is to prevent exposure to the infectious agent. However, this is only possible for a few opportunistic infections. The pathogenesis of several of these is thought to be reactivation of latent/dormant infection after prior exposure – examples include herpes simplex virus, zoster, CMV and toxoplasmosis. Preventing exposure to some of these infections is thus only of benefit if exposure has not already occurred.
HIV INFECTION AND AIDS

Safe water and food
Cryptosporidiosis, microsporidiosis and isosporiasis may be water-borne. If there is no access to safe water, then water should be boiled before drinking. Food-borne illnesses are also important in HIV infection, notably Salmonella species. Toxoplasma exposure is related to eating raw or undercooked meat. People living with HIV infection need to be informed about food hygiene and the importance of adequately cooked meat.

Tuberculosis
Preventing exposure to tuberculosis is important when there is an infectious case in the household, in clinics and in hospitals. Adequate ventilation, masks and safe coughing procedures reduce the risk of exposure.

Malaria vector control
All HIV-infected individuals living in malarious areas should practise vector control, as malaria occurs more frequently and is more severe in HIV-infected people. The most cost-effective way to achieve this is by using insecticide-impregnated bed nets. Other modalities of vector control that are of benefit to the community, such as reducing standing water and spraying with residual insecticides and larvicides, should also be implemented.

Safer sex
HIV-infected individuals should practise safer sex in order to reduce the transmission of HIV. Even if their partners are HIV-infected, condoms should be used, as HIV mutants that are more virulent or have developed antiretroviral drug resistance can be transmitted. Safer sex will also lower the risk of acquiring herpes simplex virus and human herpesvirus 8.

Pets
Toxoplasma gondii can be acquired from kittens or cat litter, and people living with HIV infection should avoid handling either. Cryptosporidiosis can be transmitted from animals, and patients should be advised to wash their hands after handling animals.

Chemoprophylaxis
Chemoprophylaxis is the use of antimicrobial agents to prevent infections. Primary prophylaxis is used to prevent opportunistic infections that have not yet occurred. Secondary prophylaxis is used to prevent recurrence of opportunistic infections because many may recur after an initial response to therapy. Secondary prophylaxis (Box 14.14) can be discontinued when ART results in immune reconstitution, with CD4 counts increasing to over 200 cells/mm³, but for CMV and MAC, prophylaxis can be stopped if CD4 counts increase to more than 100 cells/mm³.

Co-trimoxazole primary prophylaxis
Co-trimoxazole reduces the incidence of a number of opportunistic infections (Box 14.15), as well as hospitalisation and mortality rates. The indications for initiating co-trimoxazole are either clinical evidence of immune suppression (WHO clinical stages 3 or 4) or laboratory evidence of immune suppression (CD4 count below 200 cells/mm³). In low-income countries, there is considerable morbidity from infectious diseases (including malaria) in earlier HIV disease, and the WHO recommends initiating co-trimoxazole at a CD4 count of less than 350 cells/mm³, or at WHO stages 2–4. The recommended dose of co-trimoxazole is 960 mg daily, but trials have shown that half this dose is as effective and may be associated with less toxicity. Co-trimoxazole prophylaxis can be discontinued when CD4 counts increase to more than 200 cells/mm³ on ART.

Co-trimoxazole prophylaxis is well tolerated. The most common side-effect is hypersensitivity, causing a maculo-papular rash. If therapy is discontinued, desensitisation or rechallenge under antihistamine cover should be attempted, unless the rash was accompanied by systemic symptoms or mucusal involvement. Prophylactic doses of co-trimoxazole can also cause neutropenia, but this is very uncommon and routine monitoring of blood counts is not necessary. If co-trimoxazole cannot be tolerated, then dapsone 100 mg daily should be substituted. Dapsone is equally effective at reducing the incidence of P. jirovecii pneumonia, but has little or no effect on reducing the other opportunistic infections prevented by co-trimoxazole.

Isoniazid preventive therapy
Isoniazid preventive therapy (IPT) has been shown to reduce the risk of tuberculosis only in HIV-infected patients with a positive tuberculin skin test, which should be done in all patients at baseline. There is no CD4 count or clinical threshold for starting or stopping IPT. In HIV infection, induration of 5 mm or more on a Mantoux test is regarded as positive. It is important to rule out active tuberculosis before starting IPT, and symptom screening has been shown to be adequate to achieve this (Box 14.16). The usual duration of IPT is

<table>
<thead>
<tr>
<th>14.14 Secondary prophylaxis of opportunistic infections</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infection</strong></td>
</tr>
<tr>
<td>Pneumocystis jirovecii pneumonia</td>
</tr>
<tr>
<td>Toxoplasmosis</td>
</tr>
<tr>
<td>Cryptococcosis</td>
</tr>
<tr>
<td>Cytomegalovirus infection</td>
</tr>
<tr>
<td>Mycobacterium avium complex</td>
</tr>
<tr>
<td>Isospora belli infection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14.15 Opportunistic infections reduced by co-trimoxazole</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pneumocystis jirovecii pneumonia</td>
</tr>
<tr>
<td>• Cerebral toxoplasmosis</td>
</tr>
<tr>
<td>• Bacterial pneumonia</td>
</tr>
<tr>
<td>• Bacteraemia</td>
</tr>
<tr>
<td>• Isosporiasis</td>
</tr>
<tr>
<td>• Malaria</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infection Drug regimen</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumocystis jirovecii pneumonia Co-trimoxazole 960 mg daily</td>
<td></td>
</tr>
<tr>
<td>Toxoplasmosis Co-trimoxazole 960 mg daily</td>
<td></td>
</tr>
<tr>
<td>Cryptococcosis Fluconazole 200 mg daily</td>
<td></td>
</tr>
<tr>
<td>Cytomegalovirus infection Valganciclovir 900 mg daily</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium avium complex Clarithromycin 500 mg twice daily + Ethambutol 800 mg daily</td>
<td></td>
</tr>
<tr>
<td>Isospora belli infection Co-trimoxazole 960 mg daily</td>
<td></td>
</tr>
</tbody>
</table>
Antiretroviral therapy

ART that is capable of suppressing viral replication has been available since 1996. ART has transformed HIV from a progressive illness with a fatal outcome into a chronic manageable disease (Box 14.18).

The goals of ART are to:

• reduce the viral load to an undetectable level for as long as possible
• improve the CD4 count to over 200 cells/mm³ so that severe HIV-related disease is unlikely
• improve the quantity and quality of life without unacceptable drug toxicity
• reduce HIV transmission.

**EBM 14.16** Symptom screen for tuberculosis before isoniazid preventive therapy

All of the following must be absent:

• Active cough
• Weight loss
• Night sweats
• Fever

**EBM 14.17** Duration of isoniazid preventive therapy

‘Thirty-six months’ isoniazid prophylaxis was more effective for prevention of tuberculosis than 6-month prophylaxis in individuals with HIV infection, and chiefly benefited those who were tuberculin skin test-positive.’


**EBM 14.18** Life expectancy on ART

‘A mathematical model predicted that people with HIV who can access ART will have life expectancy shortened by only 7 or 10.5 years, depending on whether ART is started early (CD4 count 432) or late (CD4 count 140) respectively.’


**14.19** Commonly used antiretroviral drugs

<table>
<thead>
<tr>
<th>Classes</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleoside reverse transcriptase inhibitors (NRTIs)</td>
<td>Abacavir, emtricitabine, lamivudine, tenofovir, zidovudine</td>
</tr>
<tr>
<td>Non-nucleoside reverse transcriptase inhibitors (NNRTIs)</td>
<td>Efavirenz, nevirapine, etravirine</td>
</tr>
<tr>
<td>Protease inhibitors (PIs)</td>
<td>Atazanavir, darunavir, lopinavir</td>
</tr>
<tr>
<td>Integrase inhibitors</td>
<td>Raltegravir</td>
</tr>
<tr>
<td>Chemokine receptor inhibitor</td>
<td>Maraviroc</td>
</tr>
</tbody>
</table>

Immunisation

There are significant problems associated with vaccination in HIV infection. First, vaccination with live organisms is contraindicated in patients with severe immune suppression, as this may result in disease from the attenuated organisms. Second, immune responses to vaccination are impaired in HIV-infected patients. If the CD4 count is below 200 cells/mm³, then immune responses to immunisation are very poor. Therefore it is preferable to wait until the CD4 count has increased to more than 200 cells/mm³ on ART before immunisation is given. All patients should be given a conjugate pneumococcal vaccine (not the polysaccharide vaccine, which has been shown to be harmful) and annual influenza vaccination. Hepatitis B vaccination should be given to those who are not immune.

**ANTIRETROVIRAL THERAPY**

Many of the antiretroviral drugs which were used initially have been largely abandoned because of toxicity or poor efficacy. The drugs that are currently most commonly used are shown in Box 14.19, and their targets in the HIV life cycle in Figure 14.2 (p. 392).

**Selecting antiretroviral regimens**

The standard combination antiretroviral regimens are two NRTIs together with an NNRTI, protease inhibitor (PI) or integrase inhibitor. Dual NRTI combinations are usually emtricitabine or lamivudine (they are closely related and so are never combined) together with one of abacavir, tenofovir or zidovudine. It is possible to construct effective regimens without NRTIs if there is intolerance or resistance to the NRTIs. Currently used PIs should always be administered with ritonavir, which itself is a PI that is toxic in therapeutic doses. Low doses of ritonavir dramatically increase the concentrations and elimination half-lives of other PIs by inhibiting their metabolism by cytochrome P450. This increases drug exposure, thereby prolonging the PI’s half-life, allowing reduction in pill burden and dosing frequency, and so optimising adherence.

Most guidelines from high-income countries allow clinicians to choose a starting regimen of dual NRTIs combined with an NNRTI, or a PI or an integrase inhibitor, as these three regimens have similar efficacy. Subsequent ART regimen switches for virological failure are guided by the results of resistance testing (p. 408). For low- and middle-income countries, the WHO recommends a public health approach to using ART with standardised first-line (NNRTI plus dual NRTIs) and second-line (ritonavir-boosted PI plus dual NRTIs) regimens. NNRTIs are preferred by the WHO in first-line regimens, as they are cheaper than PIs and better tolerated. Furthermore, NNRTIs need to be given with two fully active NRTIs because they have a low genetic
barrier to resistance (see below), whereas PI-containing regimens are effective even when there are some mutations conferring resistance to the NRTIs. Therefore PIs in second-line regimens are preferable in settings where resistance testing is unavailable. The public health approach to using ART can be implemented by nurses and has been successfully implemented in resource-poor settings.

**Criteria for starting ART**

The 2010 WHO guidelines recommend starting ART in adults with either a CD4 count below 350 cells/mm³ or clinical stage 3–4. Other international guidelines are very similar, but criteria are updated regularly and it is likely that starting ART at higher CD4 counts will be recommended in the near future. HIV-infected partners in serodiscordant couples should commence ART irrespective of their CD4 count or clinical stage to reduce the risk of transmission to the uninfected partner (Box 14.20). Other categories of patients who should start ART earlier include those with chronic liver disease from viral hepatitis, non-AIDS malignancies, and conditions requiring long-term immunosuppression.

It is seldom necessary to start ART urgently. Several consultations are required to give patients insight into the need for life-long ART, to stress the importance of adherence, and to formulate a personal treatment plan. Disclosure of HIV status, joining support groups and using patient-nominated treatment supporters should be encouraged, as these have been shown to improve adherence. Management of depression and substance abuse is also important.

**Monitoring efficacy**

The most important measure of ART efficacy is the viral load. A baseline viral load should be measured prior to initiating treatment. The viral load should be repeated 4–8 weeks after starting a new ART regimen when the count should show at least a tenfold decrease. Thereafter it should be checked every 3–6 months. After 6 months of ART, the viral load should be suppressed, defined as below the limit of detection of the assay (usually less than 50 copies/mL), and this is achieved in 80–90% of patients. Failure of ART is defined by the viral load becoming detectable after suppression. (In most guidelines a threshold is used – typically more than 400 or more than 1000 copies/mL.) Adherence support should be enhanced if virological failure is detected, and the viral load repeated to confirm failure before switching to a new ART regimen.

CD4 counts are generally monitored every 3–6 months together with the viral load. Typically, the CD4 count increases rapidly in the first month, followed by a more gradual increase. In the first year, the count typically increases by 100–150 cells/mm³, and about 80 cells/mm³ per annum thereafter until the reference range is reached, provided the viral load is suppressed. If ART is stopped, the CD4 count rapidly falls to the baseline value before ART was commenced. For countries where viral load monitoring is unavailable, the WHO has defined immunological failure as a fall in CD4 count to baseline, or a 50% fall from peak on ART, or persistent count below 100 cells/mm³. However, CD4 responses are highly variable: in about 15–30% of patients the CD4 count does not increase despite virological suppression, and in a similar proportion of patients the CD4 response is good despite the presence of virological failure. Therefore it is not surprising that both unnecessary switches to second-line ART and continuation of failing first-line ART regimens (which will increase the number of resistance mutations) are common in settings where viral load monitoring is not available.

**Antiretroviral resistance**

Reverse transcription is error-prone, generating a large number of mutations. If the viral load is suppressed on ART, viral replication is suppressed and resistance mutations will not be selected. If ART is taken and there is ongoing replication, due to either resistant mutations or suboptimal adherence, mutations conferring resistance to antiretroviral drugs will be selected. Antiretroviral drugs differ in their ability to select for resistant mutations. Some drugs (e.g. emtricitabine, lamivudine, efavirenz, nevirapine) have a low genetic barrier to resistance, rapidly selecting for a single mutation conferring high-level resistance. PIs and some NRTIs (e.g. zidovudine) select for resistance mutations slowly, and multiple resistant mutations often need to accumulate before the drug’s efficacy is lost. Patients who develop antiretroviral resistance may transmit resistant virus to others and will eventually develop clinical failure.

Antiretroviral resistance is assessed by sequencing the relevant viral genes to detect mutations that are known to confer resistance. The resistant proviral DNA is archived in latent CD4 cells and will re-emerge rapidly on exposure to the antiretroviral. The patient must therefore be taking ART when the test is performed, as otherwise the wild-type virus will predominate and resistant mutations will not be detected. In regions where resistance testing is affordable, it is recommended at baseline (to detect primary resistance) and at every confirmed virological failure, in order to select the most appropriate antiretrovirals in a new regimen.

**ART complications**

**Immune reconstitution inflammatory syndrome**

Immune reconstitution inflammatory syndrome (IRIS) is a common (15–20%) early complication of ART, especially in patients who start it with CD4 counts below 50 cells/mm³. IRIS presents either with paradoxical deterioration of an existing opportunistic disease (including infections that are responding to appropriate therapy) or with the unmasking of a new infection. The clinical presentation of IRIS events is often characterised...
by an exaggerated immune response, with pronounced inflammatory features. For example, patients with CMV retinitis developing IRIS on ART develop a uveitis; and inflammatory haloes occur around KS lesions. Paradoxical tuberculosis IRIS events are common and it is important to exclude multidrug resistance, which could be responsible for the deterioration. IRIS is associated with a mortality of around 5%, but this is higher when it complicates CNS infections.

The management of IRIS is to continue ART and to ensure that the opportunistic disease is adequately treated. Symptomatic treatments are helpful. Corticosteroids are often used for more severe IRIS manifestations, but they should not be given to patients with KS, as this can result in rapid progression of KS lesions.

**Lipodystrophy**

Long-term use of ART may cause changes in body fat distribution. This can present either with fat accumulation (e.g. visceral fat, breast enlargement, ‘buffalo hump’) or with subcutaneous fat loss (‘lipoatrophy’ - *Fig. 14.16*) or with both fat loss and accumulation. The thymidine analogue NRTIs ( stavudine and, to a lesser extent, zidovudine) are associated with fat loss. Switching to the non-thymidine NRTIs abacavir or tenofovir will result in very gradual improvement of lipoatrophy.

![Image](image_url)

*Fig. 14.16 Fat loss complicating long-term use of the thymidine analogue NRTIs stavudine and zidovudine.*

Previously, PIs were thought to be the cause of fat accumulation. However, recent studies have shown that all classes of antiretrovirals are associated with fat gain to the same extent. Furthermore, longitudinal studies comparing HIV-uninfected people with HIV-infected people on long-term ART show that the extent and distribution of fat gain are similar. These data suggest that fat gain is a consequence of treating HIV rather than a side-effect of ART.

**Rashes**

These are common but must be differentiated from the other causes described above. The NRTI abacavir typically causes a systemic hypersensitivity reaction, which is HLA-associated. HLA-B*5701 has a 100% negative predictive value for abacavir hypersensitivity. HLA testing should be done before abacavir is given and the drug should not be prescribed for people who are HLA-B*5701-positive. This is rare in people of African descent. Rechallenge must never be attempted after abacavir hypersensitivity, as fatal reactions may occur.

Drug rashes are very common with NNRTIs. When the NNRTI rash is mild and not accompanied by systemic involvement, the suspected drug is often continued and antihistamines are administered. The rash usually resolves. If it worsens or if systemic features develop, the NNRTI should be discontinued.

**Other adverse effects**

The NNRTI efavirenz causes insomnia, agitation, euphoria or dysphoria in many patients, but tolerance to its neuropsychiatric effects develops in a few weeks in most. The NRTI zidovudine can cause anaemia and neutropenia, and tenofovir may cause nephrotoxicity. Pls are associated with dyslipidaemias and may increase the risk of myocardial infarction.

**ART in special situations**

**Pregnancy**

All pregnant women should routinely be recommended for HIV testing at an early stage in pregnancy, with rapid tests for those presenting in or just after labour. The CD4 count falls by about 25% during pregnancy due to haemodilution. The course of HIV disease progression is not altered by pregnancy. Pre-ART, the rate of mother-to-child transmission was 15–40%, with rates being influenced by several factors (see Box 14.3, p. 391).

ART has dramatically reduced the risk of mother-to-child transmission of HIV to less than 1%. All pregnant women who qualify for ART for their own health should start treatment at the beginning of the second trimester. If they have severe disease, they should start ART in the first trimester. Two ART strategies are currently used to prevent mother-to-child transmission for women who do not yet require ongoing ART: commence standard ART and discontinue at delivery or after weaning; or zidovudine monotherapy, usually started at 28 weeks, augmented with single-dose nevirapine at delivery. The latter approach is widely used in low- and middle-income countries. A randomised controlled trial is under way in order to assess which strategy is more effective.

Caesarean section is associated with a lower risk of mother-to-child transmission than vaginal delivery, but the mode of delivery does not affect transmission risk if the viral load is suppressed on ART.

HIV is also transmitted by breastfeeding. In high-income countries, exclusive formula feeding is generally recommended. However in resource-poor settings, formula feeding is associated with a risk of infant morbidity and mortality, which may negate the benefit of not transmitting HIV to the infant. Furthermore, providing antiretrovirals to the infant (usually nevirapine monotherapy) while they are breastfeeding has been shown to reduce the risk of transmission. Therefore breastfeeding is now encouraged in resource-poor settings. Infants should be exclusively breastfed for the first
HIV INFECTION AND AIDS

6 months, as mixed feeding (with formula or solids) is associated with a higher risk of transmission. Diagnosis of HIV in infancy requires the detection of HIV RNA by PCR as maternal antibodies to HIV, which persist for up to 15 months, will give a false-positive result on antibody assays. PCR should ideally be carried out within 6 weeks of birth to facilitate early ART initiation, which is recommended for all infants irrespective of their CD4 percentage or clinical stage. If the baby is breastfed, the PCR should be repeated 2 weeks after weaning.

Post-exposure prophylaxis

Post-exposure prophylaxis (PEP) is recommended when the risk is deemed to be significant after a careful risk assessment, in both occupational and non-occupational settings. The first dose should be given as soon as possible, preferably within 6–8 hours. There is no point in starting PEP after 72 hours. Most guidelines recommend dual NRTIs for low-risk exposures, with the addition of either a PI or efavirenz in high-risk exposures (see Boxes 14.2 and 14.3, p. 391). Tenofovir together with emtricitabine is the most widely used dual NRTI combination, as it is well tolerated and well studied in pre-exposure prophylaxis trials. PEP should not be given if the exposed person is HIV-infected. HIV antibody testing should be performed at 6, 12 and 24 weeks after exposure.

Prevention of HIV

An effective HIV vaccine remains elusive due to the extensive genetic diversity of HIV and the lack of a safe attenuated virus. Measures for the prevention of HIV transmission are shown in Box 14.22.

Further information

Websites with updated clinical guidelines

http://www.aidsinfo.nih.gov/guidelines/
http://www.bhiva.org/ClinicalGuidelines.aspx
http://www.who.int/hiv/pub/guidelines/en/
Sexually transmitted infections

Clinical examination in men 412
Clinical examination in women 413
Approach to patients with a suspected STI 414
Presenting problems in men 415
  Urethral discharge 415
  Genital itch and/or rash 415
  Genital ulceration 415
  Genital lumps 416
  Proctitis in men who have sex with men 417
Presenting problems in women 417
  Vaginal discharge 417
  Lower abdominal pain 418
  Genital ulceration 418
  Genital lumps 418
  Chronic vulval pain and/or itch 418

Prevention of STI 419
Sexually transmitted bacterial infections 419
  Syphilis 419
  Gonorrhoea 422
  Chlamydial infection 423
  Other sexually transmitted bacterial infections 423
Sexually transmitted viral infections 423
  Genital herpes simplex 423
  Human papillomavirus and anogenital warts 425
  Molluscum contagiosum 426
  Viral hepatitis 426
Sexually transmitted infection

HIV testing

It should always be standard practice to offer HIV testing as part of screening for STI (sexually transmitted infection) because the benefits of early diagnosis outweigh other considerations. Extensive pre-test counselling is not required in most instances, but it is important to establish efficient pathways for referral of patients at high risk for whom the clinician wishes specialist support, and for those diagnosed as HIV-positive.

**Observation**
- Mouth
- Eyes
- Joints
- Skin: Rash of secondary syphilis
- Scabies
- Manifestations of HIV infection (Ch. 14)

**Investigations for STIs in heterosexual males**
- First void urine (FVU)* is the specimen of choice for the combined nucleic acid amplification test (NAAT) for gonorrhoea and chlamydia
- Alternatively, for gonorrhoea, a urethral swab plated directly on a selective medium such as modified New York City (MNYC), or sent in an appropriate transport medium, can be cultured to allow for assessment of antimicrobial sensitivities
- Serological test for syphilis (STS), e.g. enzyme immunoassay (EIA) for antitreponemal immunoglobulin (Ig) G antibody
- Human immunodeficiency virus (HIV) test (see note)

*FVU,* and pharyngeal and rectal swabs for combined NAAT for gonorrhoea and chlamydia
- STS (repeat testing may be necessary in the event of negative test results in the first few weeks following exposure)
- Serological tests for hepatitis A/B (with a view to vaccination if seronegative)
- HIV test (see note)

*A urethral swab can be submitted if the patient is unable to pass urine.
CLINICAL EXAMINATION IN WOMEN

Observation
- Mouth
- Eyes
- Joints
- Skin:
  - Rash of secondary syphilis
  - Scabies
  - Manifestations of HIV infection (Ch. 14)

Inguinal glands
- Significant enlargement

Abdomen
- Abnormal masses or tenderness

Pubic area
- Pthirus pubis (crab louse)

Labia majora and minora
- Ulcers
- Vulvitis
- Warts

Vagina and cervix
- Abnormal discharge
- Warts
- Ulcers
- Inflammation
  - In women with lower abdominal pain, bimanual examination for adnexal tenderness (pelvic inflammatory disease)

Perineum and perianal skin
- Warts
- Ulcers
- Inflammation

Investigations for STIs in women
- Clinician-obtained cervical or vaginal swab, or self-taken vaginal swab for combined NAAT for gonorrhoea and chlamydia
- Alternatively, for gonorrhoea, cervical and urethral swabs plated directly on a selective medium such as MNYC, or sent in appropriate transport medium, can be cultured to allow for assessment of antimicrobial sensitivities
- Wet mount for microscopy or high vaginal swab (HVS) for culture of Trichomonas
- STS, e.g. EIA for antitreponemal IgG antibody
- HIV test (see note)

Management goals in suspected STI
- Relief of any symptoms
- Screening for treatable STI that may not be causing symptoms
- Tracing and treatment of sexual contacts who may also be infected
- Advice to reduce risk of infection in the future

Those at particular risk from STIs*
- Sex workers, male and female
- Clients of sex workers
- Men who have sex with men
- Injecting drug users (sex for money or drugs) and their partners
- Frequent travellers

*Adapted from WHO/UNAIDS, 1997.

Insets on pp. 412–413 (Perianal warts, coronal papillae, mucopus, inflammation) – see p. 426.
Sexually transmitted infections (STIs) are a group of contagious conditions whose principal mode of transmission is by intimate sexual activity involving the moist mucous membranes of the penis, vulva, vagina, cervix, anus, rectum, mouth and pharynx, along with their adjacent skin surfaces. A wide range of infections may be sexually transmitted, including syphilis, gonorrhoea, human immunodeficiency virus (HIV), genital herpes, genital warts, chlamydia and trichomoniasis. Bacterial vaginosis and genital candidiasis are not regarded as STIs, although they are common causes of vaginal discharge in sexually active women. Chancroid, lymphogranuloma venereum (LGV) and granuloma inguinale are usually seen in tropical countries. Hepatitis viruses A, B, C and D (p. 948) may be acquired sexually, as well as by other routes.

The World Health Organization estimates that 448 million curable STIs (Trichomonas vaginalis, Chlamydia trachomatis, gonorrhoea and syphilis) occur world-wide each year. In the UK in 2010, the most common treatable STIs diagnosed were chlamydia (more than 200,000 cases) and gonorrhoea (19,000 cases). Genital warts are the second most common complaint seen in genitourinary medicine (GUM) departments. In addition to causing morbidity themselves, STIs may increase the risk of transmitting or acquiring HIV infection (Ch. 14).

As coincident infection with more than one STI is seen frequently, GUM clinics routinely offer a full set of investigations at the patient’s first visit (pp. 412–413), regardless of the reason for attendance. In other settings, less comprehensive investigation may be appropriate.

The extent of the examination largely reflects the likelihood of HIV infection or syphilis. Most heterosexuals in the UK are at such low risk of these infections that less comprehensive examination is unnecessary. This is not the case in parts of the world where HIV is endemic, or for men who have sex with men (MSM) in the UK. In other words, the extent of the examination is determined by the sexual history (Box 15.1).

### Approach to Patients with a Suspected STI

Patients concerned about the possible acquisition of an STI are often anxious. Staff must be friendly, sympathetic and reassuring; they should have the ability to put patients at ease, whilst emphasising that clinic attendance is confidential. The history focuses on genital symptoms, with reference to genital ulceration, rash, irritation, pain, swelling and urinary symptoms, especially dysuria. In men, the clinician should ask about urethral discharge, and in women, vaginal discharge, pelvic pain or dyspareunia. Enquiry about general health should include menstrual and obstetric history, cervical cytology, recent medication, especially with antimicrobial or antiviral agents, previous STI and allergy. Immunisation status for hepatitis A and B should be noted, as should information about recreational drug use and alcohol intake.

A detailed sexual history is imperative (see Box 15.1), as this informs the clinician of the degree of risk for certain infections, as well as specific sites that should be sampled; for example, rectal samples should be taken from men who have had unprotected anal sex with other men. Sexual partners, whether male or female, and casual or regular, should be recorded. Sexual practices – insertive or receptive vaginal, anal, orogenital or oral – should be noted, as should information about contraceptive use for women, and condom use for both sexes.

### STI during pregnancy

Many STIs can be transmitted from mother to child in pregnancy, either transplacentally or during delivery. Possible outcomes are highlighted in Box 15.2.

### STI in children

The presence of an STI in a child may be indicative of sexual abuse, although vertical transmission may explain some presentations in the first 2 years. In an older child and in adolescents, STI may be the result of voluntary
sexual activity. Specific issues regarding the management of STI and other infections in adolescence are discussed in Box 13.27 (p. 313).

**PRESENTING PROBLEMS IN MEN**

**Urethral discharge**

In the UK the most important causes of urethral discharge are gonorrhoea and chlamydia. In a significant minority of cases, tests for both of these infections are negative, a scenario often referred to as non-specific urethritis (NSU). Some of these cases may be caused by *Trichomonas vaginalis*, herpes simplex virus (HSV), mycoplasmas or ureaplasmas. A small minority seem not to have an infectious aetiology.

Gonococcal urethritis usually causes symptoms within 7 days of exposure. The discharge is typically profuse and purulent. Chlamydial urethritis has an incubation period of 1–4 weeks, and tends to result in milder symptoms than gonorrhoea; there is overlap, however, and microbiological confirmation should always be sought.

**Investigations**

A presumptive diagnosis of urethritis can be made from a Gram-stained smear of the urethral exudate (Fig. 15.1), which will demonstrate significant numbers of polymorphonuclear leucocytes (≥ 5 per high-power field). A working diagnosis of gonococcal urethritis is made if Gram-negative intracellular diplococci (GNDC) are seen; if no GNDC are seen, a label of NSU is applied.

A Gram-stained urethral smear from a man with gonococcal urethritis. Gram-negative diplococci are seen within polymorphonuclear leucocytes.

If microscopy is not available, urine samples and/or swabs should be taken and empirical antimicrobials prescribed. A first-void urine (FVU) sample should be submitted for a combined nucleic acid amplification test (NAAT) for gonorrhoea and chlamydia; a urethral swab is an alternative if the patient cannot pass urine. When gonorrhoea is suspected, a urethral swab should be sent for culture and antimicrobial sensitivities of *Neisseria gonorrhoeae*. Tests for other potential causes of urethritis are not performed routinely.

A swab should also be taken from the pharynx because gonococcal infection here is not reliably eradicated by single-dose therapy. In MSM, swabs for gonorrhoea and chlamydia should be taken from the rectum.

**Management**

This depends on local epidemiology and the availability of diagnostic resources. Treatment is often presumptive, with prescription of multiple antimicrobials to cover the possibility of gonorrhoea and/or chlamydia. This is likely to include a single-dose treatment for gonorrhoea, which is desirable because it eliminates the risk of non-adherence. The recommended agents for treating gonorrhoea vary according to local antimicrobial resistance patterns (p. 422). Appropriate treatment for chlamydia (p. 423) should also be prescribed because concurrent infection is present in up to 50% of men with gonorrhoea. Non-gonococcal, non-chlamydial urethritis is treated as for chlamydia.

Patients should be advised to avoid sexual contact until it is confirmed that any infection has resolved and, whenever possible, recent sexual contacts should be traced. The task of contact tracing – also called partner notification – is best performed by trained nurses based in GUM clinics; it is standard practice in the UK to treat current sexual partners of men with gonococcal or non-specific urethritis without waiting for microbiological confirmation.

If symptoms clear, a routine test of cure is not necessary, but patients should be re-interviewed to confirm that there was no immediate vomiting or diarrhoea after treatment, that there has been no risk of re-infection, and that traceable partners have sought medical advice.

**Genital itch and/or rash**

Patients may present with many combinations of penile/genital symptoms, which may be acute or chronic, and infectious or non-infectious. Box 15.3 provides a guide to diagnosis.

Balanitis refers to inflammation of the glans penis, often extending to the under-surface of the prepuce, in which case it is called balanoposthitis. Tight prepuce and poor hygiene may be aggravating factors. Candidiasis is sometimes associated with immune deficiency, diabetes mellitus, and the use of broad-spectrum antimicrobials, corticosteroids or antimitic drugs. Local saline bathing is usually helpful, especially when no cause is found.

**Genital ulceration**

The most common cause of ulceration is genital herpes. Classically, multiple painful ulcers affect the glans, coronal sulcus or shaft of penis (Fig. 15.2), but solitary lesions occur rarely. Perianal ulcers may be seen in MSM. The diagnosis is made by gently scraping material from lesions and sending this in an appropriate transport medium for culture or detection of HSV DNA by polymerase chain reaction (PCR). Increasingly, laboratories will also test for *Treponema pallidum* by PCR.

In the UK, the possibility of syphilis or any other ulcerating STI is much less likely unless the patient is an MSM and/or has had a sexual partner from a region where tropical STIs are more common. The classic lesion
15.2 Penile herpes simplex (HSV-2) infection.

**Fig. 15.2** Penile herpes simplex (HSV-2) infection.

**TABLE 15.3 Differential diagnosis of genital itch and/or rash in men**

<table>
<thead>
<tr>
<th>Likely diagnosis</th>
<th>Acute or chronic</th>
<th>Itch</th>
<th>Pain</th>
<th>Discharge (non-urethral)</th>
<th>Specific characteristics</th>
<th>Diagnostic test</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subclinical urethritis</td>
<td>Either</td>
<td>±</td>
<td>−</td>
<td>±</td>
<td>Often intermittent</td>
<td>Gram stain and urethral swabs</td>
<td>As for urethral discharge</td>
</tr>
<tr>
<td>Candidiasis</td>
<td>Acute</td>
<td>✓</td>
<td>−</td>
<td>White</td>
<td>Postcoital</td>
<td>Microscopy</td>
<td>Antifungal cream, e.g. clotrimazole</td>
</tr>
<tr>
<td>Anaerobic (erosive) balanitis</td>
<td>Acute</td>
<td>±</td>
<td>−</td>
<td>Yellow</td>
<td>Offensive</td>
<td>Microscopy</td>
<td>Saline bathing ± metronidazole</td>
</tr>
<tr>
<td><em>Pthirus pubis</em> ('crab lice') infection</td>
<td>Either</td>
<td>✓</td>
<td>−</td>
<td>−</td>
<td>Lice and nits seen attached to pubic hairs</td>
<td>Can be by microscopy, but usually visual</td>
<td>According to local policy – often permethrin</td>
</tr>
<tr>
<td>Lichen planus</td>
<td>Either</td>
<td>±</td>
<td>−</td>
<td>−</td>
<td>Violaceous papules ± Wickham’s striae</td>
<td>Clinical</td>
<td>None or mild topical corticosteroid, e.g. hydrocortisone</td>
</tr>
<tr>
<td>Lichen sclerosus</td>
<td>Chronic</td>
<td>±</td>
<td>−</td>
<td>−</td>
<td>Ivory-white plaques, scarring</td>
<td>Clinical or biopsy</td>
<td>Strong topical corticosteroid, e.g. clobetasol</td>
</tr>
<tr>
<td>Plasma cell balanitis of Zoon</td>
<td>Chronic</td>
<td>✓</td>
<td>−</td>
<td>±</td>
<td>Shiny, inflamed circumscribed areas</td>
<td>Clinical or biopsy</td>
<td>Strong topical corticosteroid, e.g. clobetasol</td>
</tr>
<tr>
<td>Dermatoses, e.g. eczema or psoriasis</td>
<td>Either</td>
<td>✓</td>
<td>−</td>
<td>−</td>
<td>Similar to lesions elsewhere on skin</td>
<td>Clinical</td>
<td>Mild topical corticosteroid, e.g. hydrocortisone</td>
</tr>
<tr>
<td>Genital herpes</td>
<td>Acute</td>
<td>±</td>
<td>✓</td>
<td>−</td>
<td>Atypical ulcers are not uncommon</td>
<td>Swab for HSV PCR</td>
<td>Oral antiviral, e.g. aciclovir</td>
</tr>
<tr>
<td>Circinate balanitis</td>
<td>Either</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>Painless erosions with raised edges; usually as part of Reiter’s syndrome (p. 1107)</td>
<td>Clinical</td>
<td>Mild topical steroid, e.g. hydrocortisone</td>
</tr>
</tbody>
</table>

(HSV PSR = herpes simplex virus polymerase chain reaction)

The most common cause of genital ‘lumps’ is warts (p. 425). These are classically found in areas of friction during sex, such as the parafrenal skin and prepuce of the penis. Warts may also be seen in the urethral meatus, and less commonly on the shaft or around the base of the penis. Perianal warts are surprisingly common in men who do not have anal sex.
The differential diagnosis includes molluscum contagiosum and skin tags. Adolescent boys may confuse normal anatomical features such as coronal papillae (p. 412), parafrenal glands or sebaceous glands (Fordyce spots) with warts.

**Proctitis in men who have sex with men**

STIs that may cause proctitis in MSM include gonorrhoea, chlamydia, herpes and syphilis. The sub-strains of *Chlamydia trachomatis* that cause LGV (L1–3) have been associated with outbreaks of severe proctitis in Northern Europe, including the UK. Symptoms include mucopurulent anal discharge, rectal bleeding, pain and tenesmus.

Examination may show mucopus and erythema with contact bleeding (p. 412). In addition to the diagnostic tests on page 412, a PCR test for HSV and a request for identification of the LGV substrain should be arranged if chlamydial infection is detected. Treatment is directed at the individual infections (see below).

MSM may also present with gastrointestinal symptoms from infection with organisms such as *Entamoeba histolytica* (p. 367), *Shigella* spp. (p. 345), *Campylobacter* spp. (p. 342) and *Cryptosporidium* spp. (p. 369).

**PRESENTING PROBLEMS IN WOMEN**

**Vaginal discharge**

The natural vaginal discharge may vary considerably, especially under differing hormonal influences such as puberty, pregnancy or prescribed contraception. A sudden or recent change in discharge, especially if associated with alteration of colour and/or smell, or vulval itch/irritation, is more likely to indicate an infective cause than a gradual or long-standing change.

Local epidemiology is particularly important when assessing possible causes. In the UK, most cases of vaginal discharge are not sexually transmitted, being due to either candidal infection or bacterial vaginosis (BV). World-wide, the most common treatable STI causing vaginal discharge is trichomoniasis; other possibilities include gonorrhoea and chlamydia. HSV may cause increased discharge, although vulval pain and dysuria are usually the predominant symptoms. Non-infective causes include retained tampons, malignancy and/or fistulae.

Speculum examination often allows a relatively accurate diagnosis, with appropriate treatment to follow (Box 15.4). If the discharge is homogeneous and off-white in colour, vaginal pH is greater than 4.5, and Gram stain microscopy reveals scanty or absent lactobacilli with significant numbers of Gram-variable organisms, some of which may be coating vaginal squamous cells (so-called Clue cells, Fig. 15.3), the likely diagnosis is BV. If there is vulval and vaginal erythema, the discharge is curdy in nature, vaginal pH is less than 4.5, and Gram stain microscopy reveals fungal spores and pseudohyphae, the diagnosis is candidiasis. Trichomoniasis tends to cause a profuse yellow or green discharge and is usually associated with significant vulvovaginal inflammation. Diagnosis is made by observing motile

---

**15.4 Infections that cause vaginal discharge**

<table>
<thead>
<tr>
<th>Cause</th>
<th>Clinical features</th>
<th>Treatment (in pregnancy seek specialist advice)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Candidiasis</strong></td>
<td>Vulval and vaginal inflammation</td>
<td>Clotrimazole¹ 500 mg pessary once at night and clotrimazole cream twice daily or Econazolone³ pessary 150 mg for 3 nights and econazole cream twice daily (topical creams for 7 days) or Fluconazole⁵ 150 mg orally stat</td>
</tr>
<tr>
<td></td>
<td>Curdy white discharge adherent to walls of vagina</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low vaginal pH</td>
<td></td>
</tr>
<tr>
<td><strong>Trichomoniasis</strong></td>
<td>Vulval and vaginal inflammation</td>
<td>Metronidazole⁴ 400 mg twice daily orally for 5–7 days or Metronidazole⁴ 2 g orally as a single dose</td>
</tr>
<tr>
<td></td>
<td>Frothy yellow/green discharge</td>
<td></td>
</tr>
<tr>
<td><strong>Bacterial vaginosis</strong></td>
<td>No inflammation</td>
<td>Metronidazole⁴ 2 g stat or 400 mg twice daily orally for 5–7 days</td>
</tr>
<tr>
<td></td>
<td>White homogeneous discharge</td>
<td>Metronidazole⁴ vaginal gel 0.75% daily for 5 days</td>
</tr>
<tr>
<td></td>
<td>High vaginal pH</td>
<td>Clindamycin⁶ vaginal cream 2% daily for 7 days</td>
</tr>
<tr>
<td><strong>Streptococcal/staphylococcal infection</strong></td>
<td>Purulent vaginal discharge</td>
<td>Choice of antibiotic depends on sensitivity tests</td>
</tr>
</tbody>
</table>

¹Clotrimazole, econazole and clindamycin damage latex condoms and diaphragms. ²Avoid in pregnancy and breastfeeding. ³Avoid alcoholic drinks until 48 hours after finishing treatment. Avoid high-dose regimens in pregnancy or breastfeeding. ⁴Pseudomembranous colitis has been reported with the use of clindamycin cream.
flagellate protozoa on a wet-mount microscopy slide of vaginal material.

If examination reveals the discharge to be cervical in origin, the possibility of chlamydial or gonococcal infection is increased and appropriate cervical or vaginal swabs should be taken (p. 413). In addition, Gram stain of cervical and urethral material may reveal GNDC, allowing presumptive treatment for gonorrhoea to be given. If gonococcal cervicitis is suspected, swabs should also be taken from the pharynx and rectum; infections at these sites are not reliably eradicated by single-dose therapy and a test of cure will therefore be required.

GUM clinics in the UK may offer sexually active women presenting with vaginal discharge an STI screen (p. 413). In other settings, such as primary care or gynaecology, testing for chlamydia and gonorrhoea may be considered in young women (< 25 years old), those who have changed partner recently, and those not using a barrier method of contraception, even if a non-STI cause of discharge is suspected clinically.

Treatment of infections causing vaginal discharge is shown in Box 15.4.

### Lower abdominal pain

Pelvic inflammatory disease (PID, infection or inflammation of the Fallopian tubes and surrounding structures) is part of the extensive differential diagnosis of lower abdominal pain in women, especially those who are sexually active. The possibility of PID is increased if, in addition to acute/subacute pain, there is dyspareunia, abnormal vaginal discharge and/or bleeding. There may also be systemic features, such as fever and malaise. On examination, lower abdominal pain is usually bilateral, and vaginal examination reveals adnexal tenderness with or without cervical excitation. Unfortunately, a definitive diagnosis can only be made by laparoscopy. A pregnancy test should be performed (as well as the diagnostic tests on p. 413) because the differential diagnosis includes ectopic pregnancy.

Broad-spectrum antibiotics, including those active against gonorrhoea and chlamydia, such as ofloxacin and metronidazole, should be prescribed if PID is suspected, along with appropriate analgesia. Delaying treatment increases the likelihood of adverse sequelae, such as abscess formation, and tubal scarring that may lead to ectopic pregnancy or infertility. Hospital admission may be indicated for severe symptoms.

### Genital ulceration

The most common cause of ulceration is genital herpes. Classically, multiple painful ulcers affect the introitus, labia and perineum, but solitary lesions occur rarely. Inguinal lymphadenopathy and systemic features, such as fever and malaise, are more common than in men. Diagnosis is made by gently scraping material from lesions and sending this in an appropriate transport medium for culture or detection of HSV DNA by PCR. Increasingly, laboratories will also test such samples for *Treponema pallidum* by PCR. In the UK, the possibility of any other ulcerating STI is unlikely unless the patient has had a sexual partner from a region where tropical STIs are more common (see Box 15.12, p. 424).

Inflammatory causes include lichen sclerosus, Stevens–Johnson syndrome (pp. 1264 and 1302), Behçet’s syndrome (p. 1119) and fixed drug reactions. In older patients, malignant and pre-malignant conditions, such as squamous cell carcinoma, should be considered.

### Genital lumps

The most common cause of genital ‘lumps’ is warts. These are classically found in areas of friction during sex, such as the fourchette and perineum. Perianal warts are surprisingly common in women who do not have anal sex.

The differential diagnosis includes molluscum contagiosum, skin tags and normal papillae or sebaceous glands.

### Chronic vulval pain and/or itch

Women may present with a range of chronic symptoms that may be intermittent or continuous (Box 15.5).

Recurrent candidiasis may lead to hypersensitivity to candidal antigens, with itch and erythema becoming more prominent than increased discharge. Effective
Sexually transmitted bacterial infections

Acquired syphilis

Early syphilis

Primary syphilis

The incubation period is usually between 14 and 28 days, with a range of 9–90 days. The primary lesion or chancre (Fig. 15.4) develops at the site of infection, usually in the genital area. A dull red macule develops, becomes papular and then erodes to form an indurated ulcer (chancre). The draining inguinal lymph nodes may become moderately enlarged, mobile, discrete and rubbery. The chancre and the lymph nodes are both painless and non-tender, unless there is concurrent or secondary infection. Without treatment, the chancre will resolve within 2–6 weeks to leave a thin atrophic scar.

Chancres may develop on the vaginal wall and on the cervix. Extragenital chancres are found in about 10% of patients, affecting sites such as the finger, lip, tongue, tonsil, nipple, anus or rectum. Anal chancres often resemble fissures and may be painful.

Secondary syphilis

This occurs 6–8 weeks after the development of the chancre, when treponemes disseminate to produce a multisystem disease. Constitutional features, such as mild fever, malaise and headache, are common. Over 75% of patients present with a rash on the trunk and limbs that may later involve the palms and soles; this is initially macular but evolves to maculopapular or papular forms, which are generalised, symmetrical and non-irritable. Scales may form on the papules later. The rash affects the trunk and proximal limbs, and characteristically involves the palms, soles and face. Lesions

PREVENTION OF STI

Case-finding

Early diagnosis and treatment facilitated by active case-finding will help to reduce the spread of infection by limiting the period of infectivity; tracing and treating sexual partners will also reduce the risk of re-infection. Unfortunately, the majority of individuals with an STI are asymptomatic and therefore unlikely to seek medical attention. Improving access to diagnosis in primary care or non-medical settings, especially through opportunistic testing, may help. However, the impact of medical intervention through improved access alone is likely to be small.

Changing behaviour

The prevalence of STIs is driven largely by sexual behaviour. Primary prevention encompasses efforts to delay the onset of sexual activity and limit the number of sexual partners thereafter. Encouraging the use of barrier methods of contraception will also help to reduce the risk of transmitting or acquiring STIs. This is especially important in the setting of ‘sexual concurrency’, where sexual relationships overlap.

Unfortunately, there is contradictory evidence as to which (if any) interventions can reduce sexual activity. Knowledge alone does not translate into behaviour change, and broader issues, such as poor parental role modelling, low self-esteem, peer group pressure in the context of the increased sexualisation of our societies, gender power imbalance and homophobia, all need to be addressed. Throughout the world there is a critical need to enable women to protect themselves from undisciplined and coercive male sexual activity. Economic collapse and the turmoil of war regularly lead to situations where women are raped or must turn to prostitution to feed themselves and their children, and an inability to negotiate safe sex increases their risk of acquiring STI, including HIV.

SEXUALLY TRANSMITTED BACTERIAL INFECTIONS

Syphilis

Syphilis is caused by infection, through abrasions in the skin or mucous membranes, with the spirochaete Treponema pallidum. In adults the infection is usually sexually acquired; however, transmission by kissing, blood transfusion and percutaneous injury has been reported. Transplacental infection of the fetus can occur.

The natural history of untreated syphilis is variable. Infection may remain latent throughout, or clinical features may develop at any time. The classification of syphilis is shown in Box 15.6. All infected patients should be treated. Penicillin remains the drug of choice for all stages of infection.

15.6 Classification of syphilis

<table>
<thead>
<tr>
<th>Stage</th>
<th>Acquired</th>
<th>Congenital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>Primary</td>
<td>Clinical and latent</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Latent</td>
<td></td>
</tr>
<tr>
<td>Late</td>
<td>Latent</td>
<td>Clinical and latent</td>
</tr>
<tr>
<td></td>
<td>Benign tertiary</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cardiovascular</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neurosyphilis</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 15.4 Primary syphilis. A painless ulcer (chancre) is shown in the coronal sulcus of the penis. This is usually associated with inguinal lymphadenopathy.
Condylomata lata (papules coalescing to plaques) may develop in warm, moist sites such as the vulva or perianal area. Generalised non-tender lymphadenopathy is present in over 50% of patients. Mucosal lesions, known as mucous patches, may affect the genitalia, mouth, pharynx or larynx and are essentially modified papules, which become eroded. Rarely, confluence produces characteristic ‘snail track ulcers’ in the mouth.

Other features, such as meningitis, cranial nerve palsies, anterior or posterior uveitis, hepatitis, gastritis, glomerulonephritis or periostitis, are sometimes seen. Neurological involvement may be more common in HIV-positive patients.

The differential diagnosis of secondary syphilis can be extensive, but in the context of a suspected STI, primary HIV infection is the most important alternative condition to consider (see Ch. 14). Non-STI conditions that mimic the rash include psoriasis, pityriasis rosea, scabies, allergic drug reaction, erythema multiforme and tinea versicolor.

The clinical manifestations of secondary syphilis will resolve without treatment but relapse may occur, usually within the first year of infection. Thereafter, the disease enters the phase of latency.

**Latent syphilis**

This phase is characterised by the presence of positive syphilis serology or the diagnostic cerebrospinal fluid (CSF) abnormalities of neurosyphilis in an untreated patient with no evidence of clinical disease. It is divided into early latency (within 2 years of infection), when syphilis may be transmitted sexually, and late latency, when the patient is no longer sexually infectious. Transmission of syphilis from a pregnant woman to her fetus, and rarely by blood transfusion, is possible for several years following infection.

**Late syphilis**

This may persist for many years or for life. Without treatment, over 60% of patients might be expected to suffer little or no ill health. Coincidental prescription of antibiotics for other illnesses, such as respiratory tract or skin infections, may treat latent syphilis serendipitously.

**Benign tertiary syphilis**

This may develop between 3 and 10 years after infection but is now rarely seen in the UK. Skin, mucous membranes, bone, muscle or viscera can be involved. The characteristic feature is a chronic granulomatous lesion called a gumma, which may be single or multiple. Healing with scar formation may impair the function of the structure affected. Skin lesions may take the form of nodules or ulcers, whilst subcutaneous lesions may ulcerate with a gummy discharge. Healing occurs slowly, with the formation of characteristic tissue papule scars. Mucosal lesions may occur in the mouth, pharynx, larynx or nasal septum, appearing as punched-out ulcers. Of particular importance is gummatous involvement of the tongue, healing of which may lead to leukoplakia with the attendant risk of malignant change. Gummata of the tibia, skull, clavicle and sternum have been described, as has involvement of the brain, spinal cord, liver, testis and, rarely, other organs. Resolution of active disease should follow treatment, though some tissue damage may be permanent. Paroxysmal cold haemoglobinuria (p. 1031) may be seen.

**Cardiovascular syphilis**

This may present many years after initial infection. Aortitis, which may involve the aortic valve and/or the coronary ostia, is the key feature. Clinical features include aortic incompetence, angina and aortic aneurysm (p. 603). The condition typically affects the ascending aorta and sometimes the aortic arch; aneurysm of the descending aorta is rare. Treatment with penicillin will not correct anatomical damage and surgical intervention may be required.

**Neurosyphilis**

This may also take years to develop. Asymptomatic infection is associated with CSF abnormalities in the absence of clinical signs. Meningovascular disease, tabes dorsalis and general paralysis of the insane constitute the symptomatic forms (p. 1209). Neurosyphilis and cardiovascular syphilis may coexist and are sometimes referred to as quaternary syphilis.

**Congenital syphilis**

Congenital syphilis is rare where antenatal serological screening is practised. Antisyphilitic treatment in pregnancy: treats the fetus, if infected, as well as the mother. Treponemal infection may give rise to a variety of outcomes after 4 months of gestation, when the fetus becomes immunocompetent:

- miscarriage or stillbirth, prematurely or at term
- birth of a syphilitic baby (a very sick baby with hepatosplenomegaly, bullous rash and perhaps pneumonia)
- birth of a baby who develops signs of early congenital syphilis during the first few weeks of life (Box 15.7)
- birth of a baby with latent infection who either remains well or develops congenital syphilis/stigma later in life (see Box 15.7).

**Investigations in adult cases**

*Treponema pallidum* may be identified in serum collected from chancre, or from moist or eroded lesions in secondary syphilis using a dark-field microscope, a direct fluorescent antibody test or PCR.

The serological tests for syphilis are listed in Box 15.8. Many centres use treponemal ELAs for IgG and IgM antibodies to screen for syphilis. ELA for antitreponemal IgM becomes positive at approximately 2 weeks, whilst non-treponemal tests become positive about 4 weeks after primary syphilis. All positive results in asymptomatic patients must be confirmed by repeat tests.

Biological false positive reactions occur occasionally; these are most commonly seen with Venereal Diseases Research Laboratory (VDRL) or rapid plasma reagin (RPR) tests (when treponemal tests will be negative). Acute false-positive reactions may be associated with infections, such as infectious mononucleosis, chickenpox and malaria, and may also occur in pregnancy. Chronic false-positive reactions may be associated with autoimmune diseases. False-negative results for nontreponemal tests may be found in secondary syphilis
Phenomenon). The visualisation of the flocculation reaction (the prozone formation of the antibody–antigen lattice necessary for matic neurological disease may coexist. The CSF should indistinguishable from (non-venereal) syphilis (bejel) and pinta (pp. 332–333), diogram are useful in the investigation of cardiovascular occasional necessary.

Investigations in suspected congenital syphilis

Passively transferred maternal antibodies from an adequately treated mother may give rise to positive serological tests in her baby. In this situation, non-treponemal tests should become negative within 3–6 months of birth. A positive EIA test for antitreponemal IgM suggests early congenital syphilis. A diagnosis of congenital syphilis mandates investigation of the mother, her partner and any siblings.

Management

Penicillin is the drug of choice. Currently, a single dose of 2.4 megunits of intramuscular benzathine penicillin is recommended for early syphilis (< 2 years’ duration), with three doses at weekly intervals being recommended in late syphilis. Doxycycline is indicated for patients allergic to penicillin, except in pregnancy (see below). Azithromycin is a further alternative. All patients must be followed up to ensure cure, and partner notification is of particular importance. Resolution of clinical signs in early syphilis with declining titres for non-treponemal tests, usually to undetectable levels within 6 months for primary syphilis and 12–18 months for secondary syphilis, is an indicator of successful treatment. Specific treponemal antibody tests may remain positive for life. In patients who have had syphilis for many years there may be little serological response following treatment.

Pregnancy

Penicillin is the treatment of choice in pregnancy. Erythromycin stearate can be given if there is penicillin hypersensitivity, but crosses the placenta poorly; the newborn baby must therefore be treated with a course of penicillin and consideration given to treating the mother. Some specialists recommend penicillin desensitisation for pregnant mothers so that penicillin can be given during temporary tolerance. The author has successfully prescribed ceftriaxone 250 mg IM for 10 days in this situation. Babies should be treated in hospital with the help of a paediatrician.

Treatment reactions

• Anaphylaxis. Penicillin is a common cause; on-site facilities should be available for management (p. 91).

• Jarisch–Herxheimer reaction. This is an acute febrile reaction that follows treatment and is characterised by headache, malaise and myalgia; it resolves within 24 hours. It is common in early syphilis and rare in late syphilis. Fetal distress or premature labour can occur in pregnancy. The reaction may also cause worsening of neurological (cerebral artery occlusion) or ophthalmic (uveitis, optic neuritis) disease, myocardial ischaemia (inflammation of the coronary ostia) and laryngeal stenosis (swelling of a gumma). Prednisolone
10–20 mg orally three times daily for 3 days is recommended to prevent the reaction in patients with these forms of the disease; antisyphilitic treatment can be started 24 hours after introducing corticosteroids. In high-risk situations it is wise to initiate therapy in hospital.

• **Procaine reaction.** Fear of impending death occurs immediately after the accidental intravenous injection of procaine penicillin and may be associated with hallucinations or fits. Symptoms are short-lived, but verbal assurance and sometimes physical restraint are needed. The reaction can be prevented by aspiration before intramuscular injection to ensure the needle is not in a blood vessel.

### Gonorrhoea

Gonorrhoea is caused by infection with *Neisseria gonorrhoeae* and may involve columnar epithelium in the lower genital tract, rectum, pharynx and eyes. Transmission is usually the result of vaginal, anal or oral sex. Gonococcal conjunctivitis may be caused by accidental infection from contaminated fingers. Untreated mothers may infect babies during delivery, resulting in ophthalmia neonatorum (Fig. 15.5). Infection of children beyond the neonatal period usually indicates sexual abuse.

**Clinical features**

The incubation period is usually 2–10 days. In men the anterior urethra is commonly infected, causing urethral discharge and dysuria, but symptoms are absent in about 10% of cases. Examination will usually show a mucopurulent or purulent urethral discharge. Rectal infection in MSM is usually asymptomatic but may present with anal discomfort, discharge or rectal bleeding. Proctoscopy may reveal either no abnormality, or clinical evidence of proctitis (see p. 417) such as inflamed rectal mucosa and mucopus.

In women, the urethra, paraurethral glands/ducts, Bartholin’s glands/ducts or endocervical canal may be infected. The rectum may also be involved either due to contamination from a urogenital site or as a result of anal sex. Occasionally, the rectum is the only site infected.

About 80% of women who have gonorrhoea are asymptomatic. There may be vaginal discharge or dysuria but these symptoms are often due to additional infections, such as chlamydia (see below), trichomoniasis or candidiasis, making full investigation essential (p. 413). Lower abdominal pain, dyspareunia and intermenstrual bleeding may be indicative of PID. Clinical examination may show no abnormality, or pus may be expressed from urethra, paraurethral ducts or Bartholin’s ducts. The cervix may be inflamed, with mucopurulent discharge and contact bleeding.

Pharyngeal gonorrhoea is the result of receptive oro-genital sex and is usually symptomless. Gonococcal conjunctivitis is an uncommon complication, presenting with purulent discharge from the eye(s), severe inflammation of the conjunctivae and oedema of the eyelids, pain and photophobia. Gonococcal ophthalmia neonatorum presents similarly with purulent conjunctivitis and oedema of the eyelids. Conjunctivitis must be treated urgently to prevent corneal damage.

Disseminated gonococcal infection (DGI) is seen rarely, and typically affects women with asymptomatic genital infection. Symptoms include arthritis of one or more joints, pustular skin lesions and fever. Gonococcal endocarditis has been described.

### Investigations

Gram-negative diplococci may be seen on microscopy of smears from infected sites (see Fig. 15.1, p. 415). Pharyngeal smears are difficult to analyse due to the presence of other diplococci, so the diagnosis must be confirmed by culture or NAAT.

### Management of adults

Emerging resistance is making it increasingly difficult to cure gonorrhoea with a single oral dose of

---

**Fig. 15.5 Gonococcal ophthalmia neonatorum.** From McMillan and Scott 2000 – see p. 426.
Chlamydial infection

Chlamydial infection in men

Chlamydia is transmitted and presents in a similar way to gonorrhoea; however, urethral symptoms are usually milder and may be absent in over 50% of cases. Conjunctivitis is also milder than in gonorrhoea; pharyngitis does not occur. The incubation period varies from 1 week to a few months. Without treatment, symptoms may resolve but the patient remains infectious for several months. Complications, such as epididymo-orchitis and Reiter’s syndrome, or sexually acquired reactive arthropathy (SARA, p. 1107), are rare. Sexually transmitted pathogens, such as chlamydia or gonococci, are usually responsible for epididymo-orchitis in men aged less than 35 years, whereas bacteria such as Gram-negative enteric organisms are more commonly implicated in older men.

Treatments for chlamydia are listed in Box 15.11. NSU is treated identically. The partner(s) of men with chlamydia should be treated, even if laboratory tests for chlamydia are negative. Investigation is not mandatory but serves a useful epidemiological purpose; moreover, positive results encourage further attempts at contact-tracing.

Chlamydial infection in women

The cervix and urethra are commonly involved. Infection is asymptomatic in about 80% of patients but may cause vaginal discharge, dysuria and intermenstrual and/or postcoital bleeding. Lower abdominal pain and dyspareunia are features of PID. Examination may reveal mucopurulent cervicitis, contact bleeding from the cervix, evidence of PID or no obvious clinical signs. Treatment options are listed in Box 15.11. The patient’s male partner(s) should be investigated and treated.

Some infections may clear spontaneously but others persist. PID, with the risk of tubal damage and subsequent infertility or ectopic pregnancy, is a rare but important long-term complication. Other complications include perihepatitis, chronic pelvic pain, conjunctivitis and Reiter’s syndrome or SARA. Perinatal transmission may lead to ophthalmia neonatorum and/or pneumonia in the neonate.

Other sexually transmitted bacterial infections

Chancroid, granuloma inguinale and LGV as causes of genital ulcers in the tropics are described in Box 15.12. LGV is also a cause of proctitis in MSM (p. 417).

Genital herpes simplex

Infection with herpes simplex virus type 1 (HSV-1) or type 2 (HSV-2) produces a wide spectrum of clinical problems (p. 325), and may facilitate HIV transmission. Infection is usually acquired sexually (vaginal, anal, orogenital or oro-anal), but perinatal transmission to the neonate may also occur. Primary infection at the site of HSV entry, which may be symptomatic or asymptomatic, establishes latency in local sensory ganglia. Recurrences, either symptomatic or asymptomatic viral shedding, are a consequence of HSV reactivation. The first symptomatic episode is usually the most severe. Although HSV-1 is classically associated with orolabial herpes and HSV-2 with anogenital herpes, HSV-1 now accounts for more than 50% of anogenital infections in the UK.

Clinical features

The first symptomatic episode presents with irritable vesicles that soon rupture to form small, tender ulcers on the external genitalia (see Figs 15.6 and 15.2, p. 416). Lesions at other sites (e.g. urethra, vagina, cervix, perianal area, anus or rectum) may cause dysuria, urethral or vaginal discharge, or anal, perianal or rectal pain. Constitutional symptoms, such as fever, headache and malaise, are common. Inguinal lymph nodes become enlarged and tender, and there may be nerve root pain in the 2nd and 3rd sacral dermatomes.

Extragenital lesions may develop at other sites, such as the buttock, finger or eye, due to auto-inoculation. Oropharyngeal infection may result from orogenital sex. Complications, such as urinary retention due to autonomic neuropathy, and aseptic meningitis, are occasionally seen.

First episodes usually heal within 2–4 weeks without treatment; recurrences are usually milder and of shorter duration than the initial attack. They occur more often
in HSV-2 infection and their frequency tends to decrease with time. Prodromal symptoms, such as irritation or burning at the subsequent site of recurrence, or neuralgic pains affecting buttocks, legs or hips, are commonly seen. The first symptomatic episode may be a recurrence of a previously undiagnosed primary infection. Recurrent episodes of asymptomatic viral shedding are important in the transmission of HSV.

**Diagnosis**

Swabs are taken from vesicular fluid or ulcers for detection of DNA by PCR, or tissue culture and typing as either HSV-1 or 2. Electron microscopy of such material will only give a presumptive diagnosis, as herpes group viruses appear similar. Type-specific antibody tests are available but are not sufficiently accurate for general use.

**Management**

**First episode**

The following 5-day oral regimens are all recommended and should be started within 5 days of the beginning of the episode, or whilst lesions are still forming:

- aciclovir 200 mg five times daily
- famciclovir 250 mg three times daily
- valaciclovir 500 mg twice daily.

Analgesia may be required and saline bathing can be soothing. Treatment may be continued for longer than 5 days if new lesions develop. Occasionally, intravenous

<table>
<thead>
<tr>
<th>Infection and distribution</th>
<th>Organism</th>
<th>Incubation period</th>
<th>Genital lesion</th>
<th>Lymph nodes</th>
<th>Diagnosis</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphogranuloma venereum (LGV)</td>
<td>Chlamydia trachomatis types L1, 2, 3</td>
<td>3–30 days</td>
<td>Small, transient, painless ulcer, vesicle, papule; often unnoticed</td>
<td>Tender, usually unilateral, matted, suppurative bubo; inguinal/femoral nodes involved¹</td>
<td>Serological tests for L1–3 serotypes; swab from ulcer or bubo pus for Chlamydia</td>
<td>Doxycycline³ twice daily orally for 21 days or Erythromycin 500 mg four times daily orally</td>
</tr>
<tr>
<td>Chancroid</td>
<td>Haemophilus ducreyi (short Gram-negative bacillus)</td>
<td>3–10 days</td>
<td>Single or multiple painful ulcers with ragged undermined edges</td>
<td>As above but unilocular, suppurative bubo; inguinal nodes involved in ~50%</td>
<td>Microscopy and culture of scrapings from ulcer or pus from bubo</td>
<td>Azithromycin¹ 1 g orally once or Ceftriaxone 250 mg IM once or Ciprofloxacin² 500 mg twice daily orally for 3 days</td>
</tr>
<tr>
<td>Granuloma inguinale</td>
<td>Klebsiella granulomatis (Donovan bodies)</td>
<td>3–40 days</td>
<td>Ulcers or hypertrophic granulomatous lesions; usually painless⁴</td>
<td>Initial swelling of inguinal nodes, then spread of infection to form abscess or ulceration through adjacent skin</td>
<td>Microscopy of cellular material for intracellular bipolar-staining Donovan bodies</td>
<td>Azithromycin¹ 1 g weekly orally or 500 mg daily orally or Doxycycline¹ 100 mg twice daily orally or Ceftriaxone 1 g IM daily</td>
</tr>
</tbody>
</table>

¹The genito-ano-rectal syndrome is a late manifestation of LGV. ²Doxycycline and ciprofloxacin are contraindicated in pregnancy and breastfeeding. ³The safety of azithromycin in pregnancy and breastfeeding has not been fully assessed. ⁴Mother-to-baby transmission of granuloma inguinale may rarely occur.

**N.B.** Partners of patients with LGV, chancroid and granuloma inguinale should be investigated and treated, even if asymptomatic.
therapy may be indicated if oral therapy is poorly tolerated or aseptic meningitis occurs.

Catheterisation via the suprapubic route is advisable for urinary retention due to autonomic neuropathy because the transurethral route may introduce HSV into the bladder.

Recurrent genital herpes
Symptomatic recurrences are usually mild and may require no specific treatment other than saline bathing. For more severe episodes, patient-initiated treatment at onset, with one of the following 5-day oral regimens, should reduce the duration of the recurrence:
- aciclovir 200 mg five times daily
- famciclovir 125–250 mg twice daily
- valaciclovir 500 mg twice daily.

In a few patients, treatment started at the onset of prodromal symptoms may abort recurrence.

Suppressive therapy may be required for patients with frequent recurrences, especially if these occur at intervals of less than 4 weeks. Treatment should be given for a minimum of 1 year before stopping to assess recurrence rate. About 20% of patients will experience reduced attack rates thereafter, but for those whose recurrences remain unchanged, resumption of suppressive therapy is justified. Aciclovir 400 mg twice daily is most commonly prescribed.

Management in pregnancy
If her partner is known to be infected with HSV, a pregnant woman with no previous anogenital herpes should be advised to protect herself during sexual intercourse because the risk of disseminated infection is increased in pregnancy. Consistent condom use during pregnancy may reduce transmission of HSV. Genital herpes acquired during the first or second trimester of pregnancy is treated with aciclovir as clinically indicated. Although aciclovir is not licensed for use in pregnancy in the UK, there is considerable clinical evidence to support its safety. Third-trimester acquisition of infection has been associated with life-threatening haematogenous dissemination and should be treated with aciclovir.

Vaginal delivery should be routine in women who are symptomless in late pregnancy. Caesarean section is sometimes considered if there is a recurrence at the beginning of labour, although the risk of neonatal herpes through vaginal transmission is very low. Caesarean section is often recommended if primary infection occurs after 34 weeks because the risk of viral shedding is very high in labour.

**Human papillomavirus and anogenital warts**

Human papillomavirus (HPV) DNA typing has demonstrated over 90 genotypes (p. 1278), of which HPV-6, HPV-11, HPV-16 and HPV-18 most commonly infect the genital tract through sexual transmission. It is important to differentiate between the benign genotypes (HPV-6 and 11) that cause anogenital warts, and genotypes such as 16 and 18 that are associated with dysplastic conditions and cancers of the genital tract but are not a cause of benign warts. All genotypes usually result in subclinical infection of the genital tract rather than clinically obvious lesions affecting penis, vulva, vagina, cervix, perineum or anus.

**Clinical features**
Anogenital warts caused by HPV may be single or multiple, exophytic, papular or flat. Perianal warts (p. 412), whilst being more commonly found in MSM, are also found in heterosexual men and in women. Rarely, a giant condyloma (Buschke–Löwenstein tumour) develops with local tissue destruction. Atypical warts should be biopsied. In pregnancy, warts may dramatically increase in size and number, making treatment difficult. Rarely, they are large enough to obstruct labour and, in this case, delivery by Caesarean section will be required. Rarely, perinatal transmission of HPV leads to anogenital warts, or possibly laryngeal papillomas, in the neonate.

**Management**
The use of condoms can help prevent the transmission of HPV to non-infected partners, but HPV may affect parts of the genital area not protected by condoms. Vaccination against HPV infection has been introduced and is in routine use in several countries. There are two types of vaccine:
- A bivalent vaccine (Cervarix®) offers protection against HPV types 16 and 18, which account for approximately 75% of cervical cancers in the UK.
- A quadrivalent vaccine (Gardasil®) offers additional protection against HPV types 6 and 11, which account for over 90% of genital warts.

Both types of vaccine have been shown to be highly effective in the prevention of cervical intra-epithelial neoplasia in young women, and the quadrivalent vaccine has also been shown to be highly effective in protecting against HPV-associated genital warts (Box 15.13). It is currently recommended that HPV vaccination should be administered prior to the onset of sexual activity, typically at age 11–13, in a course of three injections. In the UK, only girls are offered vaccination, and it should be noted that this approach will not protect HPV transmission for MSM. As neither vaccine protects against all oncogenic types of HPV, cervical screening programmes will still be necessary.

A variety of treatments are available for established disease, including the following:
- Podophyllotoxin, 0.5% solution or 0.15% cream (contraindicated in pregnancy), applied twice daily for 3 days, followed by 4 days’ rest, for up to 4 weeks is suitable for home treatment of external warts.

**EBM 15.13 HPV vaccination and precancerous cervical intra-epithelial neoplasia**

‘Prophylactic HPV vaccination in women aged 15–25 years is highly effective at preventing precancerous cervical intra-epithelial neoplasia in young women who have not previously been infected with HPV.’

Sexually transmitted infections

- **Imiquimod** cream (contraindicated in pregnancy), applied 3 times weekly (and washed off after 6–10 hours) for up to 16 weeks, is also suitable for home treatment of external warts.
- **Cryotherapy** using liquid nitrogen to freeze warty tissue is suitable for external and internal warts but often requires repeated clinic visits.
- **Hyfrecation** – electrofulguration that causes superficial charring – is suitable for external and internal warts. Hyfrecation results in smoke plume which contains HPV DNA and has the potential to cause respiratory infection in the operator/patient. Masks should be worn during the procedure and adequate extraction of fumes should be provided.
- **Surgical removal**. Refractory warts, especially pedunculated perianal lesions, may be excised under local or general anaesthesia.

### Molluscum contagiosum

Infection by molluscum contagiosum virus, both sexual and non-sexual, produces flesh-coloured umbilicated hemispherical papules usually up to 5 mm in diameter after an incubation period of 3–12 weeks (Fig. 15.7). Larger lesions may be seen in HIV infection (p. 388). Lesions are often multiple and, once established in an individual, may spread by auto-inoculation. They are found on the genitalia, lower abdomen and upper thighs when sexually acquired. Facial lesions are highly suggestive of underlying HIV infection. Diagnosis is made on clinical grounds and by expression of the central core, in which the typical pox-like viral particles can be seen on electron microscopy (distinguishing molluscum contagiosum from genital warts). Typically, lesions persist for an average of 2 years before spontaneous resolution occurs. Treatment regimens are therefore cosmetic; they include cryotherapy, hyfrecation, topical applications of 0.15% podophyllotoxin cream (contraindicated in pregnancy) or expression of the central core.

---

**Viral hepatitis**

The hepatitis viruses A–D (p. 948) may be sexually transmitted:

- **Hepatitis A (HAV)**. Insertive oro-anal sex, insertive digital sex, insertive anal sex and multiple sexual partners have been linked with HAV transmission in MSM. HAV transmission in heterosexual men and women is also possible through oro-anal sex.
- **Hepatitis B (HBV)**. Insertive oro-anal sex, anal sex and multiple sexual partners are linked with HBV infection in MSM. Heterosexual transmission of HBV is well documented and commercial sex workers are at particular risk. Hepatitis D (HDV) may also be sexually transmitted.
- **Hepatitis C (HCV)**. Sexual transmission of HCV is well documented in MSM, but less so in heterosexuals. Sexual transmission is less efficient than for HBV.

The sexual partner(s) of patients with HAV and HBV should be seen as soon as possible and offered immunisation where appropriate. Patients with HAV should abstain from all forms of unprotected sex until non-infectious. Those with HBV should likewise abstain from unprotected sex until they are non-infectious or until their partners have been vaccinated successfully. No active or passive immunisation is available for protection against HCV but the consistent use of condoms is likely to protect susceptible partners. Active immunisation against HAV and HBV should be offered to susceptible people at risk of infection. Many STI clinics offer HAV immunisation to MSM along with routine HBV immunisation; a combined HAV and HBV vaccine is available.

---

**Further information and acknowledgements**

**Books and journal articles**


**Website**

www.bashh.org/guidelines Updates on treatment of all STIs.

**Figure acknowledgements**


Page 412 insets (Coronal papillae, mucopus), page 413 inset (Inflammation), Fig. 15.3 McMillan A, Young H, Ogilvie MM, Scott GR. Clinical practice in sexually transmissible infections. Edinburgh: Saunders; 2002; copyright Elsevier.
Clinical biochemistry and metabolism

Biochemical investigations 428
Integrated water and electrolyte balance 429
Water and electrolyte distribution 429
Investigation of water and electrolytes 429
Disorders of sodium balance 430
Functional anatomy and physiology of renal sodium handling 430
Presenting problems in disorders of sodium balance 432
Sodium depletion 432
Sodium excess 434
Diuretic therapy 434
Disorders of water balance 436
Functional anatomy and physiology of renal water handling 436
Presenting problems in disorders of water balance 437
Hyponatraemia 437
Hypernatraemia 439
Disorders of potassium balance 439
Functional anatomy and physiology of renal potassium handling 440
Presenting problems in disorders of potassium balance 440
Hypokalaemia 440
Hyperkalaemia 442

Disorders of acid–base balance 443
Functional anatomy and physiology of acid–base homeostasis 443
Presenting problems in disorders of acid–base balance 444
Metabolic acidosis 445
Metabolic alkalosis 446
Respiratory acidosis 447
Respiratory alkalosis 447
Mixed acid-base disorders 447

Disorders of divalent ion metabolism 447
Functional anatomy and physiology of magnesium metabolism 447
Presenting problems in disorders of magnesium metabolism 448
Hypomagnesaemia 448
Hypermagnesaemia 448
Functional anatomy and physiology of phosphate metabolism 448
Presenting problems in disorders of phosphate metabolism 448
Hypophosphataemia 448
Hyperphosphataemia 449

Disorders of amino acid metabolism 449
Disorders of carbohydrate metabolism 449
Disorders of complex lipid metabolism 450
Disorders of blood lipids and lipoproteins 450
Functional anatomy, physiology and investigation of lipid metabolism 450
Presenting problems in disorders of lipids 453
Disorders of haem synthesis – the porphyrias 458
There is a worldwide trend towards increased use of laboratory-based diagnostic investigations, and biochemical investigations in particular. In the health-care systems of developed countries, it has been estimated that 60–70% of all critical decisions taken in regard to patients, and over 90% of data stored in electronic medical records systems, involve a laboratory service or result.

This chapter covers a diverse group of disorders affecting adults not considered elsewhere in this book, whose primary manifestation is in abnormalities of biochemistry laboratory results, or whose underlying pathophysiology involves disturbance in specific biochemical pathways.

**BIOCHEMICAL INVESTIGATIONS**

There are three broad reasons why a clinician may request a biochemical laboratory investigation:

- to screen an asymptomatic subject for the presence of disease
- to assist in diagnosis of a patient’s presenting complaint
- to monitor changes in test results, as a marker of disease progression or response to treatment.

Contemporary medical practice has become increasingly reliant on laboratory investigation, and in particular, on biochemical investigation. This has been associated with extraordinary improvements in the analytical capacity and speed of laboratory instrumentation and the following operational trends:

- Large central biochemistry laboratories feature extensive use of automation and information technology. Specimens are transported from clinical areas to the laboratory using high-speed transport systems (such as pneumatic tubes) and identified with machine-readable labels (such as bar codes). Laboratory instruments have been miniaturised and integrated with robot transport systems to enable multiple rapid analyses of a single sample. Statistical process control techniques are used to assure the quality of analytical results, and increasingly to monitor other aspects of the laboratory, such as the time taken to complete the analysis (‘turn-around time’).
- **Point-of-care testing (POCT)** brings selected laboratory analytical systems into clinical areas, to the patient’s bedside or even connected to an individual patient. These systems allow the clinician to receive results almost instantaneously for immediate treatment of the patient, although often with less precision or at greater cost than using a central laboratory.
- The diversity of analyses has widened considerably with the introduction of many techniques borrowed from the chemical or other industries (Box 16.1).

Good medical practice involves the appropriate ordering of laboratory investigations and correct

<table>
<thead>
<tr>
<th>Analytical modality</th>
<th>Analyte</th>
<th>Typical applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion-selective electrodes</td>
<td>Blood gases, electrolytes (e.g. Na, K, Cl)</td>
<td>Point-of-care testing (POCT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-throughput analysers</td>
</tr>
<tr>
<td>Colorimetric chemical reaction or coupled enzymatic reaction</td>
<td>Simple mass or concentration measurement (e.g. creatinine, phosphate)</td>
<td>High-throughput analysers</td>
</tr>
<tr>
<td>Ligand assay (usually immunoassay)</td>
<td>Specific proteins</td>
<td>Increasingly available for POCT or high-throughput analysers</td>
</tr>
<tr>
<td>Ligand assay (usually immunoassay)</td>
<td>Hormones</td>
<td></td>
</tr>
<tr>
<td>Ligand assay (usually immunoassay)</td>
<td>Drugs</td>
<td></td>
</tr>
<tr>
<td>Chromatography: gas chromatography (GC), high-pressure liquid chromatography (HPLC), thin-layer chromatography (TLC)</td>
<td>Organic compounds</td>
<td>Therapeutic drug monitoring (TDM)</td>
</tr>
<tr>
<td>Mass spectroscopy (MS)</td>
<td>Haemoglobin derivatives</td>
<td>Drug screening (e.g. drugs of misuse)</td>
</tr>
<tr>
<td>Mass spectroscopy (MS)</td>
<td>Specific proteins</td>
<td>Vitamins</td>
</tr>
<tr>
<td>Mass spectroscopy (MS)</td>
<td>Immunoglobulins</td>
<td>Biochemical metabolites</td>
</tr>
<tr>
<td>Spectrophotometry, turbidimetry, nephelometry, fluorimetry</td>
<td>Xanthochromia</td>
<td></td>
</tr>
<tr>
<td>Spectrophotometry, turbidimetry, nephelometry, fluorimetry</td>
<td>Lipoproteins</td>
<td></td>
</tr>
<tr>
<td>Spectrophotometry, turbidimetry, nephelometry, fluorimetry</td>
<td>Paraproteins</td>
<td></td>
</tr>
<tr>
<td>Electrophoresis</td>
<td>Proteins</td>
<td>Paraproteins</td>
</tr>
<tr>
<td>Electrophoresis</td>
<td>Some enzymes</td>
<td>Isoenzyme analysis</td>
</tr>
<tr>
<td>Atomic absorption (AA) Inductively coupled plasma/mass spectroscopy (ICP-MS)</td>
<td>Trace elements and metals</td>
<td>Quantitation of heavy metals</td>
</tr>
<tr>
<td>Molecular diagnostics</td>
<td>Nucleic acid quantification and/or sequence</td>
<td>Inherited and somatic cell mutations (Ch. 3)</td>
</tr>
<tr>
<td>Molecular diagnostics</td>
<td></td>
<td>Genetic polymorphisms (Ch. 3)</td>
</tr>
<tr>
<td>Molecular diagnostics</td>
<td></td>
<td>Variations in rates of drug metabolism (Ch. 2)</td>
</tr>
<tr>
<td>Molecular diagnostics</td>
<td></td>
<td>Microbial diagnosis (Ch. 6)</td>
</tr>
</tbody>
</table>
interpretation of test results. The key principles, including the concepts of sensitivity and specificity, are described on page 5. Reference ranges for laboratory results are provided in Chapter 29. Many laboratory investigations can be subject to variability arising from incorrect patient preparation (for example, in the fasting or fed state), timing of sample collection (for example, in relation to diurnal variation of hormone levels, or dosage regimens for therapeutic agents), analytical factors (for example, serum versus plasma; use of the correct anticoagulant, or POCT versus central analysis) or artefact (for example, taking a venous sample proximal to the site of an intravenous infusion). It is therefore important for clinical and laboratory staff to communicate effectively and for clinicians to follow local recommendations concerning collection and transport of samples in the appropriate container and with appropriate labelling.

INTEGRATED WATER AND ELECTROLYTE BALANCE

One of the most common uses of the clinical biochemistry laboratory is to monitor electrolyte and acid–base status. The diverse clinical consequences of these biochemical disorders are illustrated in Box 16.2. Some whole-body electrolyte disturbances (notably of sodium) result in major clinical problems with minimal disturbance in measured biochemical parameters. However, these will also be considered for convenience here.

Before considering individual electrolytes and acid–base balance, it is important to review the relationships between them.

Water and electrolyte distribution

The following basic concepts are relevant to understanding the origin, consequences and therapy of many of the fluid and electrolyte disturbances discussed.

In a typical adult male, total body water (TBW) is approximately 60% of body weight (somewhat more for infants and less for women). For an average individual, TBW is about 40 L. Approximately 25 L is located inside cells (the intracellular fluid or ICF), while the remaining 15 L is in the extracellular fluid (ECF) compartment (Fig. 16.1). Most of the ECF (approximately 12 L) is interstitial fluid, which is within the tissues but outside cells, whereas the remainder (about 3 L) is in the plasma compartment.

Figure 16.1 illustrates some of the major differences in composition between the main body fluid compartments. The dominant cation in the ICF is potassium, while the dominant cation in the ECF is sodium. Phosphates and negatively charged proteins constitute the major intracellular anions, while chloride and, to a lesser extent, bicarbonate dominate the ECF anions. An important difference between the plasma and interstitial compartments of the ECF is that only plasma contains significant concentrations of protein.

The major force maintaining the difference in cation concentration between the ICF and ECF is the sodium-potassium pump (Na,K-activated ATPase), which is present in all cell membranes. Maintenance of the cation gradients across cell membranes is essential for many cell processes, including the excitability of conducting tissues such as nerve and muscle. The difference in protein content between the plasma and the interstitial fluid compartment is maintained by the impermeability of the capillary wall to protein. This protein concentration gradient (the colloid osmotic, or oncotic, pressure of the plasma) contributes to the balance of forces across the capillary wall that favour fluid retention within the plasma compartment.

Investigation of water and electrolytes

The most common biochemical test in plasma is called the urea and electrolytes (U&E) test in some parts of the world, and the electrolytes/urea/creatinine (EUC) in others. A guide to its interpretation is shown in Box 16.3. Because the blood consists of both intracellular (red cell) and extracellular (plasma) components, it is important...
Clinic al biochemistry and metabolism

The glomerulus, generating a fluid that is free from cells and protein and which resembles plasma in its electrolyte composition. This is then delivered into the renal tubules, where reabsorption of water and various electrolytes occurs (more detail on the structure and function of the glomerulus is given in Ch. 17). The glomerular filtration rate (GFR) is approximately 125 mL/min (equivalent to 180 L/day) in a typical adult. Over 99% of this filtered fluid is reabsorbed into the blood in the peritubular capillaries during its passage through successive segments of the nephron, largely as a result of tubular reabsorption of sodium. The processes mediating this sodium reabsorption, and the factors that regulate it, are key to understanding clinical disturbances and pharmacological interventions.

Nephron segments

At least four different functional segments of the nephron can be defined in terms of their mechanism for sodium reabsorption (Fig. 16.3).

Proximal tubule

This is responsible for the reabsorption of some 65% of the filtered sodium load. The cellular mechanisms are complex but some of the key features are shown in Figure 16.3A. Filtered sodium in the luminal fluid enters the cell via several sodium transporters in the apical membrane that couple sodium transport to the entry of glucose, amino acid, phosphate and other organic molecules. Entry of sodium into the tubular cells at this site is also linked to secretion of H⁺ ions, through the tubular fluid into the peritubular capillaries is 99%.

Since the kidney maintains the constancy of body fluids by adjusting urine volume and composition, it is frequently helpful to obtain a sample of urine (‘spot’ specimen or 24-hour collection) at the time of blood analysis. An example of the use of urine biochemistry is given for the differential diagnosis of hyponatraemia in Box 16.14 (p. 438).

DISORDERS OF SODIUM BALANCE

Functional anatomy and physiology of renal sodium handling

Since the great majority of the body’s sodium content is located in the ECF, where it is by far the most abundant cation, total body sodium is a principal determinant of ECF volume. Regulation of sodium excretion by the kidney is crucially important in maintaining normal ECF volume, and plasma volume, in the face of wide variations in sodium intake, which typically may range between 50 and 250 mmol/day.

The functional unit for renal excretion is the nephron (Fig. 16.2). Blood undergoes ultrafiltration in the glomerulus, generating a fluid that is free from cells and protein and which resembles plasma in its electrolyte composition. This is then delivered into the renal tubules, where reabsorption of water and various electrolytes occurs (more detail on the structure and function of the glomerulus is given in Ch. 17). The glomerular filtration rate (GFR) is approximately 125 mL/min (equivalent to 180 L/day) in a typical adult. Over 99% of this filtered fluid is reabsorbed into the blood in the peritubular capillaries during its passage through successive segments of the nephron, largely as a result of tubular reabsorption of sodium. The processes mediating this sodium reabsorption, and the factors that regulate it, are key to understanding clinical disturbances and pharmacological interventions.

Nephron segments

At least four different functional segments of the nephron can be defined in terms of their mechanism for sodium reabsorption (Fig. 16.3).

Proximal tubule

This is responsible for the reabsorption of some 65% of the filtered sodium load. The cellular mechanisms are complex but some of the key features are shown in Figure 16.3A. Filtered sodium in the luminal fluid enters the cell via several sodium transporters in the apical membrane that couple sodium transport to the entry of glucose, amino acid, phosphate and other organic molecules. Entry of sodium into the tubular cells at this site is also linked to secretion of H⁺ ions, through the tubular fluid into the peritubular capillaries is 99%.

Since the kidney maintains the constancy of body fluids by adjusting urine volume and composition, it is frequently helpful to obtain a sample of urine (‘spot’ specimen or 24-hour collection) at the time of blood analysis. An example of the use of urine biochemistry is given for the differential diagnosis of hyponatraemia in Box 16.14 (p. 438).

DISORDERS OF SODIUM BALANCE

Functional anatomy and physiology of renal sodium handling

Since the great majority of the body’s sodium content is located in the ECF, where it is by far the most abundant cation, total body sodium is a principal determinant of ECF volume. Regulation of sodium excretion by the kidney is crucially important in maintaining normal ECF volume, and plasma volume, in the face of wide variations in sodium intake, which typically may range between 50 and 250 mmol/day.

The functional unit for renal excretion is the nephron (Fig. 16.2). Blood undergoes ultrafiltration in the glomerulus, generating a fluid that is free from cells and protein and which resembles plasma in its electrolyte composition. This is then delivered into the renal tubules, where reabsorption of water and various electrolytes occurs (more detail on the structure and function of the glomerulus is given in Ch. 17). The glomerular filtration rate (GFR) is approximately 125 mL/min (equivalent to 180 L/day) in a typical adult. Over 99% of this filtered fluid is reabsorbed into the blood in the peritubular capillaries during its passage through successive segments of the nephron, largely as a result of tubular reabsorption of sodium. The processes mediating this sodium reabsorption, and the factors that regulate it, are key to understanding clinical disturbances and pharmacological interventions.

Nephron segments

At least four different functional segments of the nephron can be defined in terms of their mechanism for sodium reabsorption (Fig. 16.3).

Proximal tubule

This is responsible for the reabsorption of some 65% of the filtered sodium load. The cellular mechanisms are complex but some of the key features are shown in Figure 16.3A. Filtered sodium in the luminal fluid enters the cell via several sodium transporters in the apical membrane that couple sodium transport to the entry of glucose, amino acid, phosphate and other organic molecules. Entry of sodium into the tubular cells at this site is also linked to secretion of H⁺ ions, through the tubular fluid into the peritubular capillaries is 99%.

Since the kidney maintains the constancy of body fluids by adjusting urine volume and composition, it is frequently helpful to obtain a sample of urine (‘spot’ specimen or 24-hour collection) at the time of blood analysis. An example of the use of urine biochemistry is given for the differential diagnosis of hyponatraemia in Box 16.14 (p. 438).

DISORDERS OF SODIUM BALANCE

Functional anatomy and physiology of renal sodium handling

Since the great majority of the body’s sodium content is located in the ECF, where it is by far the most abundant cation, total body sodium is a principal determinant of ECF volume. Regulation of sodium excretion by the kidney is crucially important in maintaining normal ECF volume, and plasma volume, in the face of wide variations in sodium intake, which typically may range between 50 and 250 mmol/day.

The functional unit for renal excretion is the nephron (Fig. 16.2). Blood undergoes ultrafiltration in the glomerulus, generating a fluid that is free from cells and protein and which resembles plasma in its electrolyte composition. This is then delivered into the renal tubules, where reabsorption of water and various electrolytes occurs (more detail on the structure and function of the glomerulus is given in Ch. 17). The glomerular filtration rate (GFR) is approximately 125 mL/min (equivalent to 180 L/day) in a typical adult. Over 99% of this filtered fluid is reabsorbed into the blood in the peritubular capillaries during its passage through successive segments of the nephron, largely as a result of tubular reabsorption of sodium. The processes mediating this sodium reabsorption, and the factors that regulate it, are key to understanding clinical disturbances and pharmacological interventions.

Nephron segments

At least four different functional segments of the nephron can be defined in terms of their mechanism for sodium reabsorption (Fig. 16.3).

Proximal tubule

This is responsible for the reabsorption of some 65% of the filtered sodium load. The cellular mechanisms are complex but some of the key features are shown in Figure 16.3A. Filtered sodium in the luminal fluid enters the cell via several sodium transporters in the apical membrane that couple sodium transport to the entry of glucose, amino acid, phosphate and other organic molecules. Entry of sodium into the tubular cells at this site is also linked to secretion of H⁺ ions, through the tubular fluid into the peritubular capillaries is 99%.

Since the kidney maintains the constancy of body fluids by adjusting urine volume and composition, it is frequently helpful to obtain a sample of urine (‘spot’ specimen or 24-hour collection) at the time of blood analysis. An example of the use of urine biochemistry is given for the differential diagnosis of hyponatraemia in Box 16.14 (p. 438).
sodium–hydrogen exchanger (NHE-3). Intracellular H⁺ ions are generated within tubular cells from the breakdown of carbonic acid, which is produced from carbon dioxide and water under the influence of carbonic anhydrase. Large numbers of Na,K-ATPase pumps are present on the basolateral membrane of tubular cells that transport sodium from the cells into the blood. In addition, a large component of the transepithelial flux of sodium, water and other dissolved solutes occurs through the gaps between the cells (the ‘shunt’ pathway). Overall, fluid and electrolyte reabsorption is almost isosmotic in this segment, as water reabsorption is matched very closely to sodium fluxes. A component of this water flow also passes through the cells, via aquaporin-1 (AQP-1) water channels, which are not sensitive to hormonal regulation.

The loop of Henle
The thick ascending limb of the loop of Henle (Fig. 16.3B) reabsorbs a further 25% of the filtered sodium but is impermeable to water, resulting in dilution of the luminal fluid. Again, the primary driving force for sodium reabsorption is the Na,K-ATPase on the basolateral cell membrane, but in this segment sodium enters the cell from the lumen via a specific carrier molecule, the Na,K,2Cl co-transporter (‘triple co-transporter’, or NKCC2), which allows electroneutral entry of these ions into the renal tubular cell. Some of the potassium accumulated inside the cell recirculates across the apical membrane back into the lumen through a specific potassium channel (ROMK), providing a continuing supply of potassium to match the high concentrations of sodium and chloride available in the lumen. A small positive transepithelial potential difference exists in the lumen of this segment relative to the interstitium, and this serves to drive cations such as sodium, potassium, calcium and magnesium between the cells, forming a reabsorptive shunt pathway.

Early distal tubule
Some 6% of filtered sodium is reabsorbed in the early distal tubule (also called distal convoluted tubule) (Fig. 16.3C), again driven by the activity of the basolateral Na,K-ATPase. In this segment, entry of sodium into the cell from the luminal fluid is via a sodium–chloride co-transport carrier (NCCT). This segment is also impermeable to water, resulting in further dilution of the luminal fluid. There is no significant transepithelial flux of potassium in this segment, but calcium is reabsorbed through the mechanism shown in Figure 16.3C: a
CLINICAL BIOCHEMISTRY AND METABOLISM

basolateral sodium–calcium exchanger leads to low intracellular concentrations of calcium, promoting calcium entry from the luminal fluid through a calcium channel.

Late distal tubule and collecting ducts

The late distal tubule and cortical collecting duct are anatomically and functionally continuous (Fig. 16.3D). Here, sodium entry from the luminal fluid occurs via the epithelial sodium channel (ENaC) through which sodium passes alone, generating a substantial lumen-negative transepithelial potential difference. This sodium flux into the tubular cells is balanced by secretion of potassium and hydrogen ions into the lumen and by reabsorption of chloride ions. Potassium is accumulated in the cell by the basolateral Na,K-ATPase, and passes into the luminal fluid down its electrochemical gradient, through an apical potassium channel (ROMK). Chloride ions pass largely between cells. Hydrogen ion secretion is mediated by an H\(^+\)-ATPase located on the luminal membrane of the intercalated cells, which constitute approximately one-third of the epithelial cells in this nephron segment. This part of the nephron has a variable permeability to water, depending on the availability of antidiuretic hormone (ADH, or vasopressin) in the circulation. All ion transport processes in this segment are stimulated by the steroid hormone aldosterone, which can increase sodium reabsorption in this segment to a maximum of 2–3% of the filtered sodium load.

Less than 1% of sodium reabsorption occurs in the medullary collecting duct, where it is inhibited by atrial natriuretic peptide (ANP) and brain natriuretic peptide (BNP).

Regulation of sodium transport

A large number of interrelated mechanisms serve to maintain whole body sodium balance and hence ECF volume by matching urinary sodium excretion to sodium intake (Fig. 16.4).

Important sensing mechanisms include volume receptors in the cardiac atria and the intrathoracic veins, as well as pressure receptors located in the central arterial tree (aortic arch and carotid sinus) and the afferent arterioles within the kidney. A further afferent signal is generated within the kidney itself; the enzyme renin is released from specialised smooth muscle cells in the walls of the afferent and efferent arterioles, at the point where they make contact with the early distal tubule (at the macula densa) to form the juxtaglomerular apparatus. Renin release is stimulated by:

- reduced perfusion pressure in the afferent arteriole
- increased sympathetic nerve activity
- decreased sodium chloride concentration in the distal tubular fluid.

Renin released into the circulation activates the effector mechanisms for sodium retention, which are components of the renin–angiotensin–aldosterone (RAA) system (see Fig. 20.17, p. 771). Renin acts on the peptide substrate, angiotensinogen (manufactured in the liver), producing angiotensin I in the circulation. This in turn is cleaved by angiotensin-converting enzyme (ACE) into angiotensin II, largely in the pulmonary capillary bed. Angiotensin II has multiple actions: it stimulates proximal tubular sodium reabsorption and release of aldosterone from the zona glomerulosa of the adrenal cortex, and causes vasoconstriction of small arterioles. Aldosterone amplifies sodium retention by its action on the cortical collecting duct. The net effect is to restore ECF volume and blood pressure towards normal, thereby correcting the initiating hypovolaemic stimulus.

The sympathetic nervous system also increases sodium retention, both through haemodynamic mechanisms (afferent arteriolar vasoconstriction and GFR reduction) and by direct stimulation of proximal tubular sodium reabsorption. Other humoral mediators, such as the natriuretic peptides, inhibit sodium reabsorption, contributing to natriuresis during periods of sodium and volume excess. Hypovolaemia also has haemodynamic effects that reduce GFR and alter the peritubular physical forces around the proximal tubule, thereby decreasing sodium excretion. Conversely, increased renal perfusion in hypervolaemia and hypertension results in a compensatory increase in sodium excretion.

Fig. 16.4 Mechanisms involved in the regulation of sodium transport. (ANP = atrial natriuretic peptide; BNP = brain natriuretic peptide; ECF = extracellular fluid; GFR = glomerular filtration rate; RAA = renin–angiotensin–aldosterone system; SNS = sympathetic nervous system. ⊘ indicates an effect to stimulate Na reabsorption and hence reduce Na excretion, while ⊗ indicates an effect to inhibit Na reabsorption and hence increase Na excretion)

Presenting problems in disorders of sodium balance

When the balance of sodium intake and excretion is disturbed, any tendency for plasma sodium concentration to change is usually corrected by the osmotic mechanisms controlling water balance (p. 436). As a result, disorders in sodium balance present chiefly as alterations in the ECF volume, resulting in hypovolaemia or oedema, rather than as an alteration in plasma sodium concentration. Clinical manifestations of altered volume are illustrated in Box 16.4.

Sodium depletion

Aetiology and clinical assessment

Sodium depletion can occur occasionally under extreme environmental conditions due to inadequate intake of
Management

Management of sodium and water depletion has two main components:
- treat the cause where possible, to stop ongoing salt and water losses
- replace the salt and water deficits, and provide ongoing maintenance requirements, usually by intravenous fluid replacement when depletion is severe.

Intravenous fluid therapy

Box 16.6 shows the daily maintenance requirements for water and electrolytes in a typical adult, and Box 16.7 summarises the composition of some widely available intravenous fluids. The choice of fluid and the rate of administration depend on the clinical circumstances, as assessed at the bedside and from laboratory data, and as described in Box 16.8.

In the absence of normal oral intake (as in a fasting or post-operative patient in hospital), maintenance quantities of fluid, sodium and potassium should be provided. If any deficits or continuing pathological losses are identified, additional fluid and electrolytes will be required. In prolonged periods of fasting (more than a few days), attention also needs to be given to providing sufficient caloric and nutritional intake to prevent excessive catabolism of body energy stores (p. 120).

The choice of intravenous fluid therapy in the treatment of significant hypovolaemia relates to the concepts in Figure 16.1 (p. 429). If fluid containing neither sodium nor protein is given, it will distribute in the body fluid compartments in proportion to the normal distribution of total body water. Thus, giving 1 L of 5% dextrose will contribute relatively little (approximately 3/40 of the infused volume) towards expansion of the plasma volume. This makes 5% dextrose ineffective at restoring the circulation and perfusion of vital organs. Intravenous infusion of an isotonic (normal) saline solution, on the other hand, results in more effective expansion of the extracellular fluid, although a minority of the infused volume (some 3/15) will contribute to plasma volume.

In prolonged periods of fasting (more than a few days), attention also needs to be given to providing sufficient caloric and nutritional intake to prevent excessive catabolism of body energy stores (p. 120).

The choice of intravenous fluid therapy in the treatment of significant hypovolaemia relates to the concepts in Figure 16.1 (p. 429). If fluid containing neither sodium nor protein is given, it will distribute in the body fluid compartments in proportion to the normal distribution of total body water. Thus, giving 1 L of 5% dextrose will contribute relatively little (approximately 3/40 of the infused volume) towards expansion of the plasma volume. This makes 5% dextrose ineffective at restoring the circulation and perfusion of vital organs. Intravenous infusion of an isotonic (normal) saline solution, on the other hand, results in more effective expansion of the extracellular fluid, although a minority of the infused volume (some 3/15) will contribute to plasma volume.
16.8 How to assess fluid and electrolyte balance in hospitalised patients

**Step 1: assess clinical volume status**
- Examine patient for signs of hypovolaemia or hypervolaemia (see Box 16.4)
- Check daily weight change

**Step 2: review fluid balance chart**
- Check total volumes IN and OUT on previous day (IN–OUT is positive by ~400 mL in normal balance, reflecting insensible fluid losses of ~800 mL and metabolic water generation of ~400 mL)
- Check cumulative change in daily fluid balance over previous 3–5 days
- Correlate chart figures with weight change and clinical volume status to estimate net fluid balance

**Step 3: assess ongoing pathological process**
- Check losses from gastrointestinal tract and surgical drains
- Estimate increased insensible losses (e.g. in fever) and internal sequestration (‘third space’)

**Step 4: check plasma U&Es (see Box 16.3)**
- Check plasma Na as marker of relative water balance
- Check plasma K as a guide to extracellular K balance
- Check HCO₃⁻ as a clue to acid–base disorder
- Check urea and creatinine to monitor renal function

**Step 5: prescribe appropriate IV fluid replacement therapy**
- Replace basic water and electrolytes each day (see Box 16.6)
- Allow for anticipated oral intake and pathological fluid loss
- Adjust amounts of water (if IV, usually given as isotonic 5% dextrose), sodium and potassium according to plasma electrolyte results

16.9 Albumin infusions in hypovolaemia

"For patients with hypovolaemia there is no evidence that albumin reduces mortality when compared with cheaper alternatives such as saline."


Carrying this reasoning further, it might be expected that a solution containing plasma proteins would be largely retained within the plasma, thus maximally expanding the circulating fluid volume and improving tissue perfusion. However, recent clinical studies have not shown any overall advantage of infusions containing albumin in the treatment of acute hypovolaemia (Box 16.9). Resuscitation fluids containing synthetic colloids such as carbohydrate polymers should not be used in the acute resuscitation of volume-depleted patients since they offer no benefit over crystalloids and are associated with increased mortality (see Box 17.21, p. 482).

**Sodium excess**

**Aetiology and clinical assessment**

In patients with normal cardiac and renal function, excessive intakes of salt and water are compensated for by increased excretion and do not lead to clinically obvious features of sodium and water overload. However, patients with cardiac, renal or hepatic disease frequently present with signs and symptoms of sodium excess (Fig. 16.5). This does not always involve an increase in circulating blood volume, since the excess fluid often leaks out of the capillaries to expand the interstitial compartment of the ECF, especially in diseases like nephrotic syndrome and chronic liver disease that cause hypoalbuminaemia. Important causes of sodium excess are shown in Box 16.10.

Peripheral oedema is the most common physical sign of ECF volume expansion (p. 478). The three most common systemic disorders associated with sodium and fluid overload are cardiac failure, cirrhosis and nephrotic syndrome. In each of these, sodium retention is largely a secondary response to circulatory insufficiency caused by the primary disorder, as illustrated in Figure 16.5. The pathophysiology is different in renal failure, when the primary cause of volume expansion is the profound reduction in GFR impairing sodium and water excretion, and secondary tubular mechanisms are of lesser importance. Further detail on each of these conditions is given in other chapters of this book.

**Management**

The management of ECF volume overload involves a number of components:
- specific treatment directed at the underlying cause, such as ACE inhibitors in heart failure and corticosteroids in minimal change nephropathy
- restriction of dietary sodium (to 50–80 mmol/day) to match the diminished excretory capacity
- treatment with diuretics.

**Diuretic therapy**

Diuretics are important in the treatment of conditions of ECF expansion due to salt and water retention and in hypertension (p. 606). They act by inhibiting sodium reabsorption at various locations along the nephron (see Fig. 16.3, p. 431). Their potency and adverse effects relate to their mechanism and site of action.

**Mechanisms of action**

In the proximal tubule, carbonic anhydrase inhibitors such as acetazolamide inhibit the intracellular production of H⁺ ions, thereby reducing the fraction of sodium reabsorption that is exchanged for H⁺ by the apical
Another is the mineralocorticoid receptor, to which binding of aldosterone is blocked by spironolactone and eplerenone.

An important feature of the most commonly used diuretic drugs (furosemide, thiazides and amiloride) is that they act on their target transport molecules from the luminal side of the tubular epithelium. Since they are highly protein-bound in the plasma, very little reaches the urinary fluid by glomerular filtration, but there are active transport mechanisms for secreting organic acids and bases, including these drugs, across the proximal tubular wall into the lumen, resulting in adequate drug concentrations being delivered to later tubular segments. This secretory process may be impaired by certain other drugs, and also by accumulated organic anions as occurs in chronic renal failure and chronic liver failure, leading to resistance to diuretics.

Osmotic diuretics act independently of specific transport mechanism. They are freely filtered at the glomerulus but are not reabsorbed by any part of the tubular system. They retain fluid osmotically within the tubular lumen and limit the extent of sodium reabsorption in multiple segments. Mannitol is the most commonly used osmotic diuretic. It is given by intravenous infusion to achieve short-term diuresis in conditions such as cerebral oedema.

Clinical use of diuretics

In the selection of a diuretic drug for hypertension or oedema disorders, the following principles should be observed:

- Use the minimum effective dose.
- Use for as short a period of time as necessary.
- Monitor regularly for adverse effects.

---

**Fig. 16.5** Secondary mechanisms causing sodium excess and oedema in cardiac failure, cirrhosis and nephrotic syndrome. Primary renal retention of Na and water may also contribute to oedema formation when GFR is significantly reduced (see Box 16.10 and p. 478).
The choice of diuretic will be determined by the required potency, the presence of coexistent conditions, and the anticipated side-effect profile.

Adverse effects encountered with the most commonly used classes of diuretic (loop drugs and thiazide diuretics) are summarised in Box 16.11. Volume depletion and electrolyte disorders commonly occur, as predicted from their mechanism of action. The metabolic side-effects listed are rarely of clinical significance and may reflect effects on K⁺ channels that influence insulin secretion (p. 800). Since most drugs from these classes are sulphonamides, there is a relatively high incidence of hypersensitivity reactions, and occasional idiosyncratic side-effects in a variety of organ systems.

The side-effect profile of the potassium-sparing diuretics differs in a number of important respects from that of other diuretics. The disturbances in potassium, magnesium and acid-base balance are in the opposite direction, so that normal or increased levels of potassium and magnesium are found in the blood, and there is a tendency to metabolic acidosis, especially when renal function is impaired.

Diuretic resistance is encountered under a variety of circumstances, including impaired renal function, activation of sodium-retaining mechanisms, impaired oral bioavailability (for example, in patients with gastrointestinal disease) and decreased renal blood flow. In these circumstances, short-term intravenous therapy with a loop-acting agent such as furosemide may be useful. Combinations of diuretics administered orally may also increase potency. Either a loop or a thiazide drug can be combined with a potassium-sparing drug, and all three classes can be used together for short periods, with carefully supervised clinical and laboratory monitoring.

## DISORDERS OF WATER BALANCE

Daily water intake can vary from about 500 mL to several litres a day. While a certain amount of water is lost through the stool, sweat and the respiratory tract (‘insensible losses’, approximately 800 mL/day), and some water is generated by oxidative metabolism (‘metabolic water’, approximately 400 mL/day), the kidneys are chiefly responsible for adjusting water excretion to maintain constancy of body water content and body fluid osmolality (reference range 280–295 mmol/kg).

### Functional anatomy and physiology of renal water handling

While regulation of total ECF volume is largely achieved through renal control of sodium excretion, mechanisms also exist to allow for the excretion of a ‘pure’ water load when water intake is high, and for retention of water when access is restricted.

These functions are largely achieved by the loop of Henle and the collecting ducts. The counter-current configuration of flow in adjacent limbs of the loop (see Fig. 16.2, p. 430), involves osmotic movement of water from the descending limbs and reabsorption of solute from neighbouring ascending limbs, to set up a gradient of osmolality from isotonic (like plasma) in the renal cortex to hypertonic (around 1200 mmol/kg) in the inner part of the medulla. At the same time, the fluid emerging from the thick ascending limb is hypotonic compared to plasma, because it has been diluted by the reabsorption of sodium, but not water, from the thick ascending limb and early distal tubule. As this dilute fluid passes from the cortex through the collecting duct system to the renal pelvis, it traverses the medullary interstitial gradient of osmolality set up by the operation of the loop of Henle, and water is reabsorbed.

Further changes in the urine osmolality on passage through the collecting ducts depend on the circulating level of antidiuretic hormone (ADH), which is released by the posterior pituitary gland under conditions of increased plasma osmolality or hypovolaemia (Ch. 20).

- When water intake is high and plasma osmolality is normal or low-normal, ADH levels are suppressed and the collecting ducts remain impermeable to water. The luminal fluid osmolality remains low, resulting in the excretion of a dilute urine (minimum osmolality approximately 50 mmol/kg in a healthy young person).
- When water intake is restricted and plasma osmolality is high, or in the presence of plasma volume depletion, ADH levels rise. This causes water permeability of the collecting ducts to increase through binding of ADH to the V2 receptor, which enhances collecting duct water permeability through the insertion of aquaporin (AQP-2) channels into the luminal cell membrane. This results in osmotic reabsorption of water along the entire length of the collecting duct, with maximum urine osmolality approaching that in the medullary tip (up to 1200 mmol/kg).

Parallel to these changes in ADH release are changes in water-seeking behaviour triggered by the sensation of thirst, which also becomes activated as plasma osmolality rises.

In summary, for adequate dilution of the urine there must be:

- adequate solute delivery to the loop of Henle and early distal tubule
- normal function of the loop of Henle and early distal tubule
- absence of ADH in the circulation.

If any of these processes is faulty, water retention and hyponatraemia may result.
Conversely, to achieve concentration of the urine there must be:

- adequate solute delivery to the loop of Henle
- normal function of the loop of Henle
- ADH release into the circulation
- ADH action on the collecting ducts.

Failure of any of these steps may result in inappropriate water loss and hypernatraemia.

**Presenting problems in disorders of water balance**

Disturbances in body water balance, in the absence of changes in sodium balance, alter plasma sodium concentration and hence plasma osmolality. When extracellular osmolality changes abruptly, water flows rapidly across cell membranes with resultant cell swelling (during hypo-osmolality) or shrinkage (during hyper-osmolality). Cerebral function is very sensitive to such volume changes, particularly brain swelling during hypo-osmolality, which can lead to an increase in intracerebral pressure and reduced cerebral perfusion.

**Hyponatraemia**

**Aetiology and clinical assessment**

Hyponatraemia (plasma Na < 135 mmol/L) is a common electrolyte abnormality, which is often asymptomatic but which can also be associated with profound disturbances of cerebral function, manifesting as anorexia, nausea, vomiting, confusion, lethargy, seizures and coma. The likelihood of symptoms occurring is related more to the speed at which electrolyte abnormalities develop rather than their severity. When plasma osmolality falls rapidly, water flows into cerebral cells, which become swollen and ischaemic. However, when hyponatraemia develops gradually, cerebral neurons have time to respond by reducing intracellular osmolality, through excreting potassium and reducing synthesis of intracellular organic osmolytes (Fig. 16.6). The osmotic gradient favouring water movement into the cells is thus reduced and symptoms are avoided.

The causes of hyponatraemia are best categorised according to any associated changes in the ECF volume (Box 16.12). In all cases, there is retention of water relative to sodium, and it is clinical examination rather than the biochemical results that gives a clue to the underlying cause.

Artefactual causes of hyponatraemia should also be considered. These can occur in the presence of severe hyperlipidaemia or hyperproteininaemia, when the aqueous fraction of the plasma specimen is reduced because of the volume occupied by the macromolecules (although this artefact is dependent on the assay technology). Transient hyponatraemia may also occur due to osmotic shifts of water out of cells during hyperosmolar states caused by acute hyperglycaemia or by mannitol infusion.

**Hyponatraemia with hypovolaemia**

Patients who have hyponatraemia in association with a sodium deficit (‘depletional hyponatraemia’) have clinical features of hypovolaemia (see Box 16.4, p. 433) and supportive laboratory findings, including low urinary

---

**Fig. 16.6 Hyponatraemia and the brain.** Numbers represent osmolality (osmo) in mmol/kg.

**Table 16.12 Causes of hyponatraemia**

<table>
<thead>
<tr>
<th>Volume status</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypovolaemic (sodium deficit with a relatively smaller water deficit)</td>
<td>Renal sodium losses&lt;br&gt;Diuretic therapy (especially thiazides)&lt;br&gt;Adrenocortical failure&lt;br&gt;Gastrointestinal sodium losses&lt;br&gt;Vomiting&lt;br&gt;Diarrhoea&lt;br&gt;Skin sodium losses&lt;br&gt;Burns</td>
</tr>
<tr>
<td>Euvolaemic (water retention alone)</td>
<td>Primary polydipsia&lt;br&gt;Excessive electrolyte-free water infusion&lt;br&gt;SIADH&lt;br&gt;Hypothyroidism</td>
</tr>
<tr>
<td>Hypervolaemic (sodium retention with relatively greater water retention)</td>
<td>Congestive cardiac failure&lt;br&gt;Cirrhosis&lt;br&gt;Nephrotic syndrome&lt;br&gt;Chronic renal failure (during free water intake)</td>
</tr>
</tbody>
</table>

(SIADH = syndrome of inappropriate antidiuretic hormone secretion; see Box 16.13).
sodium concentration (< 30 mmol/L) and elevated plasma renin activity. The cause of sodium loss is usually apparent; common examples are shown in Box 16.12.

**Hyponatraemia with euolaemia**

Patients in this group (dilutional hyponatraemia) have no major disturbance of body sodium content and are clinically euolaemic. Excess body water may be the result of abnormally high intake, either orally (primary polydipsia) or as a result of medically infused fluids (as intravenous dextrose solutions, or by absorption of sodium-free bladder irrigation fluid after prostatectomy).

Water retention also occurs in the syndrome of inappropriate secretion of ADH (SIADH). In this condition, an endogenous source of ADH (either cerebral or tumour-derived) promotes water retention by the kidney in the absence of an appropriate physiological stimulus (Box 16.13). The clinical diagnosis requires the patient to be euolaemic, with no evidence of cardiac, renal or hepatic disease potentially associated with hyponatraemia. Other non-osmotic stimuli that cause release of ADH (pain, stress, nausea) should also be excluded. Supportive laboratory findings are shown in Box 16.13. In this situation, plasma concentrations of sodium, chloride, urea and uric acid are low with a correspondingly reduced osmolality. Urine osmolality, which should physiologically be maximally dilute (approximately 50 mmol/kg) in the face of low plasma osmolality, is higher than at least 100 mmol/kg and indeed is typically higher than the plasma osmolality. The urine sodium concentration is typically high (> 30 mmol/L), consistent with euolaemia and lack of compensatory factors promoting sodium retention.

**Hyponatraemia with hypervolaemia**

In this situation, excess water retention is associated with sodium retention and volume expansion, as in heart failure, liver disease or kidney disease.

**Investigations**

Plasma and urine electrolytes and osmolality (Box 16.14) are usually the only tests required to classify the hyponatraemia. Doubt about clinical signs of ECF volume may be resolved with measurement of plasma renin activity.

<table>
<thead>
<tr>
<th>Urine Na (mmol/L)</th>
<th>Urine osmolality (mmol/kg)</th>
<th>Possible diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (&lt; 30)</td>
<td>Low (&lt; 100)</td>
<td>Primary polydipsia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malnutrition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beer excess</td>
</tr>
<tr>
<td>Low</td>
<td>High (&gt; 150)</td>
<td>Salt depletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypovolaemia</td>
</tr>
<tr>
<td>High (&gt; 40)</td>
<td>Low</td>
<td>Diuretic action (acute phase)</td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>SIADH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cerebral salt-wasting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adrenal insufficiency</td>
</tr>
</tbody>
</table>

*Urine analysis may give results of indeterminate significance, and in this case the diagnosis depends on a comprehensive clinical assessment.*

Measurement of ADH is not generally helpful in distinguishing between these categories of hyponatraemia. This is because ADH is activated both in hypovolaemic states and in most chronic hypervolaemic states, as the impaired circulation in those disorders activates ADH release through non-osmotic mechanisms. Indeed, these disorders may have higher circulating ADH levels than patients with SIADH. The only disorders listed in Box 16.12 in which ADH is suppressed are primary polydipsia and iatrogenic water intoxication, where the hypo-osmolar state inhibits ADH release from the pituitary.

**Management**

The treatment of hyponatraemia is critically dependent on its rate of development, severity and underlying cause. If hyponatraemia has developed rapidly (over hours to days), and there are signs of cerebral oedema such as obtundation or convulsions, sodium levels should be restored to normal rapidly by infusion of hypertonic (3%) sodium chloride. A common approach is to give an initial bolus of 100 mL, which may be repeated once or twice over the initial hours of observation, depending on the neurological response and rise in plasma sodium.

On the other hand, rapid correction of hyponatraemia that has developed slowly (over weeks to months) can be hazardous, since brain cells adapt to slowly developing hypo-osmolality by reducing the intracellular osmolality, thus maintaining normal cell volume (see Fig. 16.6). Under these conditions, an abrupt increase in extracellular osmolality can lead to water shifting out of neurons, abruptly reducing their volume and causing them to detach from their myelin sheaths. The resulting ‘myelinolysis’ can produce permanent structural and functional damage to mid-brain structures, and is generally fatal. The rate of correction of the plasma Na concentration in chronic asymptomatic hyponatraemia should not exceed 10 mmol/L/day, and an even slower rate is generally safer.

The underlying cause should be treated. For hypervolaemic patients, this involves controlling the source of sodium loss, and administering intravenous saline if clinically warranted. Patients with dilutional hyponatraemia generally respond to fluid restriction in the range of 600–1000 mL/day, accompanied where
possible by withdrawal of the precipitating stimulus (such as drugs causing SIADH). If the response of plasma sodium is inadequate, treatment with demeclocycline (600–900 mg/day) may be of value by enhancing water excretion, through its inhibitory effect on responsiveness to ADH in the collecting duct. An effective alternative for subjects with persistent hyponatraemia due to prolonged SIADH is oral urea therapy (30–45 g/day), which provides a solute load to promote water excretion. Where available, oral vasopressin receptor antagonists such as tolvaptan may be used to block the ADH-mediated component of water retention in a range of hyponatraemic conditions. Hyperosmolar patients with hyponatraemia need treatment of the underlying condition, accompanied by cautious use of diuretics in conjunction with strict fluid restriction. Potassium-sparing diuretics may be particularly useful in this context where there is significant secondary hyperaldosteronism.

Hypernatraemia

Aetiology and clinical assessment

Just as hyponatraemia represents a failure of the mechanisms for diluting the urine during free access to water, so hypernatraemia (plasma Na > 148 mmol/L) reflects inadequate concentration of the urine in the face of restricted water intake. This can be due to failure to generate an adequate medullary concentration gradient (low GFR states, loop diuretic therapy), but more commonly it is due to failure of the ADH system, either because of pituitary damage (central or ‘cranial’ diabetes insipidus, p. 794) or because the collecting duct cells are unable to respond to circulating ADH (nephrogenic diabetes insipidus).

Patients with hypernatraemia generally have reduced cerebral function, either as a primary problem or as a consequence of the hypernatraemia itself, which results in dehydration of neurons and brain shrinkage. In the presence of an intact thirst mechanism and preserved capacity to obtain and ingest water, hypernatraemia may not progress very far. If adequate water is not obtained, dizziness, confusion, weakness and ultimately coma and death can result.

The causes of hypernatraemia are best grouped according to the associated disturbance, if any, in total body sodium content (Box 16.15). It is important to remember that hypernatraemia may be iatrogenic, and to reiterate that, whatever the underlying cause, sustained or severe hypernatraemia generally reflects an impaired thirst mechanism or responsiveness to thirst.

Management

Treatment of hypernatraemia depends on both the rate of development and the underlying cause. If there is reason to think that the condition has developed rapidly, neuronal shrinkage may be acute and relatively rapid correction may be attempted. This can be achieved by infusing an appropriate volume of intravenous fluid (isotonic 5% dextrose or hypotonic 0.45% saline) at an initial rate of 50–70 mL/hour. However, in older, institutionalised patients it is more likely that the disorder has developed slowly, and extreme caution should be exercised in lowering plasma sodium to avoid the risk of cerebral oedema. Where possible, the underlying cause should also be addressed (see Box 16.15).

16.15 Causes of hypernatraemia

<table>
<thead>
<tr>
<th>Volume status</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypovolaemic (sodium deficit with a relatively greater water deficit)</td>
<td>Renal sodium losses</td>
</tr>
<tr>
<td></td>
<td>Diuretic therapy (especially osmotic diuretic, or loop diuretic during water restriction)</td>
</tr>
<tr>
<td></td>
<td>Glycosuria (HONK, p. 814)</td>
</tr>
<tr>
<td></td>
<td>Gastrointestinal Na losses</td>
</tr>
<tr>
<td></td>
<td>Colonic diarrhoea</td>
</tr>
<tr>
<td></td>
<td>Skin sodium losses</td>
</tr>
<tr>
<td></td>
<td>Excessive sweating</td>
</tr>
<tr>
<td>Euvoalaemic (water deficit alone)</td>
<td>Diabetes insipidus (central or nephrogenic) (p. 794)</td>
</tr>
<tr>
<td>Hypovolaemic (sodium retention with relatively less water retention)</td>
<td>Enteral or parenteral feeding</td>
</tr>
<tr>
<td></td>
<td>IV or oral salt administration</td>
</tr>
<tr>
<td></td>
<td>Chronic renal failure (during water restriction)</td>
</tr>
</tbody>
</table>

16.16 Hyponatraemia and hypernatraemia in old age

- Decline in GFR: older patients are predisposed to both hyponatraemia and hypernatraemia, mainly because, as GFR declines with age, the capacity of the kidney to dilute or concentrate the urine is impaired.
- Hyponatraemia: occurs when free water intake continues in the presence of a low dietary salt intake and/or diuretic drugs (particularly thiazides).
- ADH release: water retention is aggravated by any condition that stimulates ADH release, especially heart failure. Moreover, the ADH response to non-osmotic stimuli may be brisker in older subjects. Appropriate water restriction may be a key part of management.
- Hypernatraemia: occurs when water intake is inadequate, due to physical restrictions preventing access to drinks and/or blunted thirst. Both are frequently present in patients with advanced dementia or following a severe stroke.
- Dietary salt: hypernatraemia is aggravated if dietary supplements or medications with a high sodium content (especially effervescent preparations) are administered. Appropriate prescription of fluids is a key part of management.

Elderly patients are predisposed, in different circumstances, to both hyponatraemia and hypernatraemia, and a high index of suspicion of these electrolyte disturbances is appropriate in elderly patients with recent alterations in behaviour (Box 16.16).

DISORDERS OF POTASSIUM BALANCE

Potassium is the major intracellular cation (see Fig. 16.1, p. 429), and the steep concentration gradient for potassium across the cell membrane of excitable cells plays an important part in generating the resting membrane potential and allowing the propagation of the action potential that is crucial to normal functioning of nerve, muscle and cardiac tissues. Control of body potassium balance is described below.

Changes in the distribution of potassium between the ICF and ECF compartments can alter plasma potassium concentration, without any overall change in total body potassium content. Potassium is driven into the cells by
Clinical Biochemistry and Metabolism

extracellular alkalosis and by a number of hormones, including insulin, catecholamines (through the β2 receptor) and aldosterone. Any of these factors can produce hypokalaemia, whereas extracellular acidosis, lack of insulin, and insufficiency or blockade of catecholamines or aldosterone can cause hyperkalaemia due to efflux of potassium from the intracellular compartment.

Functional anatomy and physiology of renal potassium handling

In the steady state, the kidneys excrete some 90% of the daily intake of potassium, typically 80–100 mmol/day. Potassium is freely filtered at the glomerulus; around 65% is reabsorbed in the proximal tubule and a further 25% in the thick ascending limb of the loop of Henle. Little potassium is transported in the early distal tubule but a significant secretory flux of potassium into the urine occurs in the late distal tubule and cortical collecting duct to ensure that the amount removed from the blood is proportional to the ingested load.

The mechanism for potassium secretion in the distal parts of the nephron is shown in Figure 16.3D (p. 431). Movement of potassium from blood to lumen is dependent on active uptake across the basal cell membrane by the Na,K-ATPase, followed by diffusion of potassium through a luminal membrane potassium channel (ROMK) into the tubular fluid. The electrochemical gradient for potassium movement into the lumen is contributed to both by the high intracellular potassium concentration and by the negative luminal potential difference relative to the blood.

A number of factors influence the rate of potassium secretion. Luminal influences include the rate of sodium delivery and fluid flow through the late distal tubule and cortical collecting ducts. This is a major factor responsible for the increased potassium loss that accompanies diuretic treatment. Agents interfering with the generation of the negative luminal potential also impair potassium secretion, and this is the basis of reduced potassium secretion associated with potassium-sparing diuretics such as amiloride. Factors acting on the blood side of this tubular segment include plasma potassium and pH, such that hyperkalaemia and alkalosis both enhance potassium secretion directly. However, the most important factor in the acute and chronic adjustment of potassium secretion to match metabolic potassium load is aldosterone.

As shown in Figure 16.7, a negative feedback relationship exists between the plasma potassium concentration and aldosterone. In addition to its regulation by the renin–angiotensin system (see Fig. 20.19, p. 771), aldosterone is released from the adrenal cortex in direct response to an elevated plasma potassium. Aldosterone then acts on the kidney to enhance potassium secretion, hydrogen secretion and sodium reabsorption, in the late distal tubule and cortical collecting ducts. The resulting increased excretion of potassium maintains plasma potassium within a narrow range (3.3–4.7 mmol/L). Factors that reduce angiotensin II levels may indirectly affect potassium balance by blunting the rise in aldosterone that would otherwise be provoked by hyperkalaemia. This accounts for the increased risk of hyperkalaemia during therapy with ACE inhibitors and related drugs.

Fig. 16.7 Feedback control of plasma potassium concentration.

Hypokalaemia

Aetiology and clinical assessment

Patients with mild hypokalaemia (plasma K 3.0–3.3 mmol/L) are generally asymptomatic, but more profound reductions in plasma potassium often lead to muscular weakness and associated tiredness. Ventricular ectopic beats or more serious arrhythmias may occur and the arrhythmogenic effects of digoxin may be potentiated. Typical ECG changes occur, affecting the T wave in particular (Fig. 16.8). Functional bowel obstruction may occur due to paralytic ileus. Long-standing hypokalaemia causes renal tubular damage (hypokalaemic nephropathy) and interferes with the tubular response to ADH (acquired nephrogenic diabetes insipidus), resulting in polyuria and polydipsia.

The main causes of hypokalaemia and an approach to the differential diagnosis are shown in Figure 16.9.
Redistribution of potassium into cells should be considered, since correction of the factors involved (see above) may be sufficient to correct the plasma concentration. An inadequate intake of potassium can contribute to hypokalaemia but is unlikely to be the only cause, except in extreme cases. Generally, hypokalaemia is indicative of abnormal potassium loss from the body, either through the kidney or the gastrointestinal tract. When there is no obvious clinical clue to which pathway is involved, measurement of urinary potassium may be helpful; if the kidney is the route of potassium loss, the urine potassium is high (>30 mmol/day), whereas if potassium is being lost through the gastrointestinal tract, the kidney retains potassium, resulting in a lower urinary potassium (generally <20 mmol/day). It should be noted, however, that if gastrointestinal fluid loss is also associated with hypovolaemia, activation of the renin-angiotensin-aldosterone system may occur, causing increased loss of potassium in the urine.

Renal causes of hypokalaemia can be divided into those with and those without hypertension. Hypokalaemia in the presence of hypertension may be due to increased aldosterone secretion in Conn’s syndrome (p. 780) or a genetic defect affecting sodium channels in the distal nephron (Liddle’s syndrome). Excessive intake of liquorice or treatment with carbenoxolone may result in a similar clinical picture, due to inhibition of the renal 11βHSD2 enzyme, which inactivates cortisol in peripheral tissues.

If blood pressure is normal or low, hypokalaemia can be classified according to the associated change in acid-base balance. Inherited defects in tubular transport should be suspected when hypokalaemia occurs in association with alkalosis, provided that diuretic use has been excluded. One such disease is Bartter’s syndrome, in which sodium reabsorption in the thick ascending limb of Henle is defective, usually due to a loss-of-function mutation of the NKCC2 transporter. The clinical and biochemical features are similar to chronic treatment with furosemide. In Gitelman’s syndrome there is a loss-of-function mutation affecting the NCCT transporter in the early distal tubule. The clinical and biochemical features are similar to chronic thiazide treatment. Note that while both Bartter’s and Gitelman’s syndromes are characterised by hypokalaemia and hypomagnesaemia, urinary calcium excretion is increased in Bartter’s syndrome but decreased in Gitelman’s syndrome, analogous to the effects of the loop and thiazide diuretics, respectively, on calcium transport (see Box 16.11, p. 436).

If hypokalaemia occurs in the presence of a normal blood pressure and metabolic acidosis, renal tubular acidosis (proximal or ‘classical’ distal) should be suspected (p. 446).

When hypokalaemia is due to potassium wasting through the gastrointestinal tract, the cause is usually obvious clinically. In some cases, when there is occult induction of vomiting, the hypokalaemia is characteristically associated with metabolic alkalosis, due to loss of gastric acid. If, however, potassium loss has occurred through the surreptitious use of aperients, the hypokalaemia is generally associated with metabolic acidosis. In both cases, urinary potassium excretion is low unless there is significant extracellular volume depletion, which can raise urinary potassium levels by stimulating aldosterone production.

**Investigations**

Measurement of plasma electrolytes, bicarbonate, urine potassium and sometimes of plasma calcium and magnesium is usually sufficient to establish the diagnosis. If the diagnosis remains unclear, plasma renin should be
measured. Levels are low in patients with primary hyperaldosteronism (p. 780) and other forms of mineralocorticoid excess, but raised in other causes of hypokalaemia.

The cause of hypokalaemia may remain unclear despite the above investigations when urinary potassium measurements are inconclusive and the history is incomplete or unreliable. Many such cases are associated with metabolic alkalosis, and in this setting the measurement of urine chloride concentration can be helpful. A low urine chloride (< 30 mmol/L) is characteristic of vomiting (spontaneous or self-induced, in which chloride is lost in HCl in the vomit), while a urine chloride > 40 mmol/L suggests diuretic therapy (acute phase) or a tubular disorder such as Bartter’s or Gitelman’s syndrome. Differentiation between occult diuretic use and primary tubular disorders can be achieved by performing a screen of urine for diuretic drugs.

**Management**

Treatment of hypokalaemia involves first determining the cause and then correcting this where possible. If the problem is mainly one of redistribution of potassium into cells, reversal of this (for example, correction of alkalosis) may be sufficient to restore plasma potassium without providing supplements. In most cases, however, some form of potassium replacement will be required. This can generally be achieved with slow-release potassium chloride tablets, but in more acute circumstances intravenous potassium chloride may be necessary. The rate of administration depends on the severity of hypokalaemia and the presence of cardiac or neuromuscular complications, but should generally not exceed 10 mmol of potassium per hour. In patients with severe, life-threatening hypokalaemia, the concentration of potassium in the infused fluid may be increased to 40 mmol/L if a peripheral vein is used, but higher concentrations must be infused into a large ‘central’ vein with continuous cardiac monitoring.

In the less common situation where hypokalaemia occurs in the presence of systemic acidosis, alkaline salts of potassium, such as potassium bicarbonate, can be given by mouth. If magnesium depletion is also present, replacement of magnesium may also be required for hypokalaemia to be corrected since low cell magnesium can enhance the mechanism for tubular potassium secretion, causing ongoing urinary losses. In some circumstances, potassium-sparing diuretics, such as amiloride, can assist in the correction of hypokalaemia, hypomagnesaemia and metabolic alkalosis, especially when loop or thiazide diuretics are the underlying cause.

**Hyperkalaemia**

**Aetiology and clinical assessment**

Hyperkalaemia can present with progressive muscular weakness, but sometimes there are no symptoms until cardiac arrest occurs. The typical ECG changes are shown in Figure 16.8. Peaking of the T wave is an early ECG sign, but widening of the QRS complex presages a dangerous cardiac arrhythmia.

Hyperkalaemia may occur either because of redistribution of potassium between the ICF and ECF or because intake exceeds excretion. It is also important to remember that hyperkalaemia can also be artefactual due to in vitro haemolysis of blood specimens. An approach to defining the underlying cause of hyperkalaemia is shown in Figure 16.10. Redistribution of potassium from the ICF to the ECF may occur in the presence of systemic acidosis, or when the circulating levels of insulin, catecholamines and aldosterone are reduced or when the effects of these hormones are blocked (p. 440). High potassium intake may contribute to hyperkalaemia, but is seldom the only explanation unless renal excretion mechanisms are impaired.

Impaired excretion of potassium into the urine may be associated with a reduced GFR, as in acute kidney injury or chronic kidney disease. Acute kidney injury can be associated with severe hyperkalaemia when there is a concomitant potassium load, such as in rhabdomyolysis or in sepsis, particularly when acidosis is present. In chronic kidney disease, adaptation to moderately elevated plasma potassium levels commonly occurs. However, acute rises in potassium triggered by excessive dietary intake, hypovolaemia or drugs (see below) may occur and destabilise the situation.

Hyperkalaemia can also develop when tubular potassium secretory processes are impaired, even if the GFR is well maintained. In some cases, this is due to low levels of aldosterone, as occurs in Addison’s disease or with ACE inhibitor therapy. Another cause is hyporeninemic hypoaldosteronism where the renin–angiotensin system is inactivated. This condition typically occurs in association with diabetic nephropathy with neuropathy, and is thought to be due to impaired β-adrenergic stimulation of renin release. Other causes include angiotensin receptor antagonists, non-steroidal anti-inflammatory drugs (NSAIDs) and β-blocking drugs. In another group of conditions, tubular potassium secretion is impaired as the result of aldosterone resistance. This can occur in a variety of diseases in which there is inflammation of the tubulointerstitium, such as systemic lupus erythematosus; following renal transplantation; during treatment with potassium-sparing diuretics; and in a number of inherited disorders of tubular transport.

In all conditions of aldosterone deficiency or aldosterone resistance, hyperkalaemia may be associated with acid retention, giving rise to the pattern of hyperkalaemic distal (‘type 4’) renal tubular acidosis (p. 446).

**Investigations**

Measurement of electrolytes, creatinine and bicarbonate, when combined with clinical assessment, usually provides the explanation for hyperkalaemia. In aldosterone deficiency, plasma sodium concentration is characteristically low, although this can occur in many causes of hyperkalaemia. Addison’s disease should be excluded unless there is an obvious alternative diagnosis, as described on page 777.

**Management**

Treatment of hyperkalaemia depends on its severity and the rate of development. In the absence of neuromuscular symptoms or ECG changes, reduction of potassium intake and correction of underlying abnormalities may be sufficient. However, in acute and/or severe hyperkalaemia (plasma K > 6.5–7.0 mmol/L) more urgent measures must be taken (Box 16.17).

If ECG changes are present, the first step should be infusion of 10 mL 10% calcium gluconate to stabilise conductive tissue membranes (calcium has the opposite
Disorders of acid–base balance

The pH of the arterial plasma is normally 7.40, corresponding to a $\text{H}^+$ concentration of 40 nmol/L. An increase in $\text{H}^+$ concentration corresponds to a decrease in pH. Under normal circumstances, $\text{H}^+$ concentrations do not vary outside the range of 36–44 nmol/L (pH 7.44–7.36), but abnormalities of acid–base balance occur in a wide range of diseases.

Functional anatomy and physiology of acid–base homeostasis

A variety of physiological mechanisms maintain pH of the ECF within narrow limits. The first is the action of blood and tissue buffers, of which the most important involves reaction of $\text{H}^+$ ions with bicarbonate to form carbonic acid, which, under the influence of the enzyme carbonic anhydrase (CA), dissociates to form CO$_2$ and water:

$$\text{CO}_2 + \text{H}_2\text{O} \xleftrightarrow{\text{CA}} \text{H}_2\text{CO}_3 \xleftrightarrow{} \text{H}^+ + \text{HCO}_3^-$$

This buffer system is important because bicarbonate is present at relatively high concentration in ECF (21–28 mmol/L), and two of its key components are under physiological control: CO$_2$ by the lungs, and bicarbonate by the kidneys. These relationships are illustrated in Figure 16.11 (a form of the Henderson–Hasselbalch equation).
Serving to reduce changes in the brainstem stimulate ventilatory drive, into the lumen by an H⁺-ATPase. This excreted acid is generated in the tubular cell from the hydration of CO₂ to form carbonic acid, which dissociates into an H⁺ ion secreted luminally, and a bicarbonate ion that passes across the basolateral membrane into the blood. The secreted H⁺ ions contribute to the reabsorption of any residual bicarbonate present in the luminal fluid, but also contribute net acid for removal from the body, bound to a variety of urinary buffers, of which phosphate and ammonia are the most important. Filtered phosphate (HPO₄²⁻) combines with H⁺ in the distal tubular lumen to form dihydrogen phosphate (H₂PO₄⁻), which is excreted in the urine with sodium. Ammonia (NH₃) is generated within tubular cells by deamination of the amino acid glutamine by the enzyme glutaminase. The NH₃ then reacts with secreted H⁺ in the tubular lumen to form ammonium (NH₄⁺), which becomes trapped in the luminal fluid and is excreted with chloride ions.

These two mechanisms remove approximately 1 mmol/kg of hydrogen ions from the body per day, which equates to the non-volatile acid load arising from the metabolism of dietary protein. The slightly alkaline plasma pH of 7.4 (H⁺ 40 nmol/L) that is maintained during health can be accounted for by the kidney’s ability to generate an acidic urine (pH typically 5–6), in which the net daily excess of metabolic acid produced by the body can be excreted.

**Regulation of acid–base balance**

Respiratory compensation for acid–base disturbances can occur quickly. In response to acid accumulation, pH changes in the brainstem stimulate ventilatory drive, serving to reduce PCO₂ and increase pH (p. 653). Conversely, systemic alkalosis leads to inhibition of ventilation, causing a rise in PCO₂ and reduction in pH, although it should be noted that this mechanism has limited capacity to change pH because hypoxia provides an alternative stimulus to drive ventilation.

The kidney provides a third line of defence against disturbances of arterial pH. When acid accumulates due to chronic respiratory or metabolic (non-renal) causes, the kidney has the long-term capacity to enhance urinary excretion of acid, effectively increasing the plasma bicarbonate.

**Renal control of acid–base balance**

Regulation of acid–base balance occurs at several sites in the kidney. The proximal tubule reabsors some 85% of the filtered bicarbonate ions, through the mechanism for H⁺ secretion illustrated in Figure 16.3A (p. 431). This is dependent on the enzyme carbonic anhydrase both in the cytoplasm of the proximal tubular cells and on the luminal surface of the brush border membranes. The system has a high capacity but does not lead to significant acidification of the luminal fluid.

Distal nephron segments have an important role in determining net acid excretion by the kidney. In the intercalated cells of the cortical collecting duct and the outer medullary collecting duct cells, acid is secreted into the lumen by an H⁺-ATPase. This excreted acid is generated in the tubular cell from the hydration of CO₂ to form carbonic acid, which dissociates into an H⁺ ion secreted luminally, and a bicarbonate ion that passes across the basolateral membrane into the blood. The secreted H⁺ ions contribute to the reabsorption of any residual bicarbonate present in the luminal fluid, but also contribute net acid for removal from the body.

Patients with disturbances of acid–base balance may present clinically either with the effects of tissue malfunction due to disturbed pH (such as altered cardiac and central nervous system function), or with secondary changes in respiration as a response to the underlying metabolic change (such as Kussmaul respiration during metabolic acidosis). The clinical picture is often dominated by the cause of the acid–base change, such as uncontrolled diabetes mellitus or primary lung disease. Frequently the acid–base disturbance only becomes evident when the venous plasma bicarbonate concentration is noted to be abnormal, or when a full arterial blood gas analysis shows abnormalities in pH, PCO₂ or bicarbonate. The ‘base excess’ or ‘base deficit’ can also be calculated from these data. This is the difference between the patient’s bicarbonate level and the normal bicarbonate, measured in vitro with the PCO₂ adjusted to 5.33 kPa (40 mmHg). Calculation of the base excess or deficit is particularly useful in patients with combined respiratory and metabolic disorders (p. 187).

The most common patterns of abnormality in blood gas parameters are shown in Box 16.18. (Note that the terms acidosis and alkalaemia strictly refer to the underlying direction of the acid–base change, while acidaemia and alkalaemia more correctly refer to the net change present in the blood). Interpretation of arterial blood gases is also described on page 653.

In metabolic disturbances, respiratory compensation is almost immediate, so that the predicted compensatory change in PCO₂ is achieved soon after the onset of the metabolic disturbance. In respiratory disorders, on the other hand, a small initial change in bicarbonate occurs as a result of chemical buffering of CO₂ largely within red blood cells, but over days and weeks the kidney achieves further compensatory changes in bicarbonate concentration as a result of long-term adjustments in acid secretory capacity. When the clinically obtained acid–base parameters do not accord with the predicted compensation shown, a mixed acid–base disturbance should be suspected (p. 447).
Metabolic acidosis

Aetiology and assessment

Metabolic acidosis occurs when an acid other than carbonic acid (due to CO₂ retention) accumulates in the body, resulting in a fall in the plasma bicarbonate. The pH fall that would otherwise occur is blunted by hyperventilation, resulting in a reduced PCO₂. If the kidneys are intact and the primary cause of acidosis is not renal in origin, the kidney can gradually increase acid secretion over days to weeks and restore a new steady state.

Two patterns of metabolic acidosis are recognised (Box 16.19), depending on the nature of the accumulating acid:

• In pattern A, when a mineral acid such as hydrochloric acid accumulates, or when there is a primary loss of bicarbonate buffer from the ECF, there is no addition to the plasma of a new acidic anion. In this case, the ‘anion gap’ (calculated as the difference between the main measured cations (Na⁺ + K⁺) and the anions (Cl⁻ + HCO₃⁻)) is normal, since the plasma chloride increases to replace the depleted bicarbonate levels. This ‘gap’ is normally around 12–16 mmol/L (12–16 meq/L) and is made up of anions such as phosphate, sulphate and multiple negative charges on plasma protein molecules. Normal anion gap metabolic acidosis (pattern A) is usually due either to bicarbonate loss in diarrhoea, where the clinical diagnosis is generally obvious, or to renal tubular acidosis (see below).

• In pattern B, an accumulating acid is accompanied by its corresponding anion, which adds to the unmeasured anion gap, while the chloride concentration remains normal. The cause is usually apparent from associated clinical features such as uncontrolled diabetes mellitus, renal failure or shock, or may be suggested by associated symptoms, such as visual complaints in methanol poisoning (p. 222). It is noteworthy that a number of causes of increased anion gap acidosis are associated with alcoholism, including starvation ketosis, lactic acidosis and intoxication by methanol or ethylene glycol.

Lactic acidosis

The diagnosis of lactic acidosis can be confirmed by the measurement of plasma lactate, which is increased over the normal maximal level of 2 mmol/L (20 mg/dL) by as much as tenfold. Two types of lactic acidosis have been defined:

• type 1, due to tissue hypoxia and peripheral generation of lactate, as in patients with circulatory failure and shock
• type 2, due to impaired metabolism of lactate, as in liver disease or by a number of drugs and toxins, including metformin, which inhibit lactate metabolism (p. 823).
Renal tubular acidosis

Renal tubular acidosis (RTA) should be suspected when there is a hyperchloremic acidosis with a normal anion gap and no evidence of gastrointestinal disturbance. The urine pH is inappropriately high (> 5.5) in the presence of systemic acidosis. RTA can be caused by a defect in one of three processes: impaired bicarbonate reabsorption in the proximal tubule (proximal RTA); impaired acid secretion in the late distal tubule or cortical collecting duct intercalated cells (classical distal RTA); or impaired sodium reabsorption in the late distal tubule or cortical collecting duct, which is associated with reduced secretion of both potassium and H⁺ ions (hyperkalaemic distal RTA).

Various subtypes of RTA are recognised and the most common causes are shown in Box 16.20. The inherited forms of RTA are due to mutations in the genes that regulate acid or bicarbonate transport in the renal tubules (see Fig. 16.3, p. 431). However, RTA is often an acquired disorder and in these circumstances the metabolic acidosis may serve as an early clue to the underlying diagnosis.

Sometimes distal RTA is ‘incomplete’ and the plasma bicarbonate concentration is normal under resting conditions. However, in incomplete distal RTA the urine pH fails to fall below 5.3 after an acid challenge test, involving the ingestion of ammonium chloride sufficient to lower the urine pH.

A number of features allow differentiation of types of RTA. Proximal RTA is frequently associated with urinary wasting of amino acids, phosphate and glucose (Fanconi’s syndrome), as well as bicarbonate and potassium. Patients with this disorder can lower the urine pH by treatment with loop diuretics or thiazides.

Management

The first step in management of metabolic acidosis is to identify and correct the underlying cause when possible (see Box 16.19). This may involve controlling diarrhoea, treating diabetes mellitus, correcting shock, stopping drugs that might cause the condition, or using dialysis to remove toxins. Since metabolic acidosis is frequently associated with sodium and water depletion, resuscitation with intravenous fluids is often needed. Use of intravenous bicarbonate in this setting is controversial. Because rapid correction of acidosis can induce hypokalaemia or a fall in plasma ionised calcium, the use of bicarbonate infusions is best reserved for situations where the underlying disorder cannot be readily corrected and acidosis is severe (H⁺ > 100 mmol/L, pH < 7.00) or associated with evidence of tissue dysfunction.

The acidosis in RTA can sometimes be controlled by treating the underlying cause (see Box 16.20), but supplements of sodium and potassium bicarbonate are usually also necessary in types 1 and 2 RTA to achieve a target plasma bicarbonate level of above 18 mmol/L and normokalaemia. In type 4 RTA, loop diuretics, thiazides or fludrocortisone (as appropriate to the underlying diagnosis) may be effective in correcting the acidosis and the hyperkalaemia.

Metabolic alkalosis

Aetiology and clinical assessment

Metabolic alkalosis is characterised by an increase in the plasma bicarbonate concentration and the plasma pH (see Box 16.18). There is a compensatory rise in PCO₂ due to hypoventilation but this is limited by the need to avoid hypoxia. The causes are best classified by the accompanying changes in ECF volume.

Hypovolaemic metabolic alkalosis is the most common pattern. This can be caused by sustained vomiting, in which acid-rich fluid is lost directly from the body, or by treatment with loop diuretics or thiazides. In the case of sustained vomiting, loss of gastric acid is the immediate cause of the alkalosis, but several factors act to sustain or amplify this in the context of volume depletion (Fig. 16.12). Loss of sodium and fluid leads to hypovolaemia and secondary hyperaldosteronism, triggering proximal sodium bicarbonate reabsorption and additional acid secretion by the distal tubule. Hypokalaemia occurs due to potassium loss in the vomitus and by the kidney as the result of secondary hyperaldosteronism, and itself is a stimulus to acid secretion. Additionally, the compensatory rise in PCO₂ further enhances tubular acid secretion. The net result is sustained metabolic alkalosis with an inappropriately acid urine, which cannot be corrected until the deficit in circulating volume has been replaced.

Normovolaemic (or hypervolaemic) metabolic alkalosis occurs when bicarbonate retention and volume expansion occur simultaneously. Classical causes include primary hyperaldosteronism (Conn’s syndrome, p. 780), Cushing’s syndrome (p. 773) and corticosteroid therapy (p. 776). Occasionally, overuse of antacid salts for treatment of dyspepsia produces a similar pattern.
CO₂ accumulation may itself lead to drowsiness that further depresses respiratory drive. Management involves correction of causative factors where possible, but ultimately ventilatory support may be necessary.

**Respiratory alkalosis**

Respiratory alkalosis develops when there is a period of sustained hyperventilation, resulting in a reduction of PCO₂ and increase in plasma pH. If the condition is sustained, renal compensation occurs, such that tubular acid secretion is reduced and the plasma bicarbonate falls.

Respiratory alkalosis is usually of short duration, occurring in anxiety states or as the result of over-vigorous assisted ventilation. It can be prolonged in the context of pregnancy, pulmonary embolism, chronic liver disease, and ingestion of certain drugs such as salicylates that directly stimulate the respiratory centre in the brainstem.

Clinical features are those of the underlying cause but agitation associated with perioral and digital tingling may also occur, due to a reduction in ionised calcium concentrations because of increased binding of calcium to albumin as the result of the alkalosis. In severe cases, Trousseau’s sign and Chvostek’s sign may be positive, and tetany or seizures may develop (p. 768).

Management involves correction of identifiable causes, reduction of anxiety, and a period of rebreathing into a closed bag to allow CO₂ levels to rise.

**Mixed acid–base disorders**

It is not uncommon for more than one disturbance of acid–base metabolism to be present at the same time in the same patient: for example, respiratory acidosis due to narcotic overdose with metabolic alkalosis due to vomiting. In these situations, the arterial pH will represent the net effect of all primary and compensatory changes. Indeed, the pH may be normal, but the presence of underlying acid–base disturbances can be gauged from concomitant abnormalities in the PCO₂ and bicarbonate concentration.

In assessing these disorders, all clinical influences on the patient’s acid–base status should be identified, and reference should be made to the table of predicted compensation given in Box 16.18. If the compensatory change is discrepant from the rules of thumb provided, more than one disturbance of acid–base metabolism may be suspected.

**DISORDERS OF DIVALENT ION METABOLISM**

The present section excludes discussion of calcium disorders, which are considered in Chapters 20 (pp. 767–770) and 25 (p. 1125).

**Functional anatomy and physiology of magnesium metabolism**

Like potassium, magnesium is mainly an intracellular cation. It is important to the function of many enzymes,
including the Na,K-ATPase, and can regulate both potassium and calcium channels. Its overall effect is to stabilise excitable cell membranes.

Renal handling of magnesium involves filtration of free plasma magnesium at the glomerulus (about 70% of the total) with extensive reabsorption (50–70%) in the loop of Henle, and other parts of the proximal and distal renal tubule. Magnesium reabsorption is also enhanced by parathyroid hormone (PTH).

### Presenting problems in disorders of magnesium metabolism

#### Hypomagnesaemia

**Aetiology and clinical assessment**

Hypomagnesaemia exists when plasma magnesium concentrations are below the reference range of 0.75–1.0 mmol/L (1.5–2.0 meq/L). This is usually a reflection of magnesium depletion (Box 16.21), which can be caused by excessive magnesium loss from the gastrointestinal tract (notably in chronic diarrhoea) or the kidney (during prolonged use of loop diuretics). Excessive alcohol ingestion can cause magnesium depletion through both gut and renal losses. Some inherited tubular transport disorders, such as Gitelman’s syndrome, can also result in urinary magnesium wasting (p. 440).

Hypomagnesaemia is frequently associated with hypocalcaemia, probably because magnesium is required for the normal secretion of PTH in response to a fall in serum calcium, and because hypomagnesaemia causes end-organ resistance to PTH. The clinical features of hypomagnesaemia and hypocalcaemia are similar in that tetany, cardiac arrhythmias (notably torsades de pointes, p. 570), central nervous excitation and seizures, vasoconstriction and hypertension may all occur. Magnesium depletion is also associated (through uncertain mechanisms) with hyponatraemia and hypokalaemia, which may contribute to some of the clinical manifestations.

#### Management

The underlying cause should be identified and treated where possible. When symptoms are present, the treatment of choice is intravenous magnesium chloride at a rate not exceeding 0.5 mmol/kg in the first 24 hours. When intravenous access is not available, magnesium sulphate can be given intramuscularly. Oral magnesium salts have limited effectiveness due to poor absorption and may cause diarrhoea. If hypomagnesaemia is due to diuretic treatment, adjunctive use of a potassium-sparing agent can also help by reducing magnesium loss into the urine.

#### Hypermagnesaemia

This is a much less common abnormality than hypomagnesaemia. Predisposing conditions include acute kidney injury and chronic kidney disease, and adrenocortical insufficiency. The condition is generally precipitated in patients at risk by an increased intake of magnesium, or through the use of magnesium-containing medications, such as antacids, laxatives and enemas.

Clinical features include bradycardia, hypotension, reduced consciousness and respiratory depression.

Management involves ceasing all magnesium-containing drugs and reducing dietary magnesium intake, improving renal function if possible, and promoting urinary magnesium excretion using a loop diuretic with intravenous hydration, if residual renal function allows. Calcium gluconate may be given intravenously to ameliorate cardiac effects. Dialysis may be necessary in patients with poor renal function.

#### Functional anatomy and physiology of phosphate metabolism

Inorganic phosphate (mainly present as HPO₄²⁻) is intimately involved in cell energy metabolism, intracellular signalling and bone and mineral balance (Ch. 23). The normal plasma concentration is 0.8–1.4 mmol/L (2.48–4.34 mg/dL). It is freely filtered at the glomerulus and approximately 65% is reabsorbed by the proximal tubule, via an apical sodium–phosphate co-transport carrier. A further 10–20% is reabsorbed in the distal tubules, leaving a fractional excretion of some 10% to pass into the urine, usually as H₂PO₄⁻. Proximal reabsorption is decreased by PTH, fibroblast growth factor 23 (FGF-23), volume expansion, osmotic diuretics and glucose infusion.

#### Presenting problems in disorders of phosphate metabolism

#### Hypophosphataemia

The causes of hypophosphataemia are shown in Box 16.22. Phosphate may redistribute into cells during
If renal function is normal, intravenous normal saline should be given to promote phosphate excretion. Hyperphosphataemia in patients with renal failure should be treated with dietary phosphate restriction and the use of oral phosphate binders (p. 486).

### DISORDERS OF AMINO ACID METABOLISM

Congenital disorders of amino acid metabolism usually present in the neonatal period and may involve life-long treatment regimens. However, some disorders, particularly those involved in amino acid transport, may not present until later in life.

**Phenylketonuria**

Phenylketonuria (PKU) is inherited as an autosomal recessive disorder caused by loss-of-function mutations in the PAH gene, which encodes phenylalanine hydroxylase, an enzyme required for degradation of phenylalanine. As a result, phenylalanine accumulates at high levels in the neonate’s blood, causing mental retardation.

The diagnosis of PKU is almost always made by routine neonatal screening (p. 64). Treatment involves life-long adherence to a low-phenylalanine diet. Early and adequate dietary treatment prevents major mental retardation, although there may still be a slight reduction in IQ.

**Homocystinuria**

Homocystinuria is an autosomal recessive disorder caused by loss-of-function mutations in the CBS gene, which encodes cystathionine beta-synthase. The enzyme deficiency causes accumulation of homocysteine and methionine in the blood. Many cases of homocystinuria are diagnosed through newborn screening programmes.

Clinical manifestations are wide-ranging and involve the eyes (ectopia lentis – displacement of the lens), central nervous system (mental retardation, delayed developmental milestones, seizures, psychiatric disturbances), skeleton (resembling Marfan’s syndrome, and also with generalised osteoporosis), vascular system (thrombotic lesions of arteries and veins) and skin (hypopigmentation).

Treatment is dietary, involving a methionine-restricted, cystine-supplemented diet, as well as large doses of pyridoxine.

### DISORDERS OF CARBOHYDRATE METABOLISM

The most common disorder of carbohydrate metabolism is diabetes mellitus, which is discussed in Chapter 21. There are also some rare inherited defects.

**Galactosaemia**

Galactosaemia is caused by loss-of-function mutations in the GALT gene, which encodes galactose-1-phosphate uridylyl transferase. It is usually inherited as an autosomal recessive disorder. The neonate is unable to metabolise galactose, one of the hexose sugars contained in lactose.
Vomiting or diarrhoea usually begins within a few days of ingestion of milk, and the neonate may become jaundiced. Failure to thrive is the most common clinical presentation. The classic form of the disease results in hepatomegaly, cataracts and mental retardation, and fulminant infection with *Escherichia coli* is a frequent complication. Treatment involves life-long avoidance of galactose- and lactose-containing foods.

The widespread inclusion of galactosaemia in newborn screening programmes has resulted in the identification of a number of milder variants.

**Glycogen storage diseases**

Glycogen storage diseases (GSD, or glycogenoses) result from an inherited defect in one of the many enzymes responsible for the formation or breakdown of glycogen, a complex carbohydrate that can be broken down quickly to release glucose during exercise or between meals.

There are several major types of GSD, which are classified by a number, by the name of the defective enzyme or eponymously after the physician who first described the condition (Box 16.23). Most forms of GSD are inherited as autosomal recessive disorders.

A diagnosis of GSD is made on the basis of the symptoms, physical examination and results of biochemical tests. Occasionally, a muscle or liver biopsy is required to confirm the enzyme defect. Different types of GSD present at different ages, and some may require life-long modifications of diet and lifestyle.

**DISORDERS OF COMPLEX LIPID METABOLISM**

Complex lipids are key components of the cell membrane (p. 47) that are normally catabolised in organelles called lysosomes. The lysosomal storage diseases are a heterogeneous group of disorders caused by loss-of-function mutations in various lysosomal enzymes (Box 16.24), resulting in an inability to break down complex glycolipids or other intracellular macromolecules. These disorders have diverse clinical manifestations, typically including mental retardation. Some can be treated with enzyme replacement therapy, while others (such as Tay–Sachs disease) can be prevented through community participation in genetic carrier screening programmes (p. 64).

**DISORDERS OF BLOOD LIPIDS AND LIPOPROTEINS**

The three most important classes of lipid are cholesterol, which is composed of hydrocarbon rings; triglycerides (TG), which are esters composed of glycerol linked to three long-chain fatty acids; and phospholipids, which are composed of a hydrophobic ‘tail’ consisting of two long-chain fatty acids linked through glycerol to a hydrophilic head containing a phosphate group. Phospholipids are present in cell membranes and are important signalling molecules.

Despite their poor water solubility, lipids need to be absorbed from the gastrointestinal tract and transported throughout the body. This is achieved by incorporating lipids within lipoproteins. Plasma cholesterol and TG are clinically important because they are major treatable risk factors for cardiovascular disease, whilst severe hypertriglyceridaemia also predisposes to acute pancreatitis.

**Functional anatomy, physiology and investigation of lipid metabolism**

Lipids are transported and metabolised by apolipoproteins, which combine with lipids to form spherical or disc-shaped lipoproteins, consisting of a hydrophobic core and a less hydrophobic coat (Fig. 16.13). The structure of some apolipoproteins also enables them to act as enzyme co-factors or cell receptor ligands. Variations in lipid and apolipoprotein composition result in distinct classes of lipoprotein that perform specific metabolic functions.

**Processing of dietary lipid**

The intestinal absorption of dietary lipid is described on page 841 (see also Fig. 16.14). Enterocytes lining the gut extract monoglyceride and free fatty acids from micelles and re-esterify them into TG, which are combined with a truncated form of apolipoprotein B (Apo B48) as it is synthesised. Intestinal cholesterol derived from dietary

---

**16.23 Glycogen storage diseases**

<table>
<thead>
<tr>
<th>Type</th>
<th>Eponym</th>
<th>Enzyme deficiency</th>
<th>Clinical features and complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Von Gierke</td>
<td>Glucose-6-phosphatase</td>
<td>Childhood presentation, hypoglycaemia, hepatomegaly</td>
</tr>
<tr>
<td>II</td>
<td>Pompe</td>
<td><strong>α</strong>-glucosidase (acid maltase)</td>
<td>Classical presentation in infancy, muscle weakness (may be severe)</td>
</tr>
<tr>
<td>III</td>
<td>Cori</td>
<td>Glycogen debrancher enzyme</td>
<td>Childhood presentation, hepatomegaly, mild hypoglycaemia</td>
</tr>
<tr>
<td>IV</td>
<td>Andersen</td>
<td>Brancher enzyme</td>
<td>Presentation in infancy, severe muscle weakness (may affect heart), cirrhosis</td>
</tr>
<tr>
<td>V</td>
<td>McArdle</td>
<td>Muscle glycogen phosphorylase</td>
<td>Exercise-induced fatigue and myalgia</td>
</tr>
<tr>
<td>VI</td>
<td>Hers</td>
<td>Liver phosphorylase</td>
<td>Mild hepatomegaly</td>
</tr>
<tr>
<td>VII</td>
<td>Tarui</td>
<td>Muscle phosphofructokinase</td>
<td>Exercise-induced fatigue and myalgia</td>
</tr>
<tr>
<td>IX*</td>
<td></td>
<td>Liver phosphorylase kinase</td>
<td>Mild hepatomegaly</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>Hepatic glycogen synthase</td>
<td>Fasting hypoglycaemia, post-prandial hyperglycaemia</td>
</tr>
</tbody>
</table>

*Note that type VIII has been merged into type IX and no longer exists as a separate entity.*
The main dietary determinants of plasma cholesterol concentrations are the intake of saturated and trans-unsaturated fatty acids, which reduce LDL-receptor activity (see below). Dietary cholesterol has surprisingly little effect on fasting cholesterol levels. Plant sterols and drugs that inhibit cholesterol absorption are effective because they also reduce the re-utilisation of biliary cholesterol. The dietary determinants of plasma TG concentrations are complex since excessive intake of carbohydrate, fat or alcohol may all contribute to increased plasma TG by different mechanisms.

Endogenous lipid synthesis

In the fasting state, the liver is the major source of plasma lipids (see Fig. 16.14). The liver may acquire lipids by uptake, synthesis or conversion from other macronutrients. These lipids are transported to other tissues by secretion of very low-density lipoproteins (VLDL), which are rich in TG but differ from chylomicrons in that they contain full-length Apo B100. Following secretion into the circulation, VLDL undergo metabolic processing similar to that of chylomicrons. Hydrolysis of VLDL TG releases fatty acids to tissues and converts VLDL into ‘remnant’ particles, referred to as intermediate-density lipoproteins (IDL). Most IDL are rapidly cleared by LDL receptors in the liver, but some are processed by hepatic lipase, which converts the particle to an LDL by removing TG and most materials other than Apo B100, and free and esterified cholesterol.

The LDL particles act as a source of cholesterol for cells and tissues (see Fig. 16.14). LDL cholesterol is internalised by receptor-mediated endocytosis via the LDL receptor. Delivery of cholesterol via this pathway down-regulates further expression of the LDL receptor gene and reduces the synthesis and activity of
is esterified by lecithin cholesterol acyl transferase (LCAT), thus maintaining an uptake gradient and remodelling the particle into a mature spherical HDL. These HDL release their cholesterol to the liver and other cholesterol-requiring tissues via the scavenger receptor B1 (SRB1).

The cholesterol ester transfer protein (CETP) in plasma allows transfer of cholesterol from HDL or LDL to VLDL or chylomicrons in exchange for TG. When TG is elevated, the action of CETP may reduce HDL cholesterol and remodel LDL into ‘small, dense’ LDL particles that appear to be more atherogenic in the blood vessel wall. Animal species that lack CETP are resistant to atherosclerosis.

Lipids and cardiovascular disease

Plasma lipoprotein levels are major modifiable risk factors for cardiovascular disease. Increased levels of
atherogenic lipoproteins (especially LDL, but also IDL, lipoprotein(a) and possibly chylomicron remnants) contribute to the development of atherosclerosis (p. 579). Increased plasma concentration and reduced diameter favours subendothelial accumulation of these lipoproteins. Following chemical modifications such as oxidation, Apo B-containing lipoproteins are no longer cleared by normal mechanisms. They trigger a self-perpetuating inflammatory response during which they are taken up by macrophages to form foam cells, a hallmark of atherosclerotic lesions. These processes also have an adverse effect on endothelial function.

Conversely, HDL removes cholesterol from the tissues to the liver, where it is metabolised and excreted in bile. HDL may also counteract some components of the inflammatory response, such as the expression of vascular adhesion molecules by the endothelium. Consequently, low HDL cholesterol levels, which are often associated with TG elevation, also predispose to atherosclerosis.

**Lipid measurement**

Abnormalities of lipid metabolism most commonly come to light following routine blood testing. Measurement of plasma cholesterol alone is not sufficient for comprehensive assessment. Levels of total cholesterol (TC), triglyceride (TG) and HDL cholesterol (HDL-C) need to be obtained after a 12-hour fast to permit accurate calculation of LDL cholesterol (LDL-C) according to the Friedewald formula \( \text{LDL-C} = \text{TC} - \text{HDL-C} - \frac{(\text{TG}/2.2) \times 100}{\text{TC}} \) mmol/L; or LDL-C = \( \frac{\text{TC}}{1.052} \) or LDL-C = \( \frac{\text{TC}}{1.052} - \text{HDL-C} - \frac{(\text{TG}/5) \times 100}{\text{TC}} \). The formula becomes unreliable when TG levels exceed 4 mmol/L (350 mg/dL). Other risk markers, such as NHDL-C (non-HDL-C, calculated as the difference between TC and HDL-C levels) or Apo B, may assess risk of cardiovascular disease more accurately than LDL-C when TG levels are increased. Furthermore, non-fasting samples are often used to guide therapeutic decisions since they are unaffected in terms of TC and measured LDL-C, albeit that they differ from fasting samples in terms of TG, HDL-C and, to some extent, calculated LDL-C. Consideration must be given to confounding factors, such as recent illness, after which cholesterol, LDL and HDL levels temporarily decrease in proportion to severity. Results that will affect major decisions, such as initiation of drug therapy, should be confirmed with a repeat measurement.

Elevated levels of TG are common in obesity, diabetes and insulin resistance (Chs 5 and 21), and are frequently associated with low HDL and increased ‘small, dense’ LDL. Under these circumstances, LDL-C may underestimate risk. This is one situation in which measurement of Apo B may provide additional useful information.

**Presenting problems in disorders of lipids**

Lipid measurements are usually performed for the following reasons:

- screening for primary or secondary prevention of cardiovascular disease
- investigation of patients with clinical features of lipid disorders (Fig. 16.15) and their relatives
- monitoring of response to diet, weight control and medication.

**Aetiology and clinical assessment**

The first step is to consider the effect of other diseases and drugs (Box 16.25). Overt or subclinical hypothyroidism (p. 743) may cause hypercholesterolaemia, and so measurement of thyroid function is warranted in most cases, even in the absence of typical symptoms and signs.

Once secondary causes are excluded, primary lipid abnormalities may be diagnosed. Primary lipid abnormalities can be classified according to the predominant lipid problem: hypercholesterolaemia, hypertriglyceridaemia or mixed hyperlipidaemia (Box 16.26). Although single-gene disorders are encountered in all three categories, most cases are due to multiple-gene (polygenic) loci interacting with environmental factors. Clinical consequences of dyslipidaemia vary somewhat between these causes (see Fig. 16.15).

**16.25 Causes of secondary hyperlipidaemia**

<table>
<thead>
<tr>
<th>Secondary hypercholesterolaemia</th>
<th>Moderate common</th>
<th>Less common</th>
<th>Secondary hypertriglyceridaemia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common</strong></td>
<td>Hypothyroidism</td>
<td>Nephrotic syndrome</td>
<td>Diabetes mellitus (type 2)</td>
</tr>
<tr>
<td></td>
<td>Pregnancy</td>
<td>Anorexia nervosa</td>
<td>Chronic renal disease</td>
</tr>
<tr>
<td></td>
<td>Cholestatic liver disease</td>
<td>Hyperparathyroidism</td>
<td>Abdominal obesity</td>
</tr>
<tr>
<td></td>
<td>Drugs (diuretics, ciclosporin, corticosteroids, androgens, antiretroviral agents)</td>
<td></td>
<td>Excess alcohol</td>
</tr>
</tbody>
</table>

**Predominant hypercholesterolaemia**

Polygenic hypercholesterolaemia is the most common cause of a mild to moderate increase in LDL-C (see Box 16.26). Physical signs, such as corneal arcus and xanthelasma, may be found in this as well as other forms of lipid disturbance (see Fig. 16.15). The risk of cardiovascular disease is proportional to the degree of LDL-C (or Apo B) elevation, but is modified by other major risk factors, particularly low HDL-C.

Familial hypercholesterolaemia (FH) causes moderate to severe hypercholesterolaemia and has a prevalence of at least 0.2% in most populations. It is usually caused by a loss-of-function mutation in the LDL receptor gene, which results in an autosomal dominant pattern of inheritance. A similar syndrome can arise with loss-of-function mutations in the ligand-binding domain of Apo B100 or gain-of-function mutations in the PCSK9 gene. The latter increases the activity of the PCSK9 protein, which is a sterol-sensitive protease that targets the LDL receptor for degradation. Causative mutations can be detected in one of these three genes by genetic testing in about 70% of patients with FH. Most patients with these types of FH have LDL levels that are approximately twice as high as in normal
**Clinical manifestations of hyperlipidaemia.** *Note that xanthelasma and corneal arcus may be non-specific, especially in later life.*

### Classification of hyperlipidaemia

<table>
<thead>
<tr>
<th>Disease</th>
<th>Elevated lipid results</th>
<th>Elevated lipoprotein</th>
<th>CHD risk</th>
<th>Pancreatitis risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predominant hypercholesterolaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygenic (majority)</td>
<td>TC ± TG</td>
<td>LDL ± VLDL</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia (LDL receptor defect, defective Apo B100, increased function of PCSK-9)</td>
<td>TC ± TG</td>
<td>LDL ± VLDL</td>
<td>+++</td>
<td>-</td>
</tr>
<tr>
<td>Hyperalphalipoproteinaemia</td>
<td>TC ± TG</td>
<td>HDL</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Predominant hypertriglyceridaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygenic (majority)</td>
<td>TG</td>
<td>VLDL ± LDL</td>
<td>Variable</td>
<td>+</td>
</tr>
<tr>
<td>Lipoprotein lipase deficiency</td>
<td>TG &gt;&gt; TC</td>
<td>Chylo</td>
<td>?</td>
<td>+++</td>
</tr>
<tr>
<td>Familial hypertriglyceridaemia</td>
<td>TG &gt; TC</td>
<td>VLDL ± Chylo</td>
<td>?</td>
<td>++</td>
</tr>
<tr>
<td>Mixed hyperlipidaemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polygenic (majority)</td>
<td>TC + TG</td>
<td>VLDL + LDL</td>
<td>Variable</td>
<td>+</td>
</tr>
<tr>
<td>Familial combined hyperlipidaemia*</td>
<td>TC and/or TG</td>
<td>LDL and/or VLDL</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Dysbeta lipoproteinaemia*</td>
<td>TC and/or TG</td>
<td>IDL</td>
<td>+++</td>
<td>+</td>
</tr>
</tbody>
</table>

+ = slightly increased risk; ++ = increased risk; +++ = greatly increased risk.

*Familial combined hyperlipidaemia and dysbeta lipoproteinaemia may also present as predominant hypercholesterolaemia or predominant hypertriglyceridaemia.

(CHD = coronary heart disease; Chylo = chylomicrons; TC = total cholesterol; TG = triglycerides)
subjects of the same age and gender. Affected subjects suffer from severe hypercholesterolaemia and premature cardiovascular disease. FH may be accompanied by xanthomas of the Achilles or extensor digitorum tendons (see Fig. 16.15), which are strongly suggestive of FH. The onset of corneal arcus before age 40 is also suggestive of this condition. Identification of an index case of FH (the first case of FH in a family) should trigger genetic and biochemical screening of other family members, which is a cost-effective method for case detection. Affected individuals should be managed from childhood (Box 16.27).

Familial combined hyperlipidaemia, and dysbeta-lipoproteinaemia, may present with the pattern of predominant hypercholesterolaemia (see ‘Mixed hyperlipidaemia’ below).

**Predominant hypertriglyceridaemia**

Polygenic hypertriglyceridaemia is the most common cause of a raised TG level (see Box 16.26). Other common causes include excess alcohol intake, medications (such as β-blockers and retinoids), type 2 diabetes, impaired glucose tolerance, central obesity or other manifestations of insulin resistance (p. 805) and impaired absorption of bile acids. It is often accompanied by post-prandial hypertriglyceridaemia and reduced HDL-C, both of which may contribute to cardiovascular risk. Excessive intake of alcohol or dietary fat, or other exacerbating factors may precipitate a massive increase in TG levels, which, if they exceed 10 mmol/L (880 mg/dL), may pose a risk of acute pancreatitis.

Inherited forms of hypertriglyceridaemia also occur. Loss-of-function mutations in the LPL gene, which encodes lipoprotein lipase, or the APOC2 gene, which encodes the Apo C2 protein that acts as a co-factor for lipoprotein lipase, may cause recessively inherited forms of hypertriglyceridaemia. These mutations cause massive hypertriglyceridaemia that is not readily amenable to drug treatment. It often presents in childhood and is associated with episodes of acute abdominal pain and pancreatitis. In common with other causes of severe hypertriglyceridaemia, it may result in hepatosplenomegaly, lipaemia retinalis and eruptive xanthomas (see Fig. 16.15).

Familial hypertriglyceridaemia may also be inherited in a dominant manner due to mutations in the APOA5 gene, which encodes Apo A5 – a co-factor that is essential for lipoprotein lipase activity. This disorder may be associated with high levels of TG that predispose to cardiovascular disease and pancreatitis.

Familial combined hyperlipidaemia, and dysbeta-lipoproteinaemia, may present with the pattern of predominant hypertriglyceridaemia (see ‘Mixed hyperlipidaemia’, below).

**Mixed hyperlipidaemia**

It is difficult to define quantitatively the distinction between predominant hyperlipidaemias and mixed hyperlipidaemia. The term ‘mixed’ usually implies the presence of hypertriglyceridaemia, as well as an increase in LDL or IDL. Treatment of massive hypertriglyceridaemia may improve TG faster than cholesterol, thus temporarily mimicking mixed hyperlipidaemia.

Primary mixed hyperlipidaemia is usually polygenic and, like predominant hypertriglyceridaemia, often occurs in association with type 2 diabetes, impaired glucose tolerance, central obesity or other manifestations of insulin resistance (p. 805). Both components of mixed hyperlipidaemia may contribute to the risk of cardiovascular disease.

Familial combined hyperlipidaemia is a term used to identify an inherited tendency towards the overproduction of atherogenic Apo B-containing lipoproteins. It results in elevation of cholesterol, TG or both in different family members at different times. It is associated with an increased risk of cardiovascular disease but it does not produce any pathognomonic physical signs. In practice, this relatively common condition is substantially modified by factors such as age and weight. It may not be a monogenic condition, but rather one end of a heterogeneous spectrum that overlaps insulin resistance (p. 805).

Dysbeta-lipoproteinaemia (also referred to as type 3 hyperlipidaemia, broad-beta dyslipoproteinaemia or remnant hyperlipidaemia) involves accumulation of roughly equimolar levels of cholesterol and TG. It is caused by homozygous inheritance of the Apo E2 allele, which is the isof orm least avidly recognised by the LDL receptor. In conjunction with other exacerbating factors, such as obesity and diabetes, it leads to accumulation of atherogenic IDL and chylomicron remnants. Premature cardiovascular disease is common and it may also result in the formation of palmar xanthomas, tuberous xanthomas or tendon xanthomas.

**Rare dyslipidaemias**

Several rare disturbances of lipid metabolism have been described (Box 16.28). They provide important insights into lipid metabolism and its impact on risk of cardiovascular disease.

Fish eye disease, Apo A1 Milano and LCAT deficiency demonstrate that very low HDL levels do not necessarily cause cardiovascular disease, but Apo A1 deficiency, and possibly Tangier disease, demonstrate that low HDL-C can be atherogenic under some circumstances. Autosomal recessive FH and PCSK9 gain-of-function mutations reveal the importance of proteins that chaperone the LDL receptor. Sitosterolaemia and cerebrotendinous xanthomatosis demonstrate that sterols other than cholesterol can cause xanthomas and cardiovascular disease, while PCSK9 loss-of-function mutations, abetalipoproteinaemia and hypobetalipoproteinaemia suggest that low levels of Apo B-containing lipoproteins reduce the risk of cardiovascular disease. The only adverse health outcomes associated with
16.28 Miscellaneous and rare forms of hyperlipidaemia

<table>
<thead>
<tr>
<th>Condition</th>
<th>Lipoprotein pattern</th>
<th>CVD risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangier disease</td>
<td>Very low HDL, low TC</td>
<td>+</td>
</tr>
<tr>
<td>Apo A1 deficiency</td>
<td>Very low HDL</td>
<td>++</td>
</tr>
<tr>
<td>Apo A1 Milano</td>
<td>Very low HDL</td>
<td>-</td>
</tr>
<tr>
<td>Fish eye disease</td>
<td>Very low HDL, high TG</td>
<td>-</td>
</tr>
<tr>
<td>LCAT deficiency</td>
<td>Very low HDL, high TG</td>
<td>?</td>
</tr>
<tr>
<td>Autosomal recessive FH</td>
<td>Very high LDL</td>
<td>++</td>
</tr>
<tr>
<td>Sitosterolaemia</td>
<td>High plant sterols including sitosterol</td>
<td>+</td>
</tr>
<tr>
<td>Cerebrotendinous xanthomatosis</td>
<td>Bile acid defect (cholesterol accumulation)</td>
<td>+</td>
</tr>
</tbody>
</table>

+ = slightly increased risk; ++ = increased risk.
(CVD = cardiovascular disease; FH = familial hypercholesterolaemia; HDL = high-density lipoprotein; LCAT = lecithin cholesterol acyl transferase; TC = total cholesterol; TG = triglycerides)

extremely low plasma lipid levels in the latter two conditions are attributable to fat-soluble vitamin deficiency, or impaired transport of lipid from intestine or liver.

**Management of dyslipidaemia**

Lipid-lowering therapies have a key role in the secondary and primary prevention of cardiovascular diseases (p. 581). Assessment of absolute risk, treatment of all modifiable risk factors and optimisation of lifestyle, especially diet and exercise, are central to management in all cases.

Patients with the greatest absolute risk of cardiovascular disease will derive the greatest absolute benefit from treatment. Public health organisations recommend thresholds for the introduction of lipid-lowering therapy based on the identification of patients in very high-risk categories, or those calculated to be at high absolute risk according to algorithms or tables such as the Joint British Societies Coronary Risk Prediction Chart (see Fig. 18.62, p. 582). These tables, which are based on large epidemiological studies, should be recalibrated for the local population, if possible. In general, patients who have cardiovascular disease, diabetes mellitus, chronic renal impairment, familial hypercholesterolaemia or an absolute risk of cardiovascular disease of greater than 20% in the ensuing 10 years are arbitrarily regarded as having sufficient risk to justify drug treatment.

Public health organisations also recommend target levels for patients receiving drug treatment. High-risk patients should aim for HDL-C > 1 mmol/L (38 mg/dL) and fasting TG < 2 mmol/L (approximately 180 mg/dL), whilst target levels for LDL-C have been reduced from 2.5 to 2.0 mmol/L (76 mg/dL) or less. In general, total cholesterol should be < 5 mmol/L (190 mg/dL) during treatment, and < 4 mmol/L (approximately 150 mg/dL) in high-risk patients and in secondary prevention of cardiovascular disease.

**Non-pharmacological management**

Patients with lipid abnormalities should receive medical advice and, if necessary, dietary counselling to:

- reduce intake of saturated and trans-unsaturated fat to less than 7–10% of total energy
- reduce intake of cholesterol to < 250 mg/day
- replace sources of saturated fat and cholesterol with alternative foods, such as lean meat, low-fat dairy products, polyunsaturated spreads and low glycaemic index carbohydrates
- reduce energy-dense foods such as fats and soft drinks, whilst increasing activity and exercise to maintain or lose weight
- increase consumption of cardioprotective and nutrient-dense foods, such as vegetables, unrefined carbohydrates, fish, pulses, nuts, legumes, fruit, etc.
- adjust alcohol consumption, reducing intake if excessive or if associated with hypertension, hypertriglyceridaemia or central obesity
- achieve additional benefits with supplementary intake of foods containing lipid-lowering nutrients such as n-3 fatty acids, dietary fibre and plant sterols.

The response to diet is usually apparent within 3–4 weeks but dietary adjustment may need to be introduced gradually. Although hyperlipidaemia in general, and hypertriglyceridaemia in particular, can be very responsive to these measures, LDL-C reductions are often only modest in routine clinical practice. Explanation, encouragement and persistence are often required to induce patient compliance. Even minor weight loss can substantially reduce cardiovascular risk, especially in centrally obese patients (p. 116).

All other modifiable cardiovascular risk factors should be assessed and treated. If possible, intercurrent drug treatments that adversely affect the lipid profile should be replaced.

**Pharmacological management**

The main diagnostic categories provide a useful framework for management and the selection of first-line pharmacological treatment (Fig. 16.16).

**Predominant hypercholesterolaemia**

Predominant hypercholesterolaemia can be treated with one or more of the cholesterol-lowering drugs: **Statins**. These reduce cholesterol synthesis by inhibiting the HMGCoA reductase enzyme. The reduction in cholesterol synthesis up-regulates activity of the LDL receptor, which increases clearance of LDL and its precursor, IDL, resulting in a secondary reduction in LDL synthesis. Statins reduce LDL-C by up to 60%, reduce TG by up to 40% and increase HDL-C by up to 10%. They also reduce the concentration of intermediate metabolites such as isoprenes, which may lead to other effects such as suppression of the inflammatory response. There is clear evidence of protection against total and coronary mortality, stroke and cardiovascular events across the spectrum of CVD risk (Box 16.29).

Statins are generally well tolerated and serious side-effects are rare (well below 2%). Liver function test abnormalities and muscle problems, such as myalgia, asymptomatic increase in creatine kinase (CK), myositis and, infrequently, rhabdomyolysis, are the most common. Side-effects are more likely in patients who are elderly, debilitated or receiving other drugs that interfere with statin degradation, which usually involves cytochrome P450 3A4 or glucuronidation.
Cholesterol absorption inhibitors. The only licensed drug in this class is ezetimibe, which inhibits activity of the intestinal mucosal transporter NPC1L1 that absorbs dietary and biliary cholesterol. Depletion of hepatic cholesterol up-regulates hepatic LDL receptor activity. This mechanism of action is synergistic with the effect of statins. Monotherapy with the standard 10 mg/day dose reduces LDL-C by 15–20%. Slightly greater (17–25%) incremental LDL-C reduction occurs when ezetimibe is added to statins. Ezetimibe is well tolerated, and evidence of a beneficial effect on cardiovascular disease endpoints may be inferred from a trial of combination therapy with simvastatin. Plant sterol-supplemented foods, which also reduce cholesterol absorption, lower LDL-C by 7–15%.

**Bile acid sequestering resins.** Drugs in this class include colestyramine, colestipol and colesevelam. These prevent the reabsorption of bile acids, thereby increasing de novo bile acid synthesis from hepatic cholesterol. As with ezetimibe, the resultant depletion of hepatic cholesterol up-regulates LDL receptor activity and reduces LDL-C in a manner that is synergistic with the action of statins. Resins reduce LDL-C and modestly increase HDL-C, but may increase TG. They are safe but may interfere with bioavailability of other drugs.

Colestevlam has fewer gastrointestinal effects than older preparations that are less well tolerated. Development of specific inhibitors of the intestinal bile acid transporter may further improve tolerability of this class of agent.

**Nicotinic acid.** Pharmacological doses reduce peripheral fatty acid release with the result that VLDL and LDL decline whilst HDL-C increases. Randomised clinical trials have been inconsistent regarding effects on atherosclerosis and cardiovascular events. Side-effects include flushing, gastric irritation, liver function disturbances, and exacerbation of gout and hyperglycaemia. Slow-release formulations and low-dose aspirin may reduce flushing. Combination therapy with the prostaglandin D2 receptor inhibitor laropiprant to reduce flushing further is being evaluated.

**Combination therapy.** In many patients, treatment of predominant hypercholesterolaemia can be achieved by diet plus the use of a statin in sufficient doses to achieve target LDL-C levels. Patients who do not reach LDL targets on the highest tolerated statin dose, or who are intolerant of statins, may receive ezetimibe, plant sterols, nicotinic acid or resins. Ezetimibe and resins are safe and effective in combination with a statin, but nicotinic acid with a statin requires a greater level of caution because the risk of side-effects is slightly increased.

**Predominant hypertriglyceridaemia**
Predominant hypertriglyceridaemia can be treated with one of the TG-lowering drugs (see Fig. 16.16). **Fibrates.** These stimulate peroxisome proliferator-activated receptor (PPAR) alpha, which controls the expression of gene products that mediate the metabolism of TG and HDL. As a result, synthesis of fatty acids, TG and VLDL is reduced, whilst that of lipoprotein lipase, which catabolises TG, is enhanced. In addition, production of Apo A1 and ATP binding cassette A1 is up-regulated, leading to increased reverse cholesterol transport via HDL. Consequently, fibrates reduce TG by up to 50%.

### Fig. 16.16 Flow chart for the drug treatment of hyperlipidaemia.
*Interrupt treatment if CK is more than 5–10 times the upper limit of normal, or if elevated with muscle symptoms, or if ALT is more than 2–3 times the upper limit. To convert TG in mmol/L to mg/dL, multiply by 88. To convert LDL-C in mmol/L to mg/dL, multiply by 38. (ALT = alanine aminotransferase; CK = creatine kinase)*

<table>
<thead>
<tr>
<th>Predominant hypercholesterolaemia</th>
<th>Mixed hyperlipidaemia</th>
<th>Predominant hypertriglyceridaemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-line treatment with statin</td>
<td>Combination treatment likely if TG and LDL &gt; 4 mmol/L</td>
<td>First-line treatment with fibrate</td>
</tr>
<tr>
<td>Ezetimibe if intolerant</td>
<td>Statin + fish oil</td>
<td>Fish oil if intolerant</td>
</tr>
<tr>
<td><strong>Monitor lipids,</strong> Monitor for side-effects*</td>
<td>Fibrate + ezetimibe</td>
<td>Monitor lipids,** Monitor for side-effects*</td>
</tr>
<tr>
<td><strong>Titrate dose</strong></td>
<td>Statin + nicotinate</td>
<td><strong>Titrate fish oil</strong></td>
</tr>
<tr>
<td><strong>Statin</strong></td>
<td>Statin + fibrate</td>
<td><strong>Fibrate</strong></td>
</tr>
<tr>
<td><strong>ezetimibe</strong></td>
<td><strong>Fish oil</strong></td>
<td><strong>ezestimibe</strong></td>
</tr>
<tr>
<td>** resin**</td>
<td><strong>Fish oil</strong></td>
<td><strong>ezestimibe</strong></td>
</tr>
<tr>
<td><strong>nicotinate</strong></td>
<td><strong>Fish oil</strong></td>
<td><strong>ezestimibe</strong></td>
</tr>
<tr>
<td>if resistant</td>
<td>if resistant</td>
<td>if resistant</td>
</tr>
</tbody>
</table>

### EBM 16.29 Benefits of using statins to treat patients with hypercholesterolaemia
*Meta-analysis of major randomised controlled clinical trials involving over 130 000 subjects receiving statins for approximately 5 years showed a reduced risk of a major cardiovascular event of 21% (95% confidence interval 19–23%) per 1 mmol/L reduction in LDL-C, irrespective of age, gender, baseline LDL-C or intercurrent cardiovascular disease. Relative risk reduction was equally favourable in the subjects with the lowest risk.*

and increase HDL-C by up to 20%, but LDL-C changes are variable.

Fewer large-scale trials have been conducted with fibrates than with statins and the results are less conclusive, but reduced rates of cardiovascular disease have been reported with fibrate therapy in the subgroup of patients with low HDL-C levels and elevated TG (e.g. TG > 2.3 mmol/L (200 mg/dL)). Fibrates are usually well tolerated but share a similar side-effect profile to statins. In addition, they may increase the risk of cholelithiasis and prolong the action of anticoagulants. Accumulating evidence suggests that they may also have a protective effect against diabetic microvascular complications.

**Highly polyunsaturated long-chain n-3 fatty acids.**

These include eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), which comprise approximately 30% of the fatty acids in fish oil. EPA and DHA are potent inhibitors of VLDL TG formation. Intakes of greater than 2 g n-3 fatty acid (equivalent to 6 g of most forms of fish oil) per day lower TG in a dose-dependent fashion. Up to 50% reduction in TG may be achieved with 15 g fish oil per day. Changes in LDL-C and HDL-C are variable. Fish oil fatty acids have also been shown to inhibit platelet aggregation and improve cardiac arrhythmia in animal models. Dietary and pharmacological trials suggest that n-3 fatty acids may reduce mortality from coronary heart disease. Fish oils appear to be safe and well tolerated.

Patients with predominant hypertriglyceridaemia who do not respond to lifestyle intervention can be treated with fibrates, fish oil or nicotinic acid, depending on individual response and tolerance. If target levels are not achieved, the fibrates and fish oil or nicotinic acid can be combined. Massive hypertriglyceridaemia may require more aggressive limitation of dietary fat intake (<10–20% energy as fat). Any degree of insulin deficiency should be corrected because insulin is required for optimal activity of lipoprotein lipase. The initial target for patients with massive hypertriglyceridaemia is TG < 10 mmol/L (880 mg/dL), to reduce the risk of acute pancreatitis.

**Mixed hyperlipidaemia**

Mixed hyperlipidaemia can be difficult to treat. Statins alone are less effective first-line therapy once fasting TG exceeds around 4 mmol/L (350 mg/dL). Fibrates are first-line therapy for dysbetalipoproteinaemia, but they may not control the cholesterol component in other forms of mixed hyperlipidaemia. Combination therapy is often required. Effective combinations include: statin plus fish oil when TG is not too high; fibrate plus ezetimibe; statin plus nicotinic acid; or statin plus fibrate. The risk of myopathy is increased with gemfibrozil, but fenofibrate is relatively safe in this regard.

**Monitoring of therapy**

The effect of drug therapy should be assessed after 6 weeks (12 weeks for fibrates). At this point, it is prudent to review side-effects, lipid response (see target levels above), CK and liver function tests. During longer-term follow-up, compliance with drug treatment, diet and exercise should be assessed, with monitoring of weight, blood pressure and lipid levels. The presence of cardiovascular symptoms or signs should be noted and absolute cardiovascular risk assessed periodically. It is not necessary to perform routine checks of CK and liver function unless symptoms occur, or if statins are used in combination with fibrates, nicotinic acid or other drugs that may interfere with their clearance. If myalgia or weakness occurs in association with CK elevation over 5–10 times the upper limit of normal, or if sustained alanine aminotransferase (ALT) elevation more than 2–3 times the upper limit of normal occurs that is not accounted for by fatty liver (p. 959), treatment should be discontinued and alternative therapy sought.

The principles of the management of dyslipidaemia can be applied broadly, but the objectives of treatment in the elderly (Box 16.30) and the safety of pharmacological therapy in pregnancy (Box 16.31) warrant special consideration.

**DISORDERS OF HAEM SYNTHESIS – THE PORPHYRIAS**

The porphyrias are a group of disorders caused by inherited abnormalities in the haem biosynthetic pathway (Fig. 16.17). Most of the described forms are due to partial enzyme deficiencies with a dominant mode of inheritance. They are commonly classified as either hepatic or erythropoietic, depending on whether the major site of excess porphyrin production is in the liver or in the red cell.

**Box 16.30 Management of hyperlipidaemia in old age**

- **Prevalence of atherosclerotic cardiovascular disease:** greatest in old age.
- **Associated cardiovascular risk:** lipid levels become less predictive, as do other risk factors apart from age itself.
- **Benefit of statin therapy:** maintained up to the age of 80 years but evidence is lacking beyond this.
- **Life expectancy and statin therapy:** lives saved by intervention are associated with shorter life expectancy than in younger patients, and so the impact of statins on quality-adjusted life years is smaller in old age.

**Box 16.31 Dyslipidaemia in pregnancy**

- **Lipid metabolism:** lipid and lipoprotein levels increase during pregnancy. This includes an increase in LDL-C, which resolves post-partum. Remnant dyslipidaemia and hypertriglyceridaemia may be exacerbated during pregnancy.
- **Treatment:** dyslipidaemia is rarely thought to warrant urgent treatment so pharmacological therapy is usually contraindicated when conception or pregnancy is anticipated. Teratogenicity has been reported with systemically absorbed agents, and non-absorbed agents may interfere with nutrient bioavailability.
- **Monitoring:** cardiovascular disease is very unlikely amongst women of child-bearing age, but is possible in women with severe risk factor profiles or familial hypercholesterolaemia, when pre-conception cardiovascular review can be considered to ensure that the patient will be able to withstand the demands of pregnancy and labour.
The porphyrias show a low penetrance in the order of 25%. Environmental factors are important in disease expression in some forms. In the most common of these conditions, porphyria cutanea tarda (PCT), these include alcohol, iron accumulation, exogenous oestrogens and exposure to various chemicals. Many cases are associated with hepatitis C infection and this should always be screened for on presentation.

Clinical features
The clinical features of porphyria fall into two broad categories, photosensitivity and acute neurovisceral syndrome. The enzyme defects responsible for the diseases are shown in Figure 16.17.

Photosensitive skin manifestations, attributable to excess production and accumulation of porphyrins in the skin, cause pain, erythema, bullae, skin erosions, hirsutism and hyperpigmentation, and occur predominantly on areas of the skin that are exposed to sunlight (p. 1260). The skin also becomes sensitive to damage from minimal trauma.

The other pattern of presentation is with an acute neurological syndrome. This presents with acute abdominal pain together with features of autonomic dysfunction such as tachycardia, hypertension and constipation. Neuropsychiatric manifestations, hyponatraemia due to inappropriate ADH release (p. 438), and an acute neuropathy may also occur (p. 1223). The neuropathy is usually motor and may, in severe cases, progress to respiratory failure.

There is no proven explanation for the episodic nature of the attacks in porphyria, which can relapse and remit or follow a prolonged and unremitting course. Sometimes, specific triggers can be identified, such as alcohol, fasting, or drugs such as anticonvulsants, sulphonamides, oestrogen and progesterone. The oral contraceptive pill is a common precipitating factor. In a significant number, no precipitant can be identified.

Diagnosis
The diagnosis of porphyria and classification into the various forms have traditionally relied on the pattern of
the porphyrins and porphyrin precursors found in blood, urine and faeces (Box 16.32). This is a straightforward diagnosis at clinical presentation when the metabolites are significantly elevated, but this is not always the case in asymptomatic individuals.

More recently, measurement of the enzymes that are deficient in the various porphyrias has provided further diagnostic information (for example, PBG deaminase activity in red blood cells to diagnose acute intermittent porphyria). However, there is often considerable overlap between enzyme activities in affected and normal subjects. Furthermore, some of the enzymes occur in the mitochondria, for which it is more difficult to obtain suitable specimens for analysis. All the genes of the haem biosynthetic pathway have now been characterised. This has made it possible to identify affected individuals in families by genetic testing, a significant advance considering that penetrance of porphyria is low.

Metabolite excretory patterns are always grossly abnormal during an acute attack or in the presence of cutaneous manifestations of porphyria and are diagnostic of the particular porphyria. A normal metabolite profile under these circumstances effectively excludes porphyria. Metabolites usually remain abnormal for long periods after an acute attack, and in some individuals never return to normal. The diagnosis is not so straightforward in patients who are in remission, or in asymptomatic individuals with a positive family history. Neurological porphyria rarely manifests before puberty, nor can it be readily diagnosed from metabolite patterns after the menopause. In these circumstances, genetic testing for disease-specific mutations can now clarify the situation.

**Management**

For patients predisposed to neurovisceral attacks, general management includes avoidance of any agents known to precipitate acute porphyria. Specific management includes intravenous glucose, as provision of 5000 kilojoules per day can, in some cases, terminate acute attacks through a reduction in ALA synthetase activity, leading to reduced ALA synthesis. More recently, administration of haem (in various forms such as haematin or haem arginate) has been shown to reduce metabolite excretory rates, relieve pain and accelerate recovery. Cyclical acute attacks in women sometimes respond to suppression of the menstrual cycle using gonadotrophin-releasing hormone analogues. In rare cases with frequent prolonged attacks or attacks intractable to treatment, liver transplantation has been effective.

There are few specific or effective measures to treat the photosensitive manifestations. The primary goal is to avoid sun exposure and skin trauma. Barrier sun creams containing zinc or titanium oxide are the most effective products. New colourless creams containing nanoparticle formulations have improved patient acceptance. Beta-carotene is used in some patients with erythropoietic porphyria with some efficacy. Afamelanotide, a synthetic analogue of alpha-melanocyte stimulating hormone (α-MSH), has also been shown to provide protection in erythropoietic protoporphyria. In porphyria cutanea tarda, a course of venesections to remove iron can result in long-lasting clinical and biochemical remission, especially if exposure to identified precipitants, such as alcohol or oestrogens, is reduced. Alternatively, a prolonged course of low-dose chloroquine therapy is also effective.

**Further information**

**Websites**

http://emedicine.medscape.com The Nephrology link on this site contains a useful compendium of articles.

www.lipidonline.org Summarises management strategies for dyslipidaemia.

www.ncbi.nlm.nih.gov The link to OMIM (Online Mendelian Inheritance in Man) provides updated information on the genetic basis of metabolic disorders.

www.porphyria-europe.com and www.drugs-porphyria.org Excellent resources on drug safety in porphyria.
Kidney and urinary tract disease

Clinical examination of the kidney and urinary tract 462
Functional anatomy and physiology 464
Investigation of renal and urinary tract disease 466
Glomerular filtration rate 466
Urinalysis 468
Blood tests 468
Imaging 468
Renal biopsy 471
Presenting problems in renal and urinary tract disease 471
Dysuria 471
Loin pain 471
Oliguria/anuria 471
Polyuria 472
Nocturia 472
Frequency 472
Urinary incontinence 472
Erectile dysfunction 474
Haematuria 474
Proteinuria and nephrotic syndrome 476
Oedema 478
Hypertension 478
Acute kidney injury 478
Chronic kidney disease 483
Renal replacement therapy 488
Preparing for renal replacement therapy 489
Conservative treatment 490
Haemodialysis 490
Haemofiltration 491
Haemodiafiltration 492
Peritoneal dialysis 492
Renal transplantation 492
Renal vascular diseases 494
Renal artery stenosis 494
Acute renal infarction 495
Diseases of small intrarenal vessels 495
Glomerular diseases 497
Glomerulonephritis 498
Inherited glomerular diseases 502
Tubulo-interstitial diseases 502
Acute interstitial nephritis 502
Chronic interstitial nephritis 503
Reflex nephropathy 504
Papillary necrosis 505
Sickle-cell nephropathy 505
Cystic diseases of the kidney 505
Renal stone disease 507
Isolated defects of tubular function 510
Diseases of the collecting system and ureters 510
Congenital abnormalities 510
Retroperitoneal fibrosis 511
Infections of the urinary tract 511
Benign prostatic enlargement 514
Prostatitis 515
Tumours of the kidney and urinary tract 515
Renal adenocarcinoma 515
Urothelial tumours 516
Prostate cancer 517
Testicular tumours 518
Inherited tumour syndromes affecting the renal tract 519
Renal involvement in systemic conditions 519
Pregnancy and renal disease 520
Kidney disease in adolescence 521
Drugs and the kidney 522
CLINICAL EXAMINATION OF THE KIDNEY AND URINARY TRACT

- **Kidney and urinary tract disease**

1. **Hands**
   - Yellow complexion*
   - Bruising*
   - Excoriation of pruritus*
   - Reduced skin turgor in fluid depletion

2. **Skin**
   - Hypertensive changes
   - 'Brown line' pigmentation of nails

3. **Blood pressure**
   - Often elevated

4. **Jugular venous pressure**
   - Elevated in fluid overload

5. **Fundoscopy**
   - Hypertensive changes

6. **Lungs**
   - Crepitations in fluid overload

7. **Heart**
   - Extra heart sounds in fluid overload
   - Pericardial friction rub*

8. **Abdomen**
   - Enlarged kidneys
   - Local tenderness
   - Renal or other arterial bruits in renal vascular disease
   - Rectal examination — prostate

9. **Genitalia**
   - Phimosis

10. **Sacral oedema**

11. **Ankle oedema**

12. **Peripheral neuropathy***

13. **Urinalysis for blood and protein**

14. **Urine microscopy**

**Observation**
- Tiredness
- Respiratory rate and depth increased in metabolic acidosis
- Pallor*

*Features of advanced chronic kidney disease (see also Fig. 17.13)
Many diseases of the kidney and urinary tract are clinically silent, at least in the early stages. Accordingly, it is common for these conditions to first be detected by routine blood tests or on dipstick testing of the urine. Several important abnormalities can also be picked up on physical examination, however, as summarised below.

1. Blood pressure measurements
2. Blood tests for abnormal creatinine and electrolytes
3. Dipstick testing for protein, blood, nitrate and leucocytes
4. Urine microscopy. A Erythrocytes due to bleeding from lower in the urinary tract (x400). B Dysmorphic erythrocytes due to glomerular inflammation (x400). C Hyaline casts (arrows), in normal urine (x160). D Erythrocytes and a red cell cast (arrow) in glomerulonephritis (x100). Panels A–C are phase contrast images; D is a bright field image.
5. Right kidney
6. Left kidney
7. Percussing for tenderness in renal angle
8. Checking sacrum and ankles for pitting oedema

This chapter describes the disorders of the kidneys and urinary tract that are commonly encountered in routine practice, as well as giving an overview of the highly specialised field of renal replacement therapy. Disorders of renal tubular function, which may cause alterations in electrolyte and acid–base balance, are described in Chapter 16.

FUNCTIONAL ANATOMY AND PHYSIOLOGY

The kidneys

The kidneys play a central role in excretion of many metabolic breakdown products, including ammonia, urea and creatinine from protein, and uric acid from nucleic acids, drugs and toxins. They also regulate fluid and electrolyte balance. This is achieved by making large volumes of an ultrafiltrate of plasma (120 mL/min, 170 L/day) at the glomerulus, and selectively reabsorbing components of this ultrafiltrate at points along the nephron. The rates of filtration and reabsorption are controlled by many hormonal and haemodynamic signals.

The kidneys also regulate acid–base homeostasis, calcium and phosphate homeostasis, vitamin D metabolism and production of red blood cells. They are important in regulating blood pressure. Renin is secreted from the juxtaglomerular apparatus in response to reduced afferent arteriolar pressure, stimulation of sympathetic nerves, and changes in sodium content of fluid in the distal convoluted tubule at the macula densa, and is the first step in the generation of angiotensin II and aldosterone release, which in turn regulate systemic vasoconstriction and extracellular volume.

Each kidney is approximately 11–14 cm in length in healthy adults; they are located retroperitoneally on either side of the aorta and inferior vena cava between the 12th thoracic and 3rd lumbar vertebra (Fig. 17.1A). The right kidney is usually a few centimetres lower because the liver lies above it. Both kidneys rise and descend several centimetres with respiration.

The kidneys have a rich blood supply and receive approximately 20–25% of cardiac output through the renal arteries, which arise from the abdominal aorta. The renal arteries undergo various subdivisions within the kidney, eventually forming interlobular arteries that run through the renal cortex. These eventually give rise to afferent glomerular arterioles that supply individual nephrons, which are the functional units of the kidney. The efferent arteriole, leading from the glomerulus, supplies the distal nephron and medulla in a ‘portal’ circulation (Fig. 17.1B).

The nephron

Healthy kidneys contain approximately 1 million individual nephrons. Each nephron consists of a glomerulus, which is responsible for ultrafiltration of blood, a proximal renal tubule, a loop of Henle, a distal renal tubule and a collecting duct, which together are responsible for selective reabsorption of water and electrolytes that have been filtered at the glomerulus (see Fig. 16.2, p. 430). Under normal circumstances, more than 99% of the 170 litres of glomerular filtrate that is produced each day is reabsorbed in the tubules. The remainder passes through the collecting ducts of multiple nephrons and drains into the renal pelvis and ureters.

The glomerulus

The glomerulus comprises a tightly packed loop of capillaries supplied by an afferent arteriole and drained by an efferent arteriole. It is surrounded by a cup-shaped extension of the proximal tubule termed Bowman’s capsule, which is comprised of epithelial cells. Blood that enters the glomerulus undergoes ultrafiltration across the glomerular basement membrane (GBM), which is formed by fusion of the basement membranes of tubular epithelial and vascular endothelial cells (Fig. 17.1C and D). The glomerular capillary endothelial cells contain pores (fenestrae), through which circulating molecules can pass to reach the underlying GBM. Glomerular epithelial cells (podocytes) have multiple long foot processes which interdigitate with those of the adjacent epithelial cells (Fig. 17.1E). As well as maintaining a selective barrier to filtration, podocytes are involved in regulating turnover of the GBM. Mesangial cells lie in the central region of the glomerulus. They have contractile properties similar to those of vascular smooth muscle cells but also have macrophage-like properties.

Under normal circumstances, the glomerulus is impermeable to proteins the size of albumin (67 kDa) or larger, while proteins of 20 kDa or smaller are filtered freely. The ability of molecules between 20 and 67 kDa to pass through the GBM is variable and depends on the size (smaller molecules are filtered more easily) and charge (positively charged molecules are filtered more easily). Very little lipid is filtered by the glomerulus.

Filtration pressure at the glomerulus is normally maintained at a constant level, in the face of wide variations in systemic blood pressure and cardiac output, by alterations in muscle tone within the afferent and efferent arterioles. This is known as autoregulation. When there is a reduction in renal perfusion pressure, renin is released by specialised smooth muscle cells in the juxtaglomerular apparatus. Renin cleaves angiotensinogen to release angiotensin I, which is further cleaved by angiotensin-converting enzyme (ACE) to produce angiotensin II (Fig. 17.1D). This restores glomerular perfusion pressure in the short term by causing vasoconstriction of the efferent arterioles within the kidney and by inducing systemic vasoconstriction to increase blood pressure and thus renal perfusion pressure. In the longer term, angiotensin II increases plasma volume by stimulating aldosterone release, which enhances sodium reabsorption by the renal tubules (see Fig. 20.17, p. 771).

Renal tubules, loop of Henle and collecting ducts

The proximal renal tubule, loop of Henle, distal renal tubule and collecting ducts are responsible for reabsorption of water, electrolytes and other solutes, as well as regulating acid–base balance, as described in detail on page 430 and in Figure 16.3. They also play a key role in regulating calcium homeostasis by converting 25-hydroxyvitamin D to the active metabolite 1,25-di­hydroyvitamin D (p. 1126). Failure of this process contributes to the pathogenesis of hypoccalcaemia and bone disease which occurs in chronic kidney disease (CKD, p. 483). Fibroblast-like cells that lie in the interstitium of the renal cortex are responsible for production
Fig. 17.1 Functional anatomy of the kidney. A Anatomical relationships of the kidney. B A single nephron. For the functions of different segments, see Figures 16.2 and 16.3 (pp. 430 and 431). C Histology of a normal glomerulus. D Schematic cross-section of a glomerulus, showing five capillary loops, to illustrate structure and show cell types. E Electron micrograph of the filtration barrier. (GBM = glomerular basement membrane)
of erythropoietin, which in turn is required for production of red blood cells. Erythropoietin synthesis is regulated by oxygen tension; anaemia and hypoxia increase production, whereas polycythaemia and hyperoxia inhibit it. Failure of erythropoietin production plays an important role in the pathogenesis of anaemia in CKD.

**Ureters and bladder**

The ureters drain urine from the renal pelvis (Fig. 17.1A) and deliver it to the bladder, a muscular organ that lies anteriorly in the lower part of the pelvis, just behind the pubic bone. The function of the bladder is to store and then release urine during micturition. The bladder is richly innervated. Sympathetic nerves arising from T10-L2 relay in the pelvic ganglia to cause relaxation of the detrusor muscle and contraction of the bladder neck (both via α-adrenoceptors), thereby preventing release of urine from the bladder. The distal sphincter mechanism is innervated by somatic motor fibres from sacral segments S2–4, which reach the sphincter either by the pelvic plexus or via the pudendal nerves. Afferent sensory impulses pass to the cerebral cortex, from where reflex-increased sphincter tone and suppression of detrusor contraction inhibit micturition until it is appropriate. Conversely, parasympathetic nerves arising from S2-4 stimulate detrusor contraction, promoting micturition.

The micturition cycle has a storage (filling) phase and a voiding (micturition) phase. During the filling phase, the high compliance of the detrusor muscle allows the bladder to fill steadily without a rise in intravesical pressure. As bladder volume increases, stretch receptors in its wall cause reflex bladder relaxation and increased sphincter tone. At approximately 75% bladder capacity, there is a desire to void. Voluntary control is now exerted over the desire to void, which disappears temporarily. Compliance of the detrusor allows further increase in capacity until the next desire to void. Just how often this desire needs to be inhibited depends on many factors, not the least of which is finding a suitable place in which to void.

The act of micturition is initiated first by voluntary and then by reflex relaxation of the pelvic floor and distal sphincter mechanism, followed by reflex detrusor contraction. These actions are coordinated by the pontine micturition centre. Intravesical pressure remains greater than urethral pressure until the bladder is empty.

**The prostate gland**

The prostate gland is situated at the base of the bladder, surrounding the proximal urethra. Exocrine glands within the prostate produce fluid, which comprises about 20% of the volume of ejaculated seminal fluid and is rich in zinc and proteolytic enzymes. The remainder of the ejaculate is formed in the seminal vesicles and bulbourethral glands, with spermatozoa arising from the testes.

Smooth muscle fibres within the prostate, which are under sympathetic control, play a role in controlling urine flow through the bulbourethra, and also contract at orgasm to move seminal fluid through ejaculatory ducts into the bulbourethra (emission). Contraction of the bulbocavernous muscle (via a spinal muscle reflex) then ejaculates the semen out of the urethra.

**The penis**

Blood flow into the corpus cavernosum of the penis is controlled by sympathetic nerves from the thoracolumbar plexus, which maintain smooth muscle contraction. In response to afferent input from the glans penis and from higher centres, pelvic splanchnic parasympathetic nerves actively relax the cavernosal smooth muscle via neurotransmitters such as nitric oxide, acetylcholine, vasoactive intestinal polypeptide (VIP) and prostacyclin, with consequent dilatation of the lacunar space. At the same time, draining venules are compressed, trapping blood in the lacunar space with consequent elevation of pressure and erection (tumescence) of the penis.

**INVESTIGATION OF RENAL AND URINARY TRACT DISEASE**

**Glomerular filtration rate**

The glomerular filtration rate (GFR) is the rate at which fluid passes into nephrons after filtration and is a measure of renal function. It is proportionate to body
size and the reference range is usually expressed after correction for body surface area as 120 ± 25 mL/min/1.73 m². The GFR may be measured directly by injecting and measuring the clearance of compounds such as inulin or radiolabelled ethylenediaminetetraacetic acid, which are completely filtered at the glomerulus and are not secreted or reabsorbed by the renal tubules (Box 17.1). However, this is not performed routinely and is usually reserved for special circumstances, such as the assessment of renal function in potential live kidney donors. Instead, GFR is usually assessed in clinical practice by measuring serum levels of endogenously produced compounds that are excreted by the kidney. The most widely used is serum creatinine, which is produced by muscle at a constant rate, is almost completely filtered at the glomerulus, and is not reabsorbed. Although creatinine is secreted to a small degree by the proximal tubule, this is only usually significant in terms of GFR estimation in severe renal impairment, where it accounts for a larger proportion of the creatinine excreted. Accordingly, provided muscle mass remains constant, changes in serum creatinine concentrations closely reflect changes in GFR, although the reference range for creatinine is wide due to the fact that muscle mass varies widely between different individuals (Fig. 17.2). Several methods have been developed with which to estimate GFR from serum creatinine.

17.2 Limitations of eGFR

- It is only an estimate, least reliable at extremes of body composition (malnourished, amputees) and in hospital inpatients (as it was derived from outpatients)
- Confidence intervals are wide (90% of patients will have eGFR within 30% of their measured GFR, and 98% within 50%)
- Values are consistent in individuals, so changes mean more than absolute values
- Creatinine level must be stable over days; eGFR is not valid in assessing acute kidney injury
- It tends to underestimate normal or near-normal function, so slightly low values should not be over-interpreted. Many laboratories report only up to > 60 mL/min/1.73 m² for this reason
- In the elderly, who constitute the majority of those with low eGFR, there is controversy about categorising people as having chronic kidney disease (CKD; Box 17.3) on the basis of eGFR alone, particularly at stage 3A, since there is little evidence of adverse outcomes when eGFR is > 50 unless there is also proteinuria
- eGFR is not valid in under-18s or during pregnancy
- The equation was originally validated in US patients and eGFR for any given creatinine was 21% higher in blacks. Performance in other racial groups is under investigation

17.3 Stages of chronic kidney disease (CKD)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Definition</th>
<th>Description</th>
<th>Prevalence</th>
<th>Clinical presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kidney damage with normal or high GFR (&gt; 90)</td>
<td>Mild CKD</td>
<td>6.5%</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>2</td>
<td>Kidney damage and GFR 60–89</td>
<td>Asymptomatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A</td>
<td>GFR 45–59</td>
<td>Moderate CKD</td>
<td>4.5%</td>
<td>Usually asymptomatic</td>
</tr>
<tr>
<td>3B</td>
<td>GFR 30–44</td>
<td>Anaemia in some patients at 3B</td>
<td>Most are non-progressive or progress very slowly</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>GFR 15–29</td>
<td>Severe CKD</td>
<td>0.4%</td>
<td>First symptoms often at GFR &lt; 20</td>
</tr>
<tr>
<td>5</td>
<td>GFR &lt; 15 or on dialysis</td>
<td>Kidney failure</td>
<td>Significant symptoms and complications usually present</td>
<td>Dialysis initiation varies but usually at GFR &lt; 10</td>
</tr>
</tbody>
</table>

1Stages of CKD 1–5 were originally defined by the US National Kidney Foundation Kidney Disease Quality Outcomes Initiative 2002.
2Kidney damage means pathological abnormalities or markers of damage, including abnormalities in urine tests or imaging studies. Two GFR values 3 mths apart are required to assign a stage. All GFR values are mL/min/1.73 m².
3From the NHANES III study of > 15 000 US adults (Am J Kid Dis 2003; 41:1–12).
4For further information, see page 483.
53A/3B split recommended for UK in 2007/8, plus a suffix P, indicating presence of proteinuria (albumin:creatinine ratio (ACR) > 30 or protein:creatinine ratio (PCR) > 50 mg/mmol), e.g. 3A P, in view of the prognostic importance of proteinuria.
creatinine measurements (see Box 17.1) but the most widely used is the MDRD equation, which is now the accepted standard for assessing estimated GFR (eGFR). Although the eGFR has several limitations (Box 17.2), its routine reporting by laboratories has increased recognition of moderate kidney damage and encouraged early deployment of protective therapies (Box 17.3).

A potentially more accurate assessment of GFR can be obtained by collection of a 24-hour urine sample and relating serum creatinine levels to urinary creatinine excretion (see Box 17.1).

**Urinalysis**

Screening for the presence of blood, protein, glucose, ketones, nitrates and leucocytes and to assess pH and osmolality of urine can be achieved by dipstick testing (p. 474). Urine microscopy (p. 463) or flow cytometry can detect erythrocytes, which are indicative of bleeding from the urogenital tract (anywhere from kidney to tip of penis); dysmorphic erythrocytes, which suggest the presence of nephritis; red cell casts, indicative of glomerular disease; and crystals, which may be observed in patients with renal stone disease. It should be noted that calcium oxalate and urate crystals can sometimes be found in normal urine that has been left to stand, due to crystal formation ex vivo. The presence of leucocytes and bacteria in urine is indicative of renal tract infection. White cell casts are strongly suggestive of pyelonephritis. Urine pH can provide diagnostic information in the assessment of renal tubular acidosis (p. 446). Urine collection over a 24-hour period can be performed to measure excretion of solutes, such as calcium, oxalate and urate, in patients with recurrent renal stone disease (p. 507). Proteinuria can also be measured on 24-hour collections but is usually now quantified by protein/creatinine ratio on spot urine samples.

Other dynamic tests of tubular function, including concentrating ability (p. 794), ability to excrete a water load (p. 438) and ability to excrete acid (p. 426), and calculation of fractional calcium, phosphate or sodium excretion, are valuable in some circumstances. The fractional excretion of these ions can be calculated by the general formula: (urine concentration of analyte × serum creatinine) / (serum concentration of analyte × urinary creatinine). For example, familial benign hypercalcaemia is characterised by a very low fractional excretion of calcium (p. 770), and hypophosphataemic rickets (p. 1128) by an increased fractional excretion of phosphate. Calculation of fractional excretion of sodium (FENa) can help differentiate volume depletion, when the tubules are avidly conserving sodium (FENa typically < 1.0), from acute tubular necrosis, when the tubules are damaged and are less able to conserve sodium (FENa typically > 1.0).

**Blood tests**

**Haematology**

A normochromic normocytic anaemia is common in CKD and is due in part to deficiency of erythropoietin and bone marrow suppression secondary to toxins retained in CKD. Other causes of anaemia include iron deficiency from urinary tract bleeding, and haemolytic anaemia secondary to disorders such as haemolytic uraemic syndrome (HUS) and thrombotic thrombocytopenic purpura (TTP). Other abnormalities may be observed that reflect underlying disease processes, such as neutrophilia and raised erythrocyte sedimentation rate (ESR) in vasculitis or sepsis; lymphopenia and raised ESR in systemic lupus erythematosus (SLE); and fragmented red cells in HUS and TTP.

**Biochemistry**

Abnormalities of routine biochemistry are common in renal disease. Serum levels of creatinine may be raised, reflecting reduced GFR (see above), although serum creatinine values can remain within the reference range in patients with reduced muscle mass, even when the GFR has fallen by more than 50%. Serum levels of urea are often increased in kidney disease but this analyte has limited value as a measure of GFR since levels increase with protein intake, following gastrointestinal haemorrhage and in catabolic states. Conversely, urea levels may be reduced in patients with liver failure or anorexia and in malnourished patients, independently of changes in renal function. Serum calcium tends to be reduced and phosphate increased in CKD, in association with high parathyroid hormone (PTH) levels caused by reduced production of 1,25(OH)2D by the kidney (secondary hyperparathyroidism). In some patients, this may be accompanied by raised serum alkaline phosphatase levels, which are indicative of renal osteodystrophy. Other biochemical abnormalities may be observed that reflect underlying disease processes, such as raised glucose and HbA1c levels in diabetes mellitus (p. 807) and raised levels of C-reactive protein (CRP) in sepsis and vasculitis.

**Immunology**

Antinuclear antibodies, antibodies to extractable nuclear antigens and anti-double-stranded DNA antibodies may be detected in patients with renal disease secondary to SLE (p. 1109). Antineutrophil cytoplasmic antibodies (ANCA) may be detected in patients with glomerulonephritis secondary to systemic vasculitis (p. 1115), as may antibodies to GBM in patients with Goodpasture’s syndrome (p. 497) and low levels of complement in SLE, systemic vasculitis and HUS.

**Imaging**

**Ultrasound**

Renal ultrasound is a valuable non-invasive technique that is indicated to assess renal size and to investigate patients who are suspected of having obstruction of the urinary tract (Fig. 17.3) or renal tumours, cysts or stones. It is often the only method required for renal imaging and has the advantage of showing other abdominal, pelvic and retroperitoneal pathology. Ultrasound can also be used to provide images of the prostate gland and bladder, and to estimate the completeness of emptying in patients with suspected bladder outflow obstruction. Ultrasonography may show increased density of the renal cortex with loss of distinction between cortex and medulla, which is characteristic of CKD. Doppler
imaging can be used to study blood flow in extrarenal and larger intrarenal vessels and can assess the resistivity index, which is the ratio of peak systolic and diastolic velocity. This is influenced by the resistance to flow through small intrarenal arteries and may be elevated in various diseases, including acute glomerulonephritis and rejection of a renal transplant. High peak velocities can also occur in severe renal artery stenosis. However, renal ultrasound is operator-dependent, the stored images convey only a fraction of the dynamic information gained during the investigation, and the results are often less clear in obese patients.

**Computed tomography**

Computed tomography urography (CTU) is used to evaluate cysts and mass lesions in the kidney or filling defects within the collecting systems. It usually entails an initial scan without contrast medium, and subsequent scans following injection of contrast to obtain a nephrogram image and images during the excretory phases. This technique gives more information than intravascular urography (IVU) but entails a substantially larger radiation dose. Contrast enhancement is particularly useful for characterising mass lesions within the kidney and differentiating benign from malignant lesions (see Fig. 17.26, p. 506). Magnetic resonance imaging (MRI) offers excellent resolution, and that the radiation dose is significant (Box 17.4).

**Computed tomography and angiography**

This technique (CT-angiography) involves performing computed tomography, following an intravenous injection of contrast medium, to obtain images of the renal vasculature. It produces high-quality images of the main renal vessels and is of value in patients who have suffered renal trauma and those with haemorrhage from the renal tract, and in the investigation of renal artery stenosis. Other vascular structures, such as angiomyolipomas and aneurysms, can also be detected. Drawbacks include the fact that relatively large doses of contrast medium are required, which can cause renal dysfunction, and that the radiation dose is significant (Box 17.4).

**Magnetic resonance imaging**

Magnetic resonance imaging (MRI) offers excellent resolution and gives good distinction between different tissue types (see Fig. 17.26, p. 506). It is very useful for local staging of prostate, bladder and penile cancers. Magnetic resonance angiography (MRA) provides an alternative to CT-angiography for imaging renal vessels.

---

**Fig. 17.3 Renal ultrasound.** 

A Normal kidney. The normal cortex is less echo-dense (blacker) than the adjacent liver. 

B A simple cyst occupies the upper pole of an otherwise normal kidney. 

C The renal pelvis and calyces are dilated by a chronic obstruction to urinary outflow. The thinness and increased density of the remaining renal cortex indicate chronic changes.

---

**17.4 Renal complications of radiological investigations**

**Contrast nephrotoxicity**

- Acute deterioration in renal function, sometimes life-threatening, commencing < 48 hrs after administration of IV radiographic contrast media

**Risk factors**

- Pre-existing renal impairment
- Use of high-osmolality, ionic contrast media and repetitive dosing in short time periods
- Diabetes mellitus
- Myeloma

**Prevention**

- Provide hydration with free oral fluids plus IV isotonic saline 500 mL, then 250 mL/hr during procedure
- Avoid nephrotoxic drugs; withhold non-steroidal anti-inflammatory drugs (NSAIDs). Omit metformin for 48 hrs after the procedure, in case renal impairment occurs
- N-acetylcysteine may provide some protection but data are conflicting
- If the risks are high, consider alternative methods of imaging

**Cholesterol atheroembolism**

- Typically follows days to weeks after intra-arterial investigations or interventions (p. 496)

**Nephrogenic sclerosing fibrosis after MRI contrast agents**

- Chronic progressive sclerosis of skin, deeper tissues and other organs, associated with gadolinium-based contrast agents
- Only reported in patients with renal impairment, typically on dialysis or with GFR < 15 mL/min/1.73 m², but caution is advised in patients with GFR < 30 mL/min/1.73 m²
but involves administration of gadolinium-based contrast media, which may carry risks for patients with impaired renal function (see Box 17.4). Whilst MRA gives good images of the main renal vessels, stenosis of small branch arteries may be missed.

**Renal arteriography**
Renal arteriography involves taking X-rays following an injection of contrast medium directly into the renal artery. The main indication is to investigate renal artery stenosis (p. 494) or haemorrhage. Renal angiography can often be combined with therapeutic balloon dilatation or stenting of the renal artery and can be used to occlude bleeding vessels and arteriovenous fistulae by the insertion of thin platinum wires (coils). These curl up within the vessel and promote thrombosis, thereby securing haemostasis.

**Intravenous urography**
Intravenous urography (IVU) involves taking serial plain X-rays immediately before and after an intravenous injection of contrast medium. It has largely been replaced by ultrasound, CTKUB and CTU for most renal imaging purposes but remains a useful method of viewing the renal papillae, stones and urothelial malignancies (Fig. 17.4). The initial X-rays may show the renal outlines (if perinephric fat and bowel gas shadows allow), as well as radio-opaque calculi and calcification within the renal tract. Early films taken 1 minute after injection can be used to assess renal perfusion, whereas films at later time points provide images of the collecting system, ureters and bladder. The disadvantages of this technique are the need for an injection, dependence on adequate renal function, and exposure to irradiation and contrast medium (see Box 17.4).

**Pyelography**
Pyelography involves direct injection of contrast medium into the collecting system from above or below. It offers the best views of the collecting system and upper tract, and is sometimes used to identify the cause of urinary tract obstruction (p. 472). Antegrade pyelography requires the insertion of a fine needle into the pelvicalyceal system under ultrasound or radiographic control. This approach is much more difficult and hazardous in a non-obstructed kidney. In the presence of obstruction, percutaneous nephrostomy drainage can be established, and often stents can be passed through any obstruction. Retrograde pyelography can be performed by inserting catheters into the ureteric orifices at cystoscopy (Fig. 17.5).

**Radionuclide studies**
These are functional studies requiring the injection of gamma ray-emitting radiopharmaceuticals that are taken up and excreted by the kidney, a process that can be monitored by an external gamma camera.

Dynamic radionuclide studies are performed with mercaptoacetyltriglycine labelled with technetium ($^{99m}$Tc-MAG3), which is filtered by the glomerulus and excreted into the urine. Imaging following $^{99m}$Tc-MAG3 injection can provide valuable information about the perfusion of each kidney but is not a reliable method for identifying renal artery stenosis. In patients with significant obstruction of the outflow tract, $^{99m}$Tc-MAG3...
Renal biopsy

Renal biopsy is used to establish the nature and extent of renal disease in order to judge the prognosis and need for treatment (Box 17.5). The procedure is performed transcutaneously with ultrasound or contrast radiography guidance to ensure accurate needle placement into a renal pole. Light microscopy, electron microscopy and immunohistological assessment of the specimen may all be required.

17.5 Renal biopsy

**Indications**
- Acute kidney injury that is not adequately explained
- CKD with normal-sized kidneys
- Nephrotic syndrome or glomerular proteinuria in adults
- Nephrotic syndrome in children that has atypical features or is not responding to treatment
- Isolated haematuria or proteinuria with renal characteristics or associated abnormalities

**Contraindications**
- Disordered coagulation or thrombocytopenia. Aspirin and other antiplatelet agents increase bleeding risk
- Uncontrolled hypertension
- Kidneys < 60% predicted size
- Solitary kidney* (except transplants)

**Complications**
- Pain, usually mild
- Bleeding into urine, usually minor but may produce clot colic and obstruction
- Bleeding around the kidney, occasionally massive and requiring angiography with intervention, or surgery
- Arteriovenous fistula, rarely significant clinically

*Relative contraindication.

PRESENTING PROBLEMS IN RENAL AND URINARY TRACT DISEASE

Dysuria

Dysuria refers to painful urination, often described as burning, scalding or stinging, and commonly accompanied by suprapubic pain. It is often associated with frequency of micturition and a feeling of incomplete emptying of the bladder. By far the most common cause is urinary tract infection, as described on page 511. Other diagnoses that need to be considered in patients with dysuria include sexually transmitted infections (p. 411) and bladder stones (p. 507).

Loin pain

Loin pain is often caused by musculoskeletal disease but can be a manifestation of renal tract disease; in the latter case, it may arise from renal stones, ureteric stones, renal tumours, acute pyelonephritis and urinary tract obstruction. Acute loin pain radiating anteriorly and often to the groin is termed renal colic. When combined with haematuria, this is typical of ureteric obstruction due to calculi (p. 507). When loin pain is precipitated by a large fluid intake (Dietl’s crisis), upper urinary tract obstruction caused by a congenital abnormality of the pelviureteric junction (PUJ, p. 510) is often responsible.

Oliguria/anuria

Oliguria is defined as being present when less than 300 mL urine is passed per day, whereas anuria is deemed to exist when less than 50 mL urine is passed per day.

The volume of urine produced represents a balance between the amount of fluid that is filtered at the glomerulus and that reabsorbed by the renal tubules. When GFR is low, urine volumes may still be normal if tubular reabsorption is also reduced; hence urine volume alone is a poor indicator of kidney disease. Oliguria and anuria may be caused by a reduction in urine production, as in pre-renal acute kidney injury, when GFR is reduced but tubular homeostatic mechanisms increase...
reabsorption to conserve salt and water. A high solute load or associated tubular dysfunction may, however, produce normal or high urine volumes in such cases until the pre-renal insult becomes severe and GFR is markedly reduced, such as occurs in diabetic ketoacidosis with marked glycosuria. Urine volumes are variable in acute kidney injury due to intrinsic renal disease, but a rapid decline in urine volume may be observed in patients with bilateral renal infarction, in those with a single functioning kidney, and in those with rapidly progressive glomerulonephritis.

Obstruction of the renal tract can also produce oliguria and anuria, but to do so, obstruction must be complete and occur distal to the bladder neck, be bilateral, or be unilateral on the side of a single functioning kidney. If an obstruction is not relieved, the pressure will be transmitted back to the nephrons, causing GFR to fall.

Patients with oliguria and anuria should be assessed clinically to look for evidence of hypotension or volume depletion; of bladder enlargement; and of phimosis or meatal stenosis.

The presence of pain that is exacerbated by a fluid load suggests an acute obstruction of the renal tract, and its characteristics may be of value in reaching a diagnosis. Obstruction at the bladder neck is associated with lower midline abdominal discomfort due to bladder dilatation, whereas ureteric obstruction typically presents as loin pain radiating to the groin. Obstructions at the level of the renal pelvis may present as flank pain. Chronic obstruction rarely produces pain but may give rise to a dull ache. Unilateral ureteric obstruction may not lead to any noticeable reduction in urine output, whereas bilateral ureteric obstruction will result in oliguria or anuria. Urethral strictures should be considered as a possible cause, especially in patients with a history of instrumentation of the renal tract. Urethral valves should also be considered, especially in the paediatric population.

The presence of bladder enlargement in a middle-aged or elderly man suggests benign or malignant enlargement of the prostate gland as a potential cause of oliguria or anuria (pp. 515 and 517). It is important to note that about 50% of cases of acute urinary retention are seen after general anaesthesia, particularly in patients with pre-existing prostatic enlargement. Partial obstruction can be associated with a normal or even high urine volume due to chronic tubular injury, which causes loss of tubular concentrating ability. This chronic tubular injury can also produce a type 1 renal tubular acidosis (p. 446). Over time, even partial obstruction can cause tubular atrophy and irreversible renal failure.

All patients with oliguria or anuria should have routine biochemistry and haematology checked. Catheterisation should be performed so that urine volumes can be accurately monitored and the urine analysed for evidence of proteinuria and red cell casts, which may be found in patients with glomerulonephritis and systemic diseases such as vasculitis or SLE. Isolated haematuria may be indicative of an obstructive uropathy secondary to urinary calculi (p. 507) or tumours affecting the renal tract. If oliguria persists without any other clear explanation, ultrasound examination should be undertaken promptly to look for evidence of obstruction. Rarely, however, the urinary tract may not be particularly dilated in patients with acute kidney injury due to obstruction because of lack of urine production.

Management of oliguria and anuria should be directed at the underlying cause. While relief of obstruction is usually accompanied by a rapid return of renal function, tubular function may be impaired, resulting in polyuria and failure to conserve electrolytes.

### Polyuria

Polyuria is defined as a urine volume in excess of 3 L/day. Various underlying conditions, both renal and extrarenal, may be responsible, as outlined in Box 17.6.

Investigation of polyuria includes measurement of urea, creatinine and electrolytes, glucose, calcium and albumin. A 24-hour urine collection may be helpful to confirm the severity of polyuria and the presence of nocturnal polyuria. Investigation and management of suspected diabetes insipidus are described on page 794.

#### Nocturia

Nocturia is defined as waking up at night to void urine. It may be a consequence of polyuria but may also result from increased fluid intake or diuretic use in the late evening. Nocturia also occurs in CKD, and in prostatic enlargement when it is associated with poor stream, hesitancy, incomplete bladder emptying, terminal dribbling and urinary frequency due to partial urethral obstruction (p. 515). Nocturia may also occur in sleep disturbance without functional abnormalities of the urinary tract.

#### Frequency

Frequency describes micturition more often than a patient’s expectations. It may be a consequence of polyuria, when urine volume is normal or high, but is also found in patients with dysuria and prostatic diseases, when the urine volume is low.

#### Urinary incontinence

Urinary incontinence is defined as any involuntary leakage of urine. It may occur in patients with a normal urinary tract, as the result of dementia or poor mobility, or transiently during an acute illness or hospitalisation.
especially in older people (Box 17.7). Diuretics, alcohol and caffeine may worsen incontinence.

**Pathophysiology**

As urine accumulates in the bladder during the storage phase, the sphincter tone gradually increases, but there are no significant changes in vesical pressure, detrusor pressure or intra-abdominal pressure. During voiding, intravesical pressure increases, due to detrusor contraction, and the sphincter relaxes, allowing urine to flow from the bladder until it is empty. Clinical disorders associated with incontinence are connected with various abnormalities in this cycle and these are discussed in more detail below.

**Stress incontinence**

This occurs because passive bladder pressure exceeds the urethral pressure, due to either poor pelvic floor support or a weak urethral sphincter. Most often, there is an element of both. Stress incontinence is very common in women and seen most frequently following childbirth. It is rare in men and usually follows surgery to the prostate. The presentation is with incontinence during coughing, sneezing or exertion. In women, perineal inspection may reveal leakage of urine when the patient coughs.

**Urgo incontinence**

This usually occurs because of detrusor overactivity, which produces an increased bladder pressure that overcomes the urethral sphincter. Urgency with or without incontinence may also be driven by a hypersensitive bladder resulting from urinary tract infection or a bladder stone. Detrusor overactivity is usually idiopathic, other than in patients with neurological conditions such as spina bifida or multiple sclerosis, in which case it is neurogenic (p. 1174). The incidence of urge incontinence increases with age, occurring in 10-15% of the population aged over 65 years and around 50% of patients requiring nursing home care. It is also seen in men with lower urinary tract obstruction and most often remits after the obstruction is relieved.

**Continual incontinence**

This is suggestive of a fistula, usually between the bladder and vagina (vesicovaginal), or the ureter and vagina (ureterovaginal). It is most common following gynaecological surgery but is also seen in patients with gynaecological malignancy or post-radiotherapy. In parts of the world where obstetric services are scarce, prolonged obstructed labour can be a common cause of vesicovaginal fistulae. Continual incontinence may also be seen in infants with congenital ectopic ureters. Occasionally, stress incontinence is so severe that the patient leaks continuously.

**Overflow incontinence**

This occurs when the bladder becomes chronically overdistended and may be associated with acute kidney injury (high-pressure chronic urinary retention). It is most commonly seen in men with benign prostatic enlargement or bladder neck obstruction (p. 514), but may arise in either sex as a result of failure of the detrusor muscle (atonic bladder). The latter may be idiopathic but more commonly is the result of damage to the pelvic nerves, either from surgery (commonly, hysterectomy or rectal excision), trauma or infection, or from compression of the cauda equina by disc prolapse, trauma or tumour. Incontinence due to prostatic enlargement can be regarded as a type of overflow incontinence.

**Post-micturition dribble**

This is very common in men, even in the relatively young. It is due to a small amount of urine becoming trapped in the U-bend of the bulbar urethra, which leaks out when the patient moves. It is more pronounced if associated with a urethral diverticulum or urethral stricture. It may occur in females with a urethral diverticulum and may mimic stress incontinence.

**Clinical assessment**

Patients should be encouraged to keep a voiding diary, including the measured volume voided, frequency of voiding, precipitating factors and associated features, such as urgency, since this can be of diagnostic value. The patient should be assessed for evidence of cognitive impairment and mobility. A neurological assessment should be performed to detect disorders such as multiple sclerosis that may affect the nervous supply of the bladder, and the lumbar spine should be inspected for features of spina bifida occulta. Perineal sensation and anal sphincter tone should be assessed. Rectal examination is needed to assess the prostate in men and to exclude faecal impaction as a cause of incontinence. Genital examination should be done to identify phimosis and paraphimosis in men, and vaginal mucosal atrophy, cystoceles or rectoceles in women.

**Investigations**

Urinalysis and culture should be performed in all patients. Ultrasound examination can be helpful in identifying patients with overflow incontinence who have incomplete bladder emptying, as they may reveal a significant amount of fluid in the bladder (> 100 mL) post-micturition. Urine flow rates and full urodynamic assessment by cystometryraphy may be required to diagnose the type of incontinence and are indicated in selected cases when the diagnosis is unclear on clinical grounds. An IVU should be performed in patients with continual incontinence who are suspected of having a
Management

Females with stress incontinence respond well to physiotherapy. The mainstay of treatment for urge incontinence is bladder retraining, which involves teaching patients to hold more urine voluntarily in their bladder, assisted by anticholinergic medication. Surgery may be required in patients who have severe daytime incontinence despite conservative treatment. The treatment of incontinence secondary to fistula formation is surgical. Patients with overflow incontinence due to bladder obstruction should be treated surgically or with long-term catheterisation (intermittent or continuous). Incontinence secondary to neurological diseases can be treated by intermittent self-catheterisation.

Erectile dysfunction

Causes of erectile failure are shown in Box 17.8. Vascular, neuropathic and psychological causes are most common. With the exception of diabetes mellitus, endocrine causes are relatively uncommon and are characterised by loss of libido, as well as erectile dysfunction. Erectile dysfunction and reduced libido occur in over 50% of men with advanced CKD or on dialysis. Erectile dysfunction is a markedly under-diagnosed problem. It is important to discuss matters frankly with the patient, and to establish whether there are associated features of hypogonadism (p. 760) and if erections occur at any other time. If the patient has erections on wakening, vascular and neuropathic causes are much less likely and a psychological cause should be suspected.

Investigations

Blood should be taken for glucose, prolactin, testosterone, luteinising hormone (LH) and follicle-stimulating hormone (FSH). A number of further tests are available but are rarely employed because they do not usually influence management. These include nocturnal tumescence monitoring (using a photoplethysmograph placed around the shaft of the penis overnight) to establish whether blood supply and nerve function are sufficient to allow erections to occur during sleep; intracavernosal injection of prostaglandin E1 to test the adequacy of blood supply; internal pudendal artery angiography; and tests of autonomic and peripheral sensory nerve conduction.

17.9 Interpretation of dipstick-positive haematuria

<table>
<thead>
<tr>
<th>Dipstick test positive</th>
<th>Urine microscopy</th>
<th>Suggested cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haematuria</td>
<td>White blood cells</td>
<td>Infection</td>
</tr>
<tr>
<td></td>
<td>Abnormal epithelial cells</td>
<td>Tumour</td>
</tr>
<tr>
<td></td>
<td>Red cell casts</td>
<td>Glomerular bleeding*</td>
</tr>
<tr>
<td></td>
<td>Dysmorphic erythrocytes (phase contrast microscopy)</td>
<td></td>
</tr>
<tr>
<td>Haemoglobinuria</td>
<td>No red cells</td>
<td>Intravascular haemolysis</td>
</tr>
<tr>
<td>Myoglobinuria (brown urine)</td>
<td>No red cells</td>
<td>Rhabdomyolysis</td>
</tr>
</tbody>
</table>

*Glomerular bleeding implies that the GBM is ruptured. It can occur physiologically following very strenuous exertion but usually indicates intrinsic renal disease and is an important feature of the nephritic syndrome (see Box 17.12).
of haematuria is always required, as even non-visible haematuria can be caused by malignancy.

Glomerular bleeding is characteristic of inflammatory, destructive or degenerative processes that disrupt the GBM and may cause microscopic or macroscopic haematuria. In glomerulonephritis, one or more other features of the ‘nephritic syndrome’ (Box 17.11) may be present but the full syndrome is rare (p. 497). Other benign causes include thin basement membrane disease (p. 502) and vascular malformations.

Visible haematuria is most likely to be caused by tumour, which can affect any part of the urogenital tract (p. 515 and Box 17.12). Other common causes of visible haematuria are urine infection and stones. Investigation of haematuria (Fig. 17.8), whether non-visible or visible,
renal disease (p. 497) should be considered and a renal biopsy may be indicated. Where there are no features of significant kidney disease, and malignancy and renal stones have been excluded, it may be appropriate to manage the patient by observation with periodic review, although occasionally these individuals develop significant overt renal disease during follow-up. Management of haematuria should be directed at the underlying cause.

Proteinuria and nephrotic syndrome

Whilst moderate amounts of low-molecular-weight protein pass through the healthy GBM, these proteins are normally reabsorbed by receptors on tubular cells. In healthy individuals, less than 150 mg of protein is excreted in the urine each day. A proportion of this is Tamm–Horsfall protein, secreted by the renal tubules. The presence of larger amounts of protein is usually indicative of significant renal disease (Box 17.13).

![Fig. 17.8 Investigation of haematuria.](image)

### Clinical assessment

Proteinuria is usually asymptomatic and is often picked up by urinalysis. Transient proteinuria can occur after vigorous exercise, during fever, in heart failure and in patients with urinary tract infection. Patients should be assessed for the presence of these conditions and urine testing repeated once the potential trigger has been treated or resolved. So-called orthostatic proteinuria may arise in the absence of renal disease. This occurs only during the day, in association with an upright posture, and the first morning sample is negative. Typically, less than 1 g protein per day is excreted. Orthostatic proteinuria is regarded as a benign disorder which does not require treatment. Large amounts of protein make urine froth easily and this may be noticed by the patient. Very large amounts can cause nephrotic syndrome, which presents with oedema (Boxes 17.11 and 17.14). In adults, this predominantly affects the lower limbs but extends to the genitalia and lower abdomen as it becomes more severe. The upper limbs and face may be more affected on waking in the morning. In children, ascites occurs early and oedema is often seen only in the face. Blood volume may be normal, reduced or increased. Renal sodium retention is an early and universal feature; the mechanisms of this are shown in Figure 16.5 (p. 435).

Microalbuminuria refers to the urinary excretion of small amounts of albumin. The consistent presence of albumin in the urine is abnormal and is clinically important in identifying the very early stages of glomerular disease, as occurs in conditions like diabetic nephropathy (p. 830). Because significant renal damage has already taken place before standard dipstick tests become positive, patients with diabetes mellitus should be screened regularly for microalbuminuria. Persistent microalbuminuria has also been associated with an increased risk of atherosclerosis and cardiovascular

### Table 17.13 Quantifying proteinuria in random urine samples

<table>
<thead>
<tr>
<th>ACR&lt;sup&gt;1&lt;/sup&gt;</th>
<th>PCR&lt;sup&gt;2&lt;/sup&gt;</th>
<th>Typical dipstick results&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3.5 (female)</td>
<td>&lt; 2.5 (male)</td>
<td>–</td>
<td>Normal</td>
</tr>
<tr>
<td>~3.5–15</td>
<td></td>
<td>Dipticks positive; equivalent to 24-hr protein excretion &lt; 0.5 g</td>
<td>Microalbuminuria</td>
</tr>
<tr>
<td>~15–50</td>
<td>~15–50</td>
<td>++ to +++</td>
<td>Nephrotic range: always glomerular disease, equivalent to 24-hr protein excretion &gt; 3 g</td>
</tr>
<tr>
<td>50–200</td>
<td>&gt; 250</td>
<td>++ to +++</td>
<td>Glomerular disease more likely</td>
</tr>
<tr>
<td>&gt; 200</td>
<td>&gt; 300</td>
<td>+++ to +++</td>
<td>Glomerular disease more likely</td>
</tr>
</tbody>
</table>

<sup>1</sup>Urinary albumin (mg/L)/urine creatinine (mmol/L).  
<sup>2</sup>Urine protein (mg/L)/urine creatinine (mmol/L). (If urine creatinine is measured in mg/dL, reference values for PCR and ACR can be derived by dividing by 11.31.)  
<sup>3</sup>Dipstick results are affected by urine concentration and are occasionally weakly positive on normal samples.
mortality but neither the mechanism of proteinuria nor an explanation for this association has been established.

**Investigations**

All patients with persistent proteinuria should have the amount of protein quantified to guide further investigations (Fig. 17.9). Since quantification by 24-hour urine collection is often inaccurate, the protein: creatinine ratio (PCR) in a spot sample of urine is preferred. This makes an allowance for the variable degree of urinary dilution and can be used to extrapolate to 24-hour values (see Box 17.13). Changes in PCR also give valuable information about the progression of renal disease in CKD. It is possible to measure 24-hour albumin excretion or albumin: creatinine ratio (ACR), but this requires a more expensive immunoassay and is usually reserved for detection of the early stages of diabetic nephropathy (p. 830). When assessing protein excretion by analysis of spot urine samples, greater consistency can be achieved by using first morning urine samples but this is not essential for routine clinical use.

It is sometimes helpful to identify the type of protein in the urine. Low-molecular-weight proteins may appear in the urine in quantities greater than 150 mg/day. This is usually assessed by measurement of specific low-molecular-weight proteins, such as β2-microglobulin (molecular weight 12 kDa). Large amounts of small proteins in the urine suggest renal tubular damage and are referred to as tubular proteinuria. This rarely exceeds 1.5–2 g/24 hours (maximum PCR 150–200 mg/mmol). When more than 2 g protein per day is being excreted, glomerular disease is likely and this is an indication for renal biopsy. The diseases that cause nephrotic syndrome all affect the glomerulus (see Fig. 17.21, p. 497), either directly, by damaging podocytes, or indirectly, by causing scarring or by depositing exogenous material such as amyloid into the glomerulus. A notable exception is in children, when heavy proteinuria and nephrotic syndrome are most commonly caused by minimal change glomerulonephritis. In this case, renal biopsy is not usually required unless the patient fails to respond to high-dose corticosteroid therapy.

**Box 17.13.** Changes in PCR also give valuable information about the progression of renal disease in CKD. It is possible to measure 24-hour albumin excretion or albumin: creatinine ratio (ACR), but this requires a more expensive immunoassay and is usually reserved for detection of the early stages of diabetic nephropathy (p. 830). When assessing protein excretion by analysis of spot urine samples, greater consistency can be achieved by using first morning urine samples but this is not essential for routine clinical use.

**Fig. 17.9 Investigation of proteinuria.** (ACR = albumin: creatinine ratio; PCR = protein: creatinine ratio.)

Free immunoglobulin light chains (molecular weight 25 kDa) are filtered freely at the glomerulus and can be identified as ‘Bence Jones protein’ in fresh urine samples. Bence Jones protein is poorly identified by dipstick tests and so specific immunodetection methods are required. This may occur in AL amyloidosis (p. 86) and in B-cell
**Oedema**

Oedema is caused by an excessive accumulation of fluid within the interstitial space. Clinically, this can be detected by persistency of an indentation in tissue following pressure on the affected area (pitting oedema). Non-pitting oedema is typical of lymphatic obstruction and may also occur as the result of excessive matrix deposition in tissues – for example, in hypothyroidism (p. 743) or scleroderma (p. 1112). There are various possible causes (Box 17.15). Pitting oedema tends to accumulate in the ankles during the day and improves overnight as the interstitial fluid is reabsorbed. In developed countries, the most common causes of oedema are local venous problems and heart failure (p. 546), but it is important to identify other causes.

**Clinical assessment**

The ankles and lower parts of the leg are typically affected first but oedema can be restricted to the sacrum in bed-bound patients. With increasing severity, oedema spreads to affect the upper parts of the legs, the genitalia and abdomen. Ascites is common and often an earlier feature in children or young adults, and in liver disease. Pleural effusions are common and may also occur as the result of excessive matrix deposition in tissues – for example, in hypothyroidism (p. 743) or scleroderma (p. 1112). There are various possible causes (Box 17.15). Pitting oedema tends to accumulate in the ankles during the day and improves overnight as the interstitial fluid is reabsorbed. In developed countries, the most common causes of oedema are local venous problems and heart failure (p. 546), but it is important to identify other causes.

**Management**

Management of proteinuria should be directed at the underlying cause. This may involve immunosuppressive therapy in glomerulonephritis and supportive management for nephrotic syndrome (Box 17.14).

**Investigations**

The cause of oedema is usually apparent from the history and examination of the cardiovascular system and abdomen. Blood should be taken for measurement of urea and electrolytes and serum albumin, and the urine tested for protein. Further imaging of the liver, heart or kidneys may be indicated, based on history and clinical examination. Where ascites or pleural effusions occur in isolation, aspiration of fluid with measurement of protein and glucose, and microscopy for cells, will usually help to clarify the diagnosis in differentiating a transudate (typical of oedema) from an exudate (more suggestive of local pathology, p. 662).

**Hypertension**

Hypertension is a very common feature of renal disease. Additionally, the presence of hypertension identifies a population at risk of developing CKD and current recommendations are that patients on antihypertensive medication should have renal function checked annually. Control of hypertension is very important in patients with renal impairment because of its close relationship with further decline of renal function (p. 486) and because of the exaggerated cardiovascular risk associated with CKD. Pathophysiology and management are discussed on pages 606–613.

**ACUTE KIDNEY INJURY**

Acute kidney injury (AKI), also referred to as acute renal failure, describes the situation where there is a sudden and often reversible loss of renal function, which develops over days or weeks and is usually accompanied by a reduction in urine volume. Approximately 7% of all hospitalised patients and 20% of acutely ill patients develop signs of AKI. In uncomplicated AKI, such as that due to haemorrhage or drugs, mortality is low, even when renal replacement therapy is required. In AKI associated with serious infection and multiple organ failure, mortality is 50–70% and the outcome is usually determined by the severity of the underlying disorder and other complications, rather than by renal failure itself.
Pathophysiology

There are many causes of AKI and it is frequently multifactorial. It is often classified into three subtypes: ‘pre-renal’, when perfusion to the kidney is reduced; ‘renal’, when the primary insult affects the kidney itself; and ‘post-renal’, when there is obstruction to urine flow at any point from the tubule to the urethra (Fig. 17.10). In pre-renal AKI, the kidney becomes damaged as the result of hypoperfusion, leading to acute tubular necrosis. Histologically, the kidney shows inflammatory changes, focal breaks in the tubular basement membrane and interstitial oedema. Dead tubular cells may also be shed into the tubular lumen, leading to tubular obstruction. Although tubular cell damage is the dominant feature under the microscope, there may also be profound alterations in the renal microcirculation. Renal AKI may be caused by nephrotoxic drugs (p. 522), which can cause acute tubular necrosis and a similar histological picture to pre-renal AKI or interstitial nephritis. The other common cause is glomerulonephritis, in which there is direct inflammatory damage to the glomeruli (p. 498). Post-renal AKI occurs as the result of obstruction to the renal tract, with kidney damage arising as the result of back pressure. Anaemia is common in AKI and may occur as the result of blood loss, haemolysis or decreased erythropoiesis. In established AKI, there is an increased risk of bleeding and spontaneous gastrointestinal haemorrhage due to the uremia.

Clinical features

Early recognition and intervention is important in AKI; all emergency admissions to hospital should have renal function, blood pressure, temperature and pulse checked on arrival and should undergo a risk assessment for the likelihood of developing AKI. This includes looking at coexisting diseases such as diabetes, vascular disease and liver disease, which make AKI more likely, as well as gathering information on drug treatments such as ACE inhibitors and NSAIDs, which may be associated with renal dysfunction. If a patient is found to have a high serum creatinine, it is important to establish whether this is an acute or acute-on-chronic phenomenon, or a sign of chronic kidney disease. Previous measurements of renal function can be of great value in differentiating these possibilities. Patients with AKI need to be assessed quickly to determine the likely underlying cause. Investigations that are required in all cases are shown in Box 17.16. Additional investigations that are required in some cases, depending on the clinical picture are shown in Box 17.17. The diagnosis of pre-renal AKI is usually obvious clinically (see below). Various criteria have been proposed to classify AKI and to help identify high-risk patients, guide treatment and provide information regarding prognosis. The most commonly used are the KDIGO and RIFLE criteria (Box 17.18), which use serum creatinine and urine output as biomarkers of kidney function.

Pre-renal AKI

Patients with pre-renal AKI are typically hypotensive and tachycardic with signs of poor peripheral perfusion, such as delayed capillary return. Tachycardia and postural hypotension (a fall in blood pressure of > 20/10 mmHg from lying to standing) are valuable signs of early hypovolaemia. Many patients with sepsis initially present with poor peripheral perfusion, as mentioned above, but then show evidence of peripheral vasodilatation once they have undergone initial resuscitation with intravenous fluids. However, this is accompanied by relative underfilling of the arterial tree and the kidney responds as it would to absolute hypovolaemia, with renal vasoconstriction leading to AKI with acute tubular necrosis. Biochemical assessment in pre-renal AKI usually reveals evidence of a metabolic acidosis and hyperkalaemia. It is important to note that pre-renal AKI may also occur without systemic hypotension, particularly in patients taking NSAIDs or ACE inhibitors. The cause of the hypotension is usually obvious clinically, but concealed blood loss can occur into the gastrointestinal tract, following trauma (particularly where there are fractures of the pelvis or femur), and into the pregnant uterus. Large volumes of intravascular fluid may also be lost into tissues after crush injuries or burns, and in severe inflammatory skin diseases or sepsis.

Renal AKI

Factors that can help differentiate the various causes of renal and post-renal AKI are summarised in Box 17.19. Patients with glomerulonephritis usually demonstrate significant haematuria and proteinuria and may have clinical manifestations of an underlying disease, such as SLE or systemic vasculitis. Although blood tests, including an immunological screen, should be performed to clarify the diagnosis in glomerulonephritis, a renal biopsy is usually required. Drug-induced acute interstitial nephritis is harder to spot but should be suspected in a previously well patient if there is an acute

---

**Fig. 17.10 Causes of acute kidney injury.**
### 17.16 Investigation of patients with established acute kidney injury

<table>
<thead>
<tr>
<th>Initial test</th>
<th>Interpretation and further tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urea and creatinine</td>
<td>Compare to previous results. Chronically abnormal in CKD (see Box 17.26)</td>
</tr>
<tr>
<td>Electrolytes</td>
<td>If potassium &gt; 6 mmol/L, treat urgently (p. 442)</td>
</tr>
<tr>
<td>Calcium and phosphate</td>
<td>Low calcium and high phosphate may indicate CKD (see Box 17.26)</td>
</tr>
<tr>
<td></td>
<td>Calcium low in rhabdomyolysis: measure creatine kinase</td>
</tr>
<tr>
<td></td>
<td>Hypercalcaemia in myeloma</td>
</tr>
<tr>
<td>Albumin</td>
<td>Low albumin in nephrotic syndrome (see urinalysis below)</td>
</tr>
<tr>
<td></td>
<td>Low albumin in sepsis: take blood cultures</td>
</tr>
<tr>
<td>Full blood count</td>
<td>Anaemia may indicate CKD (see Box 17.26) or myeloma (see Box 17.17)</td>
</tr>
<tr>
<td>Clotting screen</td>
<td>Anaemia and fragmented RBC on blood film with raised LDH in thrombotic microangiopathy</td>
</tr>
<tr>
<td></td>
<td>Low platelets and abnormal coagulation in DIC, including in sepsis: take blood cultures</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>ESR is misleading in renal failure</td>
</tr>
<tr>
<td></td>
<td>High CRP may indicate sepsis or inflammatory disease</td>
</tr>
<tr>
<td>Urinalysis</td>
<td>Less reliable in an oliguric catheterised patient</td>
</tr>
<tr>
<td></td>
<td>Seek earlier results if possible</td>
</tr>
<tr>
<td></td>
<td>Marked haematuria suggests glomerulonephritis, tumour of renal tract or bleeding disorder</td>
</tr>
<tr>
<td></td>
<td>Heavy proteinuria suggests glomerular disease: measure PCR or ACR</td>
</tr>
<tr>
<td>Urine microscopy</td>
<td>Casts or dysmorphic red cells suggest glomerulonephritis</td>
</tr>
<tr>
<td></td>
<td>Leucocytes suggest infection/interstitial nephritis</td>
</tr>
<tr>
<td></td>
<td>Crystals may be observed in drug-induced or uric acid nephropathy</td>
</tr>
<tr>
<td>Renal ultrasound</td>
<td>Hydronephrosis ± enlarged bladder in urinary tract obstruction: consider PSA and further imaging of urinary tract</td>
</tr>
<tr>
<td></td>
<td>Small kidneys suggest CKD (see Box 17.26)</td>
</tr>
<tr>
<td></td>
<td>Asymmetric kidneys suggest renovascular or developmental disease: consider renal artery imaging</td>
</tr>
<tr>
<td>Cultures</td>
<td>Culture blood, urine, sputum, wounds as appropriate</td>
</tr>
<tr>
<td></td>
<td>Treat all infections</td>
</tr>
<tr>
<td>Chest X-ray</td>
<td>Pulmonary oedema in fluid overload</td>
</tr>
<tr>
<td></td>
<td>Globular heart in pericardial (uraemic) effusion: perform echocardiogram</td>
</tr>
<tr>
<td></td>
<td>‘Bat wing’ appearance with normal heart size (± low Hb) may suggest pulmonary haemorrhage: measure CO transfer factor</td>
</tr>
<tr>
<td></td>
<td>Fibrotic change in systemic inflammatory disease with lung and kidney involvement: request pulmonary function and high-resolution CT</td>
</tr>
<tr>
<td>Serology</td>
<td>HIV and hepatitis serology is urgent if dialysis is needed</td>
</tr>
<tr>
<td>ECG</td>
<td>If patient is &gt; 40 yrs or has electrolyte abnormalities or risk of cardiac disease</td>
</tr>
</tbody>
</table>

(ACR = albumin:creatinine ratio; CKD = chronic kidney disease; CRP = C-reactive protein; DIC = disseminated intravascular coagulation; ESR = erythrocyte sedimentation rate; Hb = haemoglobin; LDH = lactate dehydrogenase; PCR = protein:creatinine ratio; PSA = prostate-specific antigen; RBC = red blood cells)

### 17.17 Clinical features and investigations of specific causes of acute kidney injury

<table>
<thead>
<tr>
<th>Possibility</th>
<th>Consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vascular occlusion</td>
<td>Urgent arteriography</td>
</tr>
<tr>
<td>Aorta, or renal artery to single kidney; pointers include newly missing pulses, complete anuria</td>
<td>Doppler ultrasound</td>
</tr>
<tr>
<td>Malignant hypertension</td>
<td>Examine optic fundi for typical features</td>
</tr>
<tr>
<td>BP very high; RBC fragments on blood film and haemolysis</td>
<td>Check previous BP readings</td>
</tr>
<tr>
<td>Scleroderma</td>
<td>Autoantibodies to extractable nuclear antigens</td>
</tr>
<tr>
<td>Sclerodactyly, other features of scleroderma, severe hypertension</td>
<td>Imaging for involvement of other organs (p. 1112)</td>
</tr>
<tr>
<td>Systemic inflammatory disease</td>
<td>Complement (see Box 17.40), ANCA, ANF, anti-GBM antibodies, cryoglobulins and tissue biopsy</td>
</tr>
<tr>
<td>Multi-organ involvement, rash and evidence of glomerular disease</td>
<td>Cultures, echocardiogram</td>
</tr>
<tr>
<td>Differential diagnosis includes infection, especially endocarditis or tuberculosis</td>
<td></td>
</tr>
</tbody>
</table>
Acute kidney injury

17.17 Clinical features and investigations of specific causes of acute kidney injury – cont’d

<table>
<thead>
<tr>
<th>Possibility</th>
<th>Consider</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glomerular disease</strong></td>
<td>Screen for systemic inflammatory disease as above</td>
</tr>
<tr>
<td>Heavy proteinuria and/or haematuria</td>
<td>Plus urgent renal biopsy, unless cause already known</td>
</tr>
<tr>
<td><strong>Interstitial nephritis</strong></td>
<td>Eosinophilia and urinary eosinophils</td>
</tr>
<tr>
<td>Consider if urinary abnormalities minor but leucocytes present. Usually non-oliguric in early stages. Take detailed history of exposure to drugs and other possible causes</td>
<td>Renal biopsy</td>
</tr>
<tr>
<td><strong>Myeloma</strong></td>
<td>Uric acid if tumour lysis possible</td>
</tr>
<tr>
<td>Extra renal features may be present, including bone pain and hypercalcaemia. Renal disease can occur without overt myeloma. Often presents with interstitial nephritis but cast formation can be acute, so patients frequently oliguric</td>
<td>FBC, calcium, skeletal survey, bone marrow aspirate</td>
</tr>
<tr>
<td><strong>Infections</strong></td>
<td>Serum immunoglobulins and protein electrophoresis</td>
</tr>
<tr>
<td>Leptospirosis, hantavirus, syphilis, post-streptococcal glomerulonephritis</td>
<td>Check urinary light chains</td>
</tr>
<tr>
<td>ASO titre and other serological tests for infection</td>
<td></td>
</tr>
</tbody>
</table>

(ANCA = antineutrophil cytoplasmic antibodies; ANF = antinuclear factor; ASO = anti-streptolysin O; BP = blood pressure; FBC = full blood count; GBM = glomerular basement membrane; RBC = red blood cells)

17.18 Criteria for acute kidney injury

<table>
<thead>
<tr>
<th>KDIGO (Kidney disease: improving global outcomes)</th>
<th>RIFLE (Risk, injury, failure, loss, end-stage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1: serum creatinine increase &gt; 1.5–1.9-fold, urine production of &lt; 0.5 mL/kg/hr for 6–12 hrs</td>
<td>= Risk</td>
</tr>
<tr>
<td>Stage 2: serum creatinine increase &gt; 2.0–2.9-fold, urine production of &lt; 0.5 mL/kg/hr for ≥ 12 hrs</td>
<td>= Injury</td>
</tr>
<tr>
<td>Stage 3: serum creatinine increase &gt; 3.0-fold, urine production of &lt; 0.3 mL/kg/hr for ≥ 24 hrs or absolute anuria for ≥ 12 hrs, or absolute serum creatinine &gt; 354 μmol/L with an acute rise of &gt; 44 μmol/L</td>
<td>= Failure</td>
</tr>
<tr>
<td>Loss: persistent AKI, or complete loss of kidney function for &gt; 4 wks</td>
<td></td>
</tr>
<tr>
<td>End-stage renal disease: need for renal replacement therapy for &gt; 3 mths</td>
<td></td>
</tr>
</tbody>
</table>

17.19 Acute kidney injury in a haemodynamically stable, non-septic patient

**Urinary tract obstruction**
- Suggested by a history of loin pain, haematuria, renal colic or difficulty in micturition but often clinically silent
- Can usually be excluded by renal ultrasound: essential in any patient with unexplained AKI
- Prompt relief of the obstruction restores renal function

**Drugs and toxins**
- Poisoning, paraphenylenediamine hair dye, snake bite, paracetamol, herbal medicines, *Cortinarius* mushrooms
- Therapeutic agents: direct toxicity (aminoglycosides, amphotericin, tenofovir), or haemodynamic effects (NSAIDs, ACE inhibitors), often with other factors. Phosphate crystallisation after IV administration or from bowel preparation
- Sometimes associated with systemic vasculitis, systemic lupus erythematosus (SLE) and Goodpasture’s (anti-GBM) disease
- Useful blood tests include: antineutrophil cytoplasmic antibodies (ANCA), antinuclear antibodies (ANA), anti-GBM antibodies, complement, immunoglobulins
- Renal biopsy shows aggressive glomerular inflammation, usually with crescent formation

**Acute interstitial nephritis**
- Usually caused by an adverse drug reaction
- Characterised by small amounts of blood and protein in urine, often with leucocyturia
- Kidneys are normal size
- Requires cessation of drug and often prednisolone treatment

Deterioration of renal function coinciding with introduction of a new drug treatment. Drugs that are commonly implicated include gentamicin, omeprazole, cisplatin and amphotericin B.

**Post-renal AKI**

Patients should be examined clinically to look for evidence of bladder enlargement and should also undergo imaging with ultrasound to detect evidence of obstruction above the level of the bladder. Post-renal AKI is usually accompanied by hydronephrosis, but this can be absent if the ureters are affected by fibrosis or malignancy, or if obstruction of the renal tract occurs in combination with a renal disorder such as acute tubular necrosis that causes reduced urinary flow.

**Management**

Management options common to all forms of AKI are discussed in more detail below and summarised in Box 17.20.
Haemodynamic status

If hypovolaemia is present, it should be corrected by replacement of intravenous fluid or blood; excessive administration of fluid should be avoided, since this can worsen outcome in AKI due to the development of pulmonary oedema. Monitoring of the central venous pressure may be of value in determining the rate of administration of fluid in these circumstances. Balanced salt solutions, such as Hartmann’s or Ringer’s lactate, may be preferable to isotonic (0.9%) saline when large volumes of fluid resuscitation are required, in order to avoid hyperchloraemic acidosis, but whether this substantially influences outcome remains unclear. Administration of hydroxyethyl starch solutions should be avoided, since they have been associated with higher rates of established AKI (Box 17.21). Critically ill patients may require inotropic drugs to restore an effective blood pressure but clinical trials do not support a specific role for low-dose dopamine (Box 17.22).

Hyperkalaemia and acidosis

Hyperkalaemia is common, particularly in patients with sepsis, burns, haemolysis or metabolic acidosis (p. 442). If serum K⁺ concentration is > 6.5 mmol/L, this should be treated immediately, as described in Box 16.17 (p. 443), to prevent life-threatening cardiac arrhythmias. Metabolic acidosis develops unless prevented by loss of hydrogen ions through vomiting or aspiration of gastric contents. Severe acidosis can be ameliorated with sodium bicarbonate if volume status allows. Restoration of blood volume will correct acidosis by restoring kidney function. Infusions of sodium bicarbonate (50 mL of 8.4%) may also be used, if acidosis is severe, to reduce life-threatening hyperkalaemia.

Cardiopulmonary complications

Pulmonary oedema (Fig. 17.11) may result from the administration of excessive amounts of fluids relative to urine output and because of increased pulmonary capillary permeability. If pulmonary oedema is present and urine output cannot be rapidly restored, treatment with dialysis may be required to remove excess fluid. Temporary respiratory support may also be necessary with continuous positive airways pressure (CPAP) or intermittent positive pressure ventilation (IPPV). Once initial resuscitation has been performed, fluid intake should be matched to urine output plus 500 mL to cover insensible losses, unless diarrhoea is present, in which case additional fluids might be required.

Electrolyte disturbances

Electrolyte disturbances, such as dilutional hyponatraemia, may occur if the patient has continued to drink freely despite oliguria or has received inappropriate amounts of intravenous dextrose. They can be avoided by paying careful attention to fluid balance and by giving intravenous fluids slowly. Modest hypercalcaemia is common but rarely requires treatment. Serum phosphate levels are usually high but may fall to

---

**Fig. 17.11 Pulmonary oedema in acute kidney injury.** The appearances are indistinguishable from left ventricular failure but the heart size is usually normal. Blood pressure is often high.

---

**EBM 17.20 Management of acute kidney injury**

- Correct hypovolaemia and optimise systemic haemodynamic status with inotropic drugs if necessary
- Administer glucose and insulin to correct hyperkalaemia if K⁺ > 6.5 mmol/L
- Consider administering sodium bicarbonate (100 mmol) to correct acidosis if pH < 7.0 (> 100 mmol/L)
- Discontinue potentially nephrotoxic drugs and reduce doses of therapeutic drugs according to level of renal function
- Match fluid intake to urine output plus an additional 500 mL to cover insensible losses once patient is euovolaemic
- Measure body weight on a regular basis as a guide to fluid requirements
- Ensure adequate nutritional support
- Administer proton pump antagonists to reduce the risk of upper gastrointestinal bleeding
- Screen for intercurrent infections and treat promptly if present

**EBM 17.21 Colloid resuscitation fluids in the critically ill**

‘A systematic review and meta-analysis of randomised controlled trials comparing hydroxyethyl starch (HES) with crystalloid or albumin solutions in the resuscitation of critically ill patients with sepsis revealed that, in trials with a low risk of bias, HES was associated with increased mortality (relative risk 1.11, 95% confidence interval 1.0–1.23), a greater requirement for renal replacement therapy (1.36 [1.03–1.80]) and more serious adverse events (1.30 [1.02–1.67]).’


**EBM 17.22 Low-dose dopamine in acute kidney injury**

‘Dopamine at low, “renal” doses has been used in the belief that it may increase renal blood flow in critically ill patients (as it does in normal individuals) and prevent AKI. However, meta-analysis of clinical trials does not support its use in patients with, or at risk of, acute kidney injury.’

dangerously low levels in patients on daily or continuous dialysis or haemofiltration, necessitating phosphate replacement.

Dietary measures

Adequate nutritional support should be ensured and it is important to give sufficient amounts of energy and adequate amounts of protein; high protein intake should be avoided. This is particularly important in patients with sepsis and burns who are hypercatabolic. Enteral or parenteral nutrition may be required (p. 123).

Infection

Patients with AKI are at substantial risk of intercurrent infection because humoral and cellular immune mechanisms are depressed. Regular clinical examination, supplemented by microbiological investigation where appropriate, is required to diagnose infection. If infection is discovered, it should be treated promptly according to standard principles (Ch. 6).

Immunosuppression

Patients with glomerulonephritis may require immunosuppressive drugs (p. 498), plasma infusion and plasma exchange (p. 501).

Renal tract obstruction

In post-recovery AKI, the obstruction should be relieved as soon as possible. This may involve catheterisation in urethral obstruction, or correction of ureteric obstruction with a ureteric stent or percutaneous nephrostomy.

Renal replacement therapy

Conservative management can be successful in AKI with meticulous attention to fluid balance, electrolytes and nutrition, but renal replacement therapy (RRT) may be required in patients who are not showing signs of recovery with these measures. Typically, the decision to start RRT is driven by hyperkalaemia, fluid overload or acidosis. Severe uraemia with pericarditis and neurological signs (uraemic encephalopathy) is uncommon in AKI but, when present, is a strong indication for RRT. No specific cut-off values for serum urea or creatinine have been identified at which RRT should be commenced, and clinical trials of earlier versus later RRT in unselected patients with AKI have not shown differences in outcome. Furthermore, RRT can be a risky intervention in patients with comorbidity, since it requires the placement of large intravenous catheters that may become infected and can also represent a major haemodynamic challenge in unstable patients. Accordingly, the decision to institute RRT should be made on an individual basis, taking account of the potential risks and benefits, comorbidity and other aspects of the patient’s care, including an assessment of whether early or delayed recovery is likely. The two main options for RRT in AKI are haemodialysis and high-volume haemofiltration, or the hybrid approach of haemodiafiltration. Peritoneal dialysis is also an option if haemodialysis is not available (p. 490).

Recovery from AKI

This is usually heralded by a gradual return of urine output and a steady improvement in plasma biochemistry. During recovery, there is often a diuretic phase in which urine output increases rapidly and remains excessive for several days before returning to normal. This may be due in part to tubular damage and to temporary loss of the medullary concentration gradient. This gradient plays a key role in concentrating urine in the collecting duct, and depends on continued delivery of filtrate to the ascending limb of the loop of Henle and active tubular transport. After a few days, urine volume falls to normal as the concentrating mechanism and tubular reabsorption are restored. During the recovery phase of AKI, it may be necessary to provide supplements of sodium chloride, sodium bicarbonate, potassium chloride and sometimes phosphate temporarily, to compensate for increased urinary losses.

17.23 Acute kidney injury in old age

- **Physiological change**: nephrons decline in number with age and average GFR falls progressively.
- **Creatinine**: as muscle mass falls with age, less creatinine is produced each day. Serum creatinine can be misleading as a guide to renal function.
- **Renal tubular function**: declines with age, leading to loss of urinary concentrating ability.
- **Drugs**: increased drug prescription in older people (diuretics, ACE inhibitors and NSAIDs) may contribute to risk of AKI.
- **Causes**: infection, renal vascular disease, prostatic obstruction, hypovolaemia and severe cardiac dysfunction are common.
- **Mortality**: rises with age, primarily because of comorbid conditions.

Chronic kidney disease

Chronic kidney disease (CKD), previously termed chronic renal failure, refers to an irreversible deterioration in renal function which usually develops over a period of years (see Box 17.3, p. 467). Initially, it is manifest only as a biochemical abnormality but, eventually, loss of the excretory, metabolic and endocrine functions of the kidney leads to the clinical symptoms and signs of renal failure, collectively referred to as uraemia. When death is likely without RRT (CKD stage 5), it is called end-stage renal disease or failure (ESRD or ESRF).

Epidemiology

The social and economic consequences of CKD are considerable. In most countries, estimates of the prevalence of CKD stage 3–5 (eGFR < 60) are around 5–7%, mostly affecting people aged 65 years and above. The
prevalence of CKD in hypertension, diabetes and vascular disease is substantially higher, and targeted screening for CKD should be considered in these and other high-risk groups. The great majority of patients with earlier CKD (stages 1–3) never develop ESRD, which is fortunate, given the numbers (see Box 17.3).

Pathophysiology
Common causes of CKD are shown in Box 17.24. In many cases, the underlying diagnosis is unclear, especially among the large number of elderly patients with moderate GFR reductions (stage 3 CKD; see Box 17.3), in whom the additional information gained from a biopsy would not alter treatment. Many patients diagnosed at a late stage have bilateral small kidneys; renal biopsy is rarely undertaken in this group since it is more risky, less likely to provide a histological diagnosis because of the severity of damage, and unlikely to alter management.

Clinical features
The typical presentation is with a raised urea and creatinine found during routine blood tests, frequently accompanied by hypertension, proteinuria or anaemia. The rate of change in renal function varies between patients but is relatively constant for an individual and provides useful prognostic information. A plot of GFR, or of the reciprocal of the plasma creatinine concentration against time (Fig. 17.12), can demonstrate whether a patient has a stable GFR over time, predict when ESRF will be reached if decline is progressive, and can detect any unexpected worsening of kidney disease. It can also be used to monitor the success of interventions.

General symptoms
Most patients with slowly progressive disease are asymptomatic until GFR falls below 30 mL/min/1.73 m² (stage 4 or 5) and some can remain asymptomatic with much lower GFR values than this. An early symptom is nocturia, due to the loss of concentrating ability and increased osmotic load per nephron, but this is non-specific. When GFR falls below 15–20 mL/min/1.73 m², symptoms and signs are common and can affect almost all body systems (Fig. 17.13). They typically include tiredness or breathlessness, which may, in part, be related to renal anaemia, pruritus, anorexia, weight loss, nausea and vomiting. With further deterioration in renal function, patients may suffer hiccups, experience unusually deep respiration related to metabolic acidosis (Kussmaul’s respiration), and develop muscular twitching, fits, drowsiness and coma.
Chronic kidney disease

1 7
485

sodium salts, to prevent fluid depletion and worsening of renal function. Acidosis is common. Although it is usually asymptomatic, it may be associated with increased tissue catabolism and decreased protein synthesis, and may exacerbate bone disease and the rate of decline in renal function.

Endocrine function

A number of hormonal abnormalities may also be observed. In both genders, there is loss of libido related, at least in part, to hypogonadism as a consequence of hyperprolactinaemia (p. 790). The half-life of insulin is prolonged in CKD due to reduced tubular metabolism of insulin but there is also insulin resistance and reduced appetite. Because of this, insulin requirements are unpredictable in diabetic patients in advanced CKD.

Neurological and muscle function

Generalised myopathy may occur due to a combination of poor nutrition, hyperparathyroidism, vitamin D deficiency and disorders of electrolyte metabolism. Muscle cramps are common. The ‘restless leg syndrome’, in which the patient’s legs are jumpy during the night, may

Immune dysfunction

Cellular and humoral immunity is impaired in advanced CKD and there is increased susceptibility to infections, the second most common cause of death in dialysis patients, after cardiovascular disease. Many infections are associated with access devices but some are common infections, such as pneumonia.

Haematological

There is an increased bleeding tendency in advanced CKD, which manifests as cutaneous ecchymoses and mucosal bleeds. Platelet function is impaired and bleeding time prolonged. Adequate dialysis partially corrects the bleeding tendency in those with severe uraemia, but these patients are at significantly increased risk of complications from all anticoagulants, including those that are required during dialysis. Anaemia is common and is due in part to reduced erythropoietin production. Haemoglobin can be as low as 50–70 g/L in CKD stage 5, although it is often less severe or absent in patients with polycystic kidney disease. Several mechanisms are implicated, as summarised in Box 17.25.

Electrolyte abnormalities

Patients with CKD often develop electrolyte abnormalities and metabolic acidosis (p. 445). Fluid retention is common in advanced CKD and disproportionate fluid retention may occur in milder disease, sometimes leading to episodic pulmonary oedema. This is particularly associated with renal artery stenosis. Conversely, some patients with tubulo-interstitial disease can develop ‘salt-wasting’ disease and may require a high sodium and water intake, including supplements of...
be troublesome. Both sensory and motor neuropathy can arise, presenting as paraesthesia and foot drop, respectively, but appear late during the course of CKD. They are now unusual, given the widespread availability of RRT; if present, they can often improve once dialysis is established.

**Cardiovascular disease**

The risk of cardiovascular disease is substantially increased in patients with CKD stage 3 or worse (GFR < 60 mL/min/1.73 m²) and those with proteinuria or microalbuminuria. Left ventricular hypertrophy may occur, secondary to hypertension, and may account for the increased risk of sudden death (presumed to be caused by dysrhythmias) in this patient group. Pericarditis may complicate untreated or inadequately treated ESRD and cause pericardial tamponade or constrictive pericarditis. Medial vascular calcification is common and may be due, in part, to the high serum phosphate levels which are present in stage 3b CKD and above. Hyperphosphataemia may also contribute to the itching that arises in advanced CKD and probably contributes in an important way to the increased risk of cardiovascular disease. Reflecting this fact, serum fibroblast growth factor 23 (FGF23) levels (which increase in response to serum phosphate – see below) are an independent predictor of mortality in CKD.

**Metabolic bone disease**

Disturbances of calcium and phosphate metabolism are almost universal in advanced CKD, and various types of metabolic bone disease may also occur, including osteitis fibrosa cystica, osteomalacia and osteoporosis (Fig. 17.14). The sequence of events that leads to renal bone disease is complex. There is impaired conversion of 25-hydroxyvitamin D to its active metabolite, 1,25-dihydroxyvitamin D, due in part to renal tubular cell damage and elevated FGF23 levels. The reduced 1,25-dihydroxyvitamin D levels impair intestinal absorption of calcium, thereby causing hypocalcaemia, which leads in turn to increased PTH production by the parathyroid glands. Serum phosphate levels also start to rise because of the reduction in GFR. This leads to increased production of the hormone FGF23 from osteocytes (p. 1061). The FGF23 promotes phosphate excretion, thereby compensating in part for the reduced glomerular filtration of phosphate. This homeostatic response eventually fails, however, as renal failure progresses and hyperphosphataemia develops. The raised levels of serum phosphate complex with calcium in the extracellular space, causing ectopic calcification in blood vessels and other tissues. Patients with CKD commonly develop parathyroid gland hypertrophy and secondary hyperparathyroidism. In some cases, tertiary hyperparathyroidism supervenes, due to autonomous production of PTH by the enlarged parathyroid glands; this presents with hypercalcaemia. The histological picture in renal bone disease is complex. In osteitis fibrosa cystica, there is increased bone turnover due to the high levels of PTH, whereas low bone turnover (adynamic bone disease) may be observed in patients who have been over-treated with vitamin D metabolites.

**Fig. 17.14 Pathogenesis of renal osteodystrophy.** Low 1,25(OH)₂D levels cause calcium malabsorption and this, combined with high phosphate levels, causes hypocalcaemia, which increases PTH production by the parathyroid glands. The raised level of PTH increases osteoclastic bone resorption and bone formation. Although production of FGF23 from osteocytes also increases, promoting phosphate excretion, this is insufficient to prevent hyperphosphataemia in advanced CKD.
Osteomalacia can occur with overtreatment of hyperphosphataemia (p. 449).

**Investigations**

The recommended investigations in patients with CKD are shown in Box 17.26. Their main aims are:

- to identify the underlying cause where possible, since this may influence the treatment
- to identify reversible factors that may worsen renal function, such as hypertension, urinary tract obstruction, nephrotic drugs, and salt and water depletion
- to screen for complications of CKD, such as anaemia and renal osteodystrophy
- to screen for cardiovascular risk factors.

Referral to a nephrologist is appropriate for patients with potentially treatable underlying disease and those who are likely to progress to ESRD. Suggested referral criteria are listed in Box 17.27.

**Management**

The aims of management in CKD are to prevent or slow further renal damage; to limit the adverse physiological effects of renal impairment on the skeleton and on haematopoiesis; to treat risk factors for cardiovascular disease; and to prepare for RRT, if appropriate (p. 489).

**Antihypertensive therapy**

Lowering of blood pressure slows the rate at which renal function declines in CKD, independently of the agent used (Box 17.28), and has additional benefits in lowering the risk of hypertensive heart failure, stroke and peripheral vascular disease, as well as reducing proteinuria (see below). No threshold for beneficial effects has been identified and any reduction of blood pressure appears to be beneficial. Various targets have been suggested, such as 130/80 mmHg for uncomplicated CKD, and 125/75 mmHg for CKD complicated by significant proteinuria of more than 1 g/day (PCR > 100 mg/mmol or ACR > 70 mg/mmol). Achieving these blood pressure targets often requires multiple drugs, and therapeutic success may be limited by adverse effects and poor compliance.

**Reduction of proteinuria**

There is a clear relationship between the degree of proteinuria and the rate of progression of renal disease, and strong evidence that reducing proteinuria reduces the risk of progression. Angiotensin-converting enzyme (ACE) inhibitors and angiotensin II receptor blockers (ARBs) reduce proteinuria and retard the progression of CKD (Box 17.28). These effects are partly due to the reduction in blood pressure but there is evidence for a specific beneficial effect in patients with proteinuria (PCR > 50 mg/mmol or ACR > 30 mg/mmol) and those with incipient or overt diabetic nephropathy. In addition, ACE inhibitors have been shown to reduce the risk of cardiovascular events and all-cause mortality in CKD. Treatment with ACE inhibitors and ARBs may be accompanied by an immediate reduction in GFR when treatment is initiated, due to a reduction in glomerular perfusion pressure. Treatment can be continued so long
as the reduction in GFR is less than 20% and is not progressive. Accordingly ACE inhibitors and/or ARBs should be prescribed to all patients with diabetic nephropathy and those with proteinuria, irrespective of whether or not hypertension is present, providing that hyperkalaemia does not occur.

Dietary and lifestyle interventions
There is experimental evidence that restricting dietary protein can reduce progression of CKD in animal models but the results are less clear-cut in humans. All patients with stage 4 and 5 CKD should be given dietetic advice aimed at preventing excessive consumption of protein, ensuring adequate calorific intake and limiting potassium and phosphate intake. Severe protein restriction is not recommended. All patients should be advised to stop smoking, since there is evidence that this slows the decline in renal function in addition to reducing cardiovascular risk. Exercise and weight loss may also reduce proteinuria and have beneficial effects on cardiovascular risk profile.

Lipid-lowering therapy
Hypercholesterolaemia is almost universal in patients with significant proteinuria, and increased triglyceride levels are also common in patients with CKD. Lipid lowering has been shown to reduce vascular events in non-dialysis CKD patients (Box 17.29). There is some evidence that control of dyslipidaemia with statins may slow the rate of progression of renal disease.

Treatment of anaemia
Anaemia is common in patients with a GFR below 30 mL/min/1.73 m² and contributes to many of the non-specific symptoms of CKD. Recombinant human erythropoietin is effective in correcting the anaemia of CKD and improving the associated morbidity. Erythropoietin treatment does not influence mortality, however, and correcting haemoglobin to normal levels may carry some extra risk, including hypertension and thrombosis (including thrombosis of the arteriovenous fistulae used for haemodialysis). The target haemoglobin is usually between 100 and 120 g/L (10–20 g/dL). Erythropoietin is less effective in the presence of iron deficiency, active inflammation or malignancy, and in patients with aluminium overload, which may occur in dialysis.

Maintaining fluid and electrolyte balance
Patients with evidence of fluid retention should have dietary sodium intake limited to about 100 mmol/day, but often loop diuretics may also be required to treat fluid overload. If hyperkalaemia occurs, drug therapy should be reviewed, to reduce or stop potassium-sparing diuretics, ACE inhibitors and ARBs. Correction of acidosis may be helpful, and limiting potassium intake to about 70 mmol/day may be necessary in late CKD. Potassium-binding resins, such as calcium resonium, may be useful in the short term but should not be used chronically. The plasma bicarbonate should be maintained above 22 mmol/L by giving sodium bicarbonate supplements (starting dose of 1 g 3 times daily, increasing as required). If the increased sodium intake induces hypertension or oedema, calcium carbonate (up to 3 g daily) may be used as an alternative, since this has the advantage of also binding dietary phosphate.

Renal bone disease
Treatment should be initiated with active vitamin D metabolites (either 1-α-hydroxyvitamin D or 1,25-dihydroxyvitamin D) in patients who are found to have hypocalcaemia or serum PTH levels more than twice the upper limit of normal. The dose should be adjusted to try to reduce PTH levels to between 2 and 4 times the upper limit of normal to avoid over-suppression of bone turnover and adynamic bone disease, but care must be exercised in order to avoid hypercalcaemia. Hyperphosphataemia should be treated by dietary restriction of foods with high phosphate content (milk, cheese, eggs and protein-rich foods) and by the use of phosphate-binding drugs. Various drugs are available, including calcium carbonate, aluminium hydroxide, lanthanum carbonate and polymer-based phosphate binders such as sevelamer. The aim is to maintain serum phosphate values at 1.8 mmol/L (5.6 mg/dL) or below if possible, but many of these drugs are difficult to take and compliance may be a problem.

Parathyroidectomy may be required for the treatment of tertiary hyperparathyroidism. An alternative is to employ calcimimetic agents, such as cinacalcet, which bind to the calcium-sensing receptor and reduce PTH secretion. They have a place if parathyroidectomy is unsuccessful or not possible.

RENAL REPLACEMENT THERAPY

Renal replacement therapy (RRT) may be required on a temporary basis in patients with AKI or on a permanent
basis for those with CKD secondary to the many different types of renal disease discussed in this chapter.

Since the advent of long-term RRT in the 1960s, the numbers of patients with ESRD who are kept alive by dialysis and transplantation have increased considerably. For example, in the UK, there was a 3.3% increase per year between 2005 and 2010. By the end of 2010, over 51000 patients (832 per million) were on RRT, and the incidence of new patients starting RRT was 107 per million of the adult population. Whilst these numbers have been relatively steady over the last few years in the UK, they are still rising in the USA. The median age of patients starting RRT in the UK is now 65 years and 24% have a primary renal diagnosis of diabetic nephropathy. There are variations in the numbers of patients receiving RRT in different countries due to differences in the incidence of predisposing disease, as well as differences in medical practice. For example, in the USA, incidence rates for RRT in 2010 were about three times higher than in the UK at 348 per million population, and prevalence rates more than twice as high at 1752 per million population. Nearly half of these patients had a primary diagnosis of diabetes mellitus.

At the present time, RRT is commenced in about 1 in 10000 of the general population in the UK each year. Of patients starting dialysis in 2010, about 68% were still being treated with haemodialysis 3 months later – 18% by peritoneal dialysis, 8% by renal transplant (often before dialysis has been initiated) – and 6% had died or stopped treatment. Of prevalent patients established on RRT, about 48% had been transplanted, 44% were maintained on haemodialysis (approximately 1.5% on home haemodialysis) and 8% were on peritoneal dialysis.

Following initiation of dialysis in the UK, the survival is 84% at 1 year and 50% after 5 years. Mortality is strongly influenced by age; patients over 65 have a 76% survival at 1 year and 29% survival at 5 years, whereas corresponding figures for patients aged less than 65 are 91% and 71%, respectively. Although many young patients without extrarenal disease lead normal and active lives on RRT, those aged 30–34 have a mortality rate 25 times higher than age-matched controls. Comorbid conditions, such as diabetes mellitus (47% 5-year survival versus 66% for non-diabetics starting at age 45–64 years) and generalised vascular disease, also have a strong influence on mortality.

The aim of RRT is to replace the excretory functions of the kidney, and to maintain normal electrolyte concentrations and fluid balance. Various options are available, including haemodialysis, haemofiltration, haemodiafiltration, peritoneal dialysis and renal transplantation, and each of these is discussed in more detail below.

Preparation for renal replacement therapy

It is crucial that patients who are known to have progressive CKD are prepared well in advance for the institution of RRT. This involves ensuring that they are referred to a nephrologist in a timely manner, as those who are referred late, when they are either at the stage of or very close to requiring dialysis (about 20% of referrals in the UK), tend to have poorer outcomes. It is often possible to predict when RRT will be required from serial measurements of serum creatinine in patients with progressive CKD (see Box 17.3, p. 467). In such patients, preparations for the initiation of RRT should be started at least 12 months before the predicted start date.

At the present time, the average eGFR at the time of initiating RRT in the UK is about 8 mL/min/1.73 m² but there is wide variation. Since there is no evidence that early initiation of RRT improves outcome, the overall aim is to commence RRT by the time symptoms of CKD have started to appear but before serious complications have occurred. Preparation for RRT involves providing the patient with psychological and social support, assessing home circumstances and discussing the various choices of treatment (Fig. 17.15). Depression is common in patients who are on or approaching RRT, and support from the renal multidisciplinary team should be provided for both them and their relatives, to explain and help them adapt to the changes to lifestyle that may be necessary once RRT starts; this may help to reduce their anxieties about these changes.

Several decisions need to be taken in discussion with the patient and their family. The first is to decide whether RRT is an appropriate choice or whether conservative treatment might be preferable (p. 490). This is especially relevant in older people with significant comorbidity. For those that decide to go ahead with RRT, there are further choices between haemodialysis and peritoneal dialysis (Box 17.30), between hospital and home treatment, and on referral for renal transplantation.

### Comparison of haemodialysis and peritoneal dialysis

<table>
<thead>
<tr>
<th>Haemodialysis</th>
<th>Peritoneal dialysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient; 4 hrs three times per wk is usually adequate</td>
<td>Less efficient; 4 exchanges per day are usually required, each taking 30–60 mins (continuous ambulatory peritoneal dialysis) or 8–10 hrs each night (automated peritoneal dialysis)</td>
</tr>
<tr>
<td>Requires visits to hospital (although home treatment is possible for some patients)</td>
<td>Performed at home</td>
</tr>
<tr>
<td>Requires adequate venous circulation for vascular access</td>
<td>Requires an intact peritoneal cavity without major scarring from previous surgery</td>
</tr>
<tr>
<td>Careful compliance with diet and fluid restrictions required between treatments</td>
<td>Diet and fluid less restricted</td>
</tr>
<tr>
<td>Fluid removal compressed into treatment periods; may cause symptoms and haemodynamic instability</td>
<td>Slow continuous fluid removal, usually asymptomatic</td>
</tr>
<tr>
<td>Infections related to vascular access may occur</td>
<td>Peritonitis and catheter-related infections may occur</td>
</tr>
<tr>
<td>Patients are usually dependent on others</td>
<td>Patients can take full responsibility for their treatment</td>
</tr>
</tbody>
</table>
Kidney and urinary tract disease

Conservative treatment

In older patients with multiple comorbidities, conservative treatment of stage 5 CKD, aimed at limiting the adverse symptomatic effects of ESRD without commencing RRT, is increasingly viewed as a positive choice (Box 17.31). Current evidence suggests that survival of these patients without dialysis can be similar or only slightly shorter than that of patients who undergo RRT, but they avoid the hospitalisation and interventions associated with dialysis. Patients are offered full medical, psychological and social support to optimise and sustain their existing renal function and to treat complications, such as anaemia, for as long as possible, with appropriate palliative care in the terminal phase of their disease. Many of these patients enjoy a good quality of life for several years. It is also appropriate to discontinue dialysis treatment, with the consent of the patient, and to offer conservative therapy and palliative care when quality of life on dialysis is inadequate.

Haemodialysis

Haemodialysis is the most common form of RRT in ESRD and is also used in AKI. Haemodialysis involves gaining access to the circulation, either through an arteriovenous fistula, a central venous catheter or an arteriovenous shunt, such as a Scribner shunt. The patient’s blood is pumped through a haemodialyser, which allows bidirectional diffusion of solutes between blood and the dialysate across a semipermeable membrane down a concentration gradient (Fig. 17.15A). The composition of the dialysate can be varied to achieve the desired gradient and fluid can be removed by applying negative pressure to the dialysate side.
Haemodialysis offers the best rate of small solute clearance in AKI, as compared with other techniques such as haemofiltration, but should be started gradually because of the risk of confusion and convulsions due to cerebral oedema (dialysis disequilibrium). Typically, 1–2 hours of dialysis is prescribed initially but, subsequently, patients with AKI who are haemodynamically stable can be treated by 4–5 hours of haemodialysis on alternate days, or 2–3 hours every day. During dialysis, it is standard practice to anticoagulate patients with heparin but the dose may be reduced if there is a bleeding risk. Epoprostenol can be used as an alternative but carries a risk of hypotension. For short dialyses and in patients with abnormal clotting, it may be possible to avoid anticoagulation altogether. In AKI, dialysis is performed through a large-bore, dual-lumen catheter inserted into the femoral or internal jugular vein. Subclavian lines are avoided where possible, as thromboses or stenoses here will compromise the ability to form a functioning fistula in the arm if the patient fails to recover renal function and needs chronic dialysis.

Haemodialysis in CKD

In CKD, vascular access for haemodialysis is gained by formation of an arteriovenous fistula, usually in the forearm, up to a year before dialysis is contemplated. After 4–6 weeks, increased pressure transmitted from the artery to the vein leading from the fistula causes distension and thickening of the vessel wall (arterialisation). Large-bore needles can then be inserted into the vein to provide access for each haemodialysis treatment. Figure 17.16 shows a patient undergoing haemodialysis through such a fistula. Preservation of arm veins is thus very important in patients with progressive renal disease who may require haemodialysis in the future. If this access is not possible, venous or Gortex grafts may be fashioned or plastic catheters in central veins can be used for short-term access. These may be tunnelled under the skin to reduce infection risk. All patients must be screened in advance for hepatitis B, hepatitis C and human immunodeficiency virus (HIV), and vaccinated against hepatitis B if they are not immune. All dialysis units should have segregation facilities for hepatitis B-positive patients, given its easy transmissibility. Patients with hepatitis C and HIV are less infectious and can be treated satisfactorily using machine segregation and standard infection control measures.

Haemodialysis is usually carried out for 3–5 hours three times weekly, either at home or in an outpatient dialysis unit. The intensity and frequency of dialysis should be adjusted to achieve a reduction in urea during dialysis (urea reduction ratio) of over 65%. Most patients notice an improvement in symptoms during the first 6 weeks of treatment. Plasma urea and creatinine are lowered by each treatment but do not return to normal. The intensity of dialysis can be increased by escalating the number of standard sessions to four or more per week; by performing short, frequent dialysis sessions of 2–3 hours 5–7 times per week; or by performing nocturnal haemodialysis, when low blood-pump speeds and single-needle dialysis are used for approximately 8 hours overnight 5–6 times per week.

More frequent dialysis and nocturnal dialysis can achieve better fluid balance and electrolyte control than standard dialysis and, in particular, better control of serum phosphate levels. Studies are currently in progress to determine whether these different modes of dialysis improve clinical outcome. Box 17.32 summarises some of the problems related to haemodialysis.

Haemofiltration

This technique is principally used in the treatment of AKI. Water and solutes are filtered from blood across a porous semipermeable membrane under a pressure gradient. Replacement fluid of a suitable electrolyte composition is added to the blood after it exits from the haemofilter. If removal of fluid is required, then less
**17.32 Problems with haemodialysis**

<table>
<thead>
<tr>
<th>Problem</th>
<th>Clinical features</th>
<th>Cause</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotension during dialysis</td>
<td>Sudden ↓BP; often leg cramps; sometimes chest pain</td>
<td>Fluid removal and hypovolaemia</td>
<td>Saline infusion; exclude cardiac ischaemia; quinine may help cramp</td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>Hypotension, sometimes chest pain</td>
<td>Potassium and acid–base shifts</td>
<td>Check K⁺ and arterial blood gases; review dialysis prescription; stop dialysis</td>
</tr>
<tr>
<td>Haemorrhage</td>
<td>Blood loss (overt or occult); hypotension</td>
<td>Anticoagulation, Venous needle disconnection</td>
<td>Stop dialysis; seek source; consider heparin-free treatment</td>
</tr>
<tr>
<td>Air embolism</td>
<td>Circulatory collapse; cardiac arrest</td>
<td>Disconnected or faulty lines and equipment malfunction</td>
<td>Stop dialysis</td>
</tr>
<tr>
<td>Dialyser hypersensitivity</td>
<td>Acute circulatory collapse</td>
<td>Allergic reaction to dialysis membrane or sterilisant</td>
<td>Stop dialysis; change to different artificial kidney</td>
</tr>
<tr>
<td>Between treatments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary oedema</td>
<td>Breathlessness</td>
<td>Fluid overload</td>
<td>Ultrafiltration ± dialysis</td>
</tr>
<tr>
<td>Systemic sepsis</td>
<td>Rigors; fever; ↓BP</td>
<td>Usually involves vascular access devices (catheter or fistula)</td>
<td>Blood cultures; antibiotics</td>
</tr>
</tbody>
</table>

Fluid is added back than is removed (Fig. 17.15B). Haemofiltration may be either intermittent or continuous, and typically 1–2 litres of filtrate is replaced per hour (equivalent to a GFR of 15–30 mL/min); higher rates of filtration may be of benefit in patients with sepsis and multi-organ failure. In continuous arteriovenous haemofiltration (CAVH), the extracorporeal blood circuit is driven by the arteriovenous pressure difference, but poor filtration rates and clotting of the filter are common and this treatment has fallen out of favour. Continuous venovenous haemofiltration (CVVH) is pump-driven, providing a reliable extracorporeal circulation. Issues concerning anticoagulation are similar to those for haemodialysis, but may be more problematic because longer or continuous anticoagulation is necessary.

**Haemodiafiltration**

This technique combines haemodialysis with approximately 20–30 litres of ultrafiltration (with replacement of filtrate) over a 3–5-hour treatment. It uses a large-pore membrane and combines the improved middle-molecule clearance of haemofiltration with the higher small-solute clearance of haemodialysis. It is sometimes used in the treatment of AKI but is increasingly favoured in the treatment of CKD. It is more expensive than haemodialysis, however, and the long-term benefits are not yet established.

**Peritoneal dialysis**

Peritoneal dialysis is principally used in the treatment of CKD. It requires the insertion of a permanent Silastic catheter into the peritoneal cavity (Fig. 17.15C). Two types are in common use. In continuous ambulatory peritoneal dialysis (CAPD), about 2 litres of sterile, isotonic dialysis fluid are introduced and left in place for approximately 4–6 hours. Metabolic waste products diffuse from peritoneal capillaries into the dialysis fluid down a concentration gradient. The fluid is then drained and fresh dialysis fluid introduced, in a continuous four-times-daily cycle. The inflow fluid is rendered hyperosmolar by the addition of glucose or glucose polymer; this results in net removal of fluid from the patient during each cycle, due to diffusion of water from the blood through the peritoneal membrane down an osmotic gradient (ultrafiltration). The patient is mobile and able to undertake normal daily activities. Automated peritoneal dialysis (APD) is similar to CAPD but uses a mechanical device to perform the fluid exchanges during the night, leaving the patient free, or with only a single exchange to perform, during the day.

CAPD is particularly useful in children, as a first treatment in adults with residual renal function, and as a treatment for elderly patients with cardiovascular instability. The long-term use of peritoneal dialysis may be limited by episodes of bacterial peritonitis and damage to the peritoneal membrane, including sclerosing peritonitis, but some patients have been treated successfully for more than 10 years. Box 17.33 summarises some of the problems related to CAPD treatment.

**Renal transplantation**

Renal transplantation offers the best chance of long-term survival in ESRD, and is the most cost-effective treatment. Transplantation can restore normal kidney function and correct all the metabolic abnormalities of CKD but requires long-term immunosuppression with its attendant risks (see below). All patients with ESRD should be considered for transplantation, unless there are contraindications (Box 17.34).

Kidney grafts may be taken from a cadaver in the UK after brain death (51%) or circulatory death (11%), or from a living donor (38%). As described on page 94, matching of a donor to a specific recipient is strongly influenced by immunological factors, since graft rejection is the major cause of transplant failure. Compatibility of ABO blood group between donor and recipient is usually required and the degree of matching for major histocompatibility (MHC) antigens, particularly HLA-DR, influences the incidence of rejection. Immediately prior to transplantation, tests should be performed for
17.33 Problems with continuous ambulatory peritoneal dialysis

<table>
<thead>
<tr>
<th>Problem</th>
<th>Clinical features</th>
<th>Cause</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peritonitis</td>
<td>Cloudy drainage fluid; abdominal pain and systemic sepsis are variable</td>
<td>Usually entry of skin contaminants via catheter; bowel organisms less common</td>
<td>Culture of peritoneal dialysis fluid&lt;br&gt; Intraperitoneal antibiotics, tobramycin, vancomycin&lt;br&gt; Catheter removal sometimes required</td>
</tr>
<tr>
<td>Catheter exit site infection</td>
<td>Erythema and pus around exit site</td>
<td>Usually skin organisms</td>
<td>Antibiotics; sometimes surgical drainage</td>
</tr>
<tr>
<td>Ultrafiltration failure</td>
<td>Fluid overload</td>
<td>Damage to peritoneal membrane, leading to rapid transport of glucose and loss of osmotic gradient</td>
<td>Replace glucose with synthetic, poorly absorbed polymers for some exchanges (icodextrin)</td>
</tr>
<tr>
<td>Peritoneal membrane failure</td>
<td>Inadequate clearance of urea etc.</td>
<td>Scarring/damage to peritoneal membrane</td>
<td>Increase exchange volumes; consider automated peritoneal dialysis or switch to haemodialysis</td>
</tr>
<tr>
<td>Sclerosing peritonitis</td>
<td>Intermittent bowel obstruction Malnutrition</td>
<td>Unknown; typically occurs after many years</td>
<td>Switch to haemodialysis (may still progress) Surgery and tamoxifen may be used</td>
</tr>
</tbody>
</table>

17.34 Contraindications to renal transplantation

Absolute

- Active malignancy: a period of at least 2 yrs of complete remission recommended for most tumours
- Active vasculitis or recent anti-GBM disease
- Severe heart disease
- Severe occlusive aorto-iliac vascular disease

Relative

- Age: while practice varies, transplants are not routinely offered to very young children (< 1 yr) or older people (> 75 yrs)
- High risk of disease recurrence in the transplant kidney
- Disease of the lower urinary tract: in patients with impaired bladder function, an ileal conduit may be considered
- Significant comorbidity

- Primary graft non-function. Causes include hypovolaemia, preservation injury/acute tubular necrosis during storage and transfer, other pre-existing renal damage, hyperacute rejection, vascular occlusion and urinary tract obstruction.
- Sepsis. In addition to risks of sepsis associated with any operation, there is an increased risk due to the uremia and immunosuppression.

Once the graft begins to function, near-normal biochemistry is usually achieved within days to weeks. The median eGFR of patients in the UK receiving a deceased donor transplant at 1 year is 51 mL/min/1.73 m².

All transplant patients require regular life-long follow-up to monitor renal function and immunosuppression. Life-long immunosuppressive therapy (see Box 4.25, p. 96) is required to prevent rejection but needs to be more intensive in the early post-transplantation period, when the risk is highest. A common regimen is triple therapy with prednisolone; cyclosporin or tacrolimus; and azathioprine or mycophenolate mofetil. Sirolimus (rapamycin) is an alternative that can be introduced later. Antibodies to deplete or modulate specific lymphocyte populations are increasingly used; targeting the lymphocyte interleukin (IL)-2 receptor is particularly effective for preventing rejection. Acute rejection is usually treated, in the first instance, by short courses of high-dose corticosteroids, such as methylprednisolone 500 mg IV on 3 consecutive days. Other therapies, such as antilymphocyte antibodies, intravenous immunoglobulin and plasma exchange, can be used for episodes of acute rejection that do not respond to high-dose corticosteroids.

Complications of immunosuppression include infections and malignancy (p. 95). Approximately 50% of white patients develop skin malignancy by 15 years after transplantation.

The prognosis after kidney transplantation is good. Recent UK statistics for transplants from cadaver donors indicate 96% patient survival and 93% graft survival at 1 year, and 88% patient survival and 84% graft survival at 5 years. Even better figures are obtained with living donor transplantation (91% graft survival at 5 years).
Advances in immunosuppression have greatly improved results from using genetically unrelated donors, such as spouses.

**RENAL VASCULAR DISEASES**

Diseases which affect renal blood vessels may cause renal ischaemia, leading to acute or chronic kidney disease or secondary hypertension. The rising prevalence of atherosclerosis and diabetes mellitus in ageing populations has made renovascular disease an important cause of ESRD.

**Renal artery stenosis**

Renal artery stenosis is a relatively uncommon disorder, which presents clinically with hypertension. It has been estimated to occur in about 2% of unselected patients with hypertension but may affect up to 4% of older patients with hypertension who have evidence of atherosclerotic disease elsewhere. Most cases of renal artery stenosis are caused by atherosclerosis but fibromuscular dysplasia involving the vessel wall may be responsible in younger patients. Rare causes include vasculitis, thromboembolism and aneurysms of the renal artery.

**Pathophysiology**

Renal artery stenosis results in a reduction in renal perfusion pressure, which activates the renin-angiotensin system, leading to increased circulating levels of angiotensin II. This provokes vasoconstriction and increases aldosterone production by the adrenal, causing sodium retention by the renal tubules. Significant reduction of renal blood flow occurs when there is greater than 70% narrowing of the artery, and this is commonly associated with a dilated region more distally (post-stenotic dilatation). Atherosclerosis is the most common cause. The characteristic lesion is an ostial stenosis that is associated with atherosclerosis within the aorta and affects other major branches, particularly the iliac vessels. The picture is often complicated by small-vessel disease in affected kidneys, which may be related to subclinical atheroemboli or other vascular disease. As the stenosis becomes more severe, global renal ischaemia leads to shrinkage of the affected kidney and may cause renal failure (ischaemic nephropathy). However, the progression of stenosis is not easily predictable, and many patients die from coronary, cerebral or other vascular disease rather than renal failure.

In younger patients, fibromuscular dysplasia is a more likely cause of renal artery stenosis. This is an uncommon disorder of unknown cause. It is characterised by hypertrophy of the media (medial fibroplasia), which narrows the artery but rarely leads to total occlusion. It may be associated with disease in other arteries; for example, those who have carotid artery dissections are more likely to have renal arteries with this appearance. It most commonly presents with hypertension in patients aged 15–30 years, and women are affected more frequently than men. Irregular narrowing (beading) may occur in the distal renal artery and this sometimes extends into the intrarenal branches of the vessel. Rarely, renal artery stenosis may occur as a complication of large-vessel vasculitis, such as Takayasu’s arteritis and polyarteritis nodosa (pp. 1116 and 1117). Untreated, atheromatous renal artery stenosis is thought to progress to complete arterial occlusion in about 15% of cases. This figure increases with more severe degrees of stenosis. If the progression is gradual, collateral vessels may develop and some function may be preserved, preventing infarction and loss of kidney structure. Conversely, at least 85% of patients with renal artery stenosis will not develop progressive renal impairment, and indeed, in many patients, the stenosis may be haemodynamically insignificant and not responsible for coexisting essential hypertension. Unfortunately, methods of predicting which patients are at risk of progression or who will respond to treatment are still imperfect.

**Clinical features**

Renal artery stenosis can present in various ways (Box 17.35), including hypertension, renal failure (with bilateral disease), a deterioration in renal function when ACE inhibitors or ARBs are used, or acute pulmonary oedema. Although many patients experience a slight drop in GFR when commencing these drugs, an increase in serum creatinine of 25% or more raises the possibility of renal artery stenosis. Acute pulmonary oedema is particularly characteristic of bilateral renovascular disease. It is associated with severe hypertension, occurring without other obvious cause in patients with normal or only mildly impaired renal function. Clinical evidence of generalised vascular disease may be observed, particularly in the legs, in older patients with atherosclerotic renal artery stenosis. Clinical features associated with an increased risk of renal artery stenosis in hypertensive patients are summarised in Box 17.35. The presence of one or more of these features should prompt investigation for possible renal artery stenosis, as described below, provided that intervention is being contemplated to improve renal perfusion.

**Investigations**

When appropriate, imaging of the renal vasculature with either CT angiography or MR angiography should be performed to confirm the diagnosis (Fig. 17.17). Both give good views of the main renal arteries, the vessels predominantly involved and the most amenable to intervention. Biochemical testing may reveal impaired renal function and an elevated plasma renin activity, sometimes with hypokalaemia due to hyperaldosteronism. Ultrasound may also reveal a discrepancy in size between the two kidneys. While these investigations

<table>
<thead>
<tr>
<th>17.35 Presentation and clinical features of renal artery stenosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal artery stenosis is more likely if:</td>
</tr>
<tr>
<td>• hypertension is severe, of recent onset or difficult to control</td>
</tr>
<tr>
<td>• kidneys are asymmetrical in size</td>
</tr>
<tr>
<td>• flash pulmonary oedema occurs repeatedly*</td>
</tr>
<tr>
<td>• there is peripheral vascular disease of the lower limbs</td>
</tr>
<tr>
<td>• there is renal impairment*</td>
</tr>
<tr>
<td>• renal function has deteriorated on ACE inhibitors or angiotensin II receptor blockers</td>
</tr>
</tbody>
</table>

*Particularly with bilateral disease.
vascular disease in hypertensive patients.

provide supportive information, they are insufficiently sensitive or specific to be of value in diagnosis of renovascular disease in hypertensive patients.

**Management**

The first-line management in patients with renal artery stenosis is medical therapy with antihypertensive drugs, supplemented, where appropriate, by statins and low-dose aspirin in those with atherosclerotic disease. Interventions to correct the vessel narrowing should be considered in young patients (age below 40) suspected of having renal artery stenosis; those in whom blood pressure cannot easily be controlled with antihypertensive agents; those who have a history of ‘flash’ pulmonary oedema or accelerated phase (malignant) hypertension; and those in whom renal function is deteriorating. The most commonly used technique is angioplasty. The best results are obtained in non-atheromatous fibromuscular dysplasia, where correction of the stenosis has a high chance of success in improving blood pressure and protecting renal function. Angioplasty and stenting can sometimes be successful in atherosclerotic disease but randomised trials have produced no convincing evidence for overall benefit in terms of renal function or blood pressure control (Box 17.36). The risks of angioplasty and stenting include renal artery occlusion, renal infarction, and atheroemboli (p. 496) from manipulations in a severely diseased aorta. Small-vessel disease distal to the stenosis may preclude substantial functional recovery.

**Acute renal infarction**

This is an uncommon condition that occurs as the result of sudden occlusion of the renal arteries. The presentation is typically with loin pain of acute onset, usually in association with haematuria detected on dipstick testing or urine microscopy, but pain may be absent in some cases. Severe hypertension is common but not universal, presumably because some residual renal perfusion is required to generate renin release. Blood levels of lactate dehydrogenase (LDH) and CRP are commonly raised. The condition may be caused by local atherosclerosis (atheroembolic) or by thromboemboli from a distant source, where occlusion may occur in branch arteries distal to the main renal artery. This can cause multiple infarcts within the renal parenchyma of both kidneys, which may be visualised by CT scanning. If occlusion of the main renal arteries is bilateral or if there is occlusion of a single functioning kidney, the presentation is with AKI and the patient is typically anuric. Patients with bilateral occlusion usually have evidence of widespread vascular disease and may show evidence of aortic occlusion, with absent femoral pulses and reduced lower limb perfusion. Management is largely supportive, and includes anticoagulation if a source of thromboembolism is identified. It is sometimes possible to carry out stenting of an acutely blocked main renal artery to try to restore renal blood flow and kidney function.

**Diseases of small intrarenal vessels**

A number of conditions are associated with acute damage and occlusion of small blood vessels (arterioles and capillaries) in the kidney (Box 17.37). They are often found in conjunction with similar changes elsewhere in the body. A common feature of these syndromes is microangiopathic haemolytic anaemia, in which haemolysis and red cell fragmentation arise as consequences of damage incurred to red blood cells during passage through the abnormal vessels.

Haemolytic uraemic syndrome

Haemolytic uraemic syndrome (HUS) is characterised by thrombotic microangiopathy that causes damage to...
endothelial cells of the microcirculation. This is accompanied by swelling of the endothelial cells, increased platelet adherence and intravascular thrombosis. There is a marked reduction in the platelet count and anaemia, with features of intravascular haemolysis (p. 1026), such as a raised unconjugated bilirubin level, raised levels of LDH and decreased circulating levels of haemoglobin. A reticulocytosis is often seen. The kidney microcirculation tends to be most affected in HUS, with involvement of other organs (including the brain) in more severe cases. The most common cause of classical HUS is infection associated with organisms that produce enterotoxins called verotoxins: so-called diarrhoea-positive HUS (D+HUS). In about 10% of cases, however, no infective cause can be identified and this is termed atypical HUS.

The organisms most commonly implicated in D+HUS are enterohaemorrhagic Escherichia coli (p. 342). The O157:H7 serotype is the best known but other serotypes that produce verotoxins can also be responsible. Although these bacteria live as commensals in the gut of cattle and other domestic livestock, they can cause haemorrhagic diarrhoea in humans when the infection is contracted from contaminated food products, water or other infected individuals. In a proportion of cases, verotoxin produced by the organisms enters the circulation and binds to specific glycolipid receptors that are expressed on the surface of microvascular endothelial cells. In developed countries, D+HUS is now the most common cause of AKI in children. Recovery is good in most patients but sometimes RRT may be required for up to 14 days. No other specific treatments have been shown to help the renal lesion.

Atypical HUS is subclassified into familial and sporadic subtypes but both are associated with abnormalities of the complement system. Up to 70% of sporadic cases are associated with the development of autoantibodies to complement factor H, whereas the inherited forms are due to mutations in various genes that encode components of the complement cascade, including factor H (CFH), factor B (CFB) and complement component 3 (C3). The penetrance of familial HUS is incomplete, indicating that environmental triggers are also involved, but the nature of these triggers is poorly understood. Management of atypical HUS involves supportive care, optimising fluid and electrolyte balance, transfusion and RRT if necessary. Infusion of fresh frozen plasma can be helpful (presumably by replacing complement components), as can plasma exchange (presumably by removing pathogenic autoantibodies). Recently, impressive results have been reported with the anti-C5 monoclonal antibody, eculizumab, which binds to C5, thereby preventing activation of the terminal complement cascade.

**Thrombotic thrombocytopenic purpura**

Like HUS, thrombotic thrombocytopenic purpura (TTP) is characterised by thrombotic microangiopathy, which causes damage to endothelial cells of the microcirculation. This leads to swelling of the endothelial cells, increased platelet adherence and intravascular thrombosis. In contradistinction to HUS, the brain is more commonly affected in TTP and involvement of the kidney is usually less prominent. TTP is an autoimmune disorder caused by antibodies against ADAMTS-13, which is involved in regulating platelet aggregation. More details are provided on page 1056.

**Disseminated intravascular coagulation**

This may occur as a complication of a range of illnesses and can be accompanied by multi-organ failure and renal failure. Pathogenesis and management are discussed in more detail on pages 201 and 1055.

**Systemic sclerosis**

Renal involvement is a serious complication of systemic sclerosis, which is more likely to occur in diffuse cutaneous systemic sclerosis (DCSS) than in limited cutaneous systemic sclerosis (LCSS). The renal lesion is caused by intimal cell proliferation and luminal narrowing of intrarenal arteries and arterioles. There is intense intrarenal vasospasm and plasma renin activity is markedly elevated. Renal involvement usually presents clinically with severe hypertension, microangiopathic features and progressive oliguric renal failure (scleroderma renal crisis). Use of ACE inhibitors to control the hypertension has improved the 1-year survival from 20% to 75% but about 50% of patients continue to require RRT. Onset or acceleration of the syndrome after stopping ACE inhibitors is now well described.

**Cholesterol emboli**

These present with renal impairment, haematuria, proteinuria and sometimes eosinophilia with inflammatory features that can mimic a small-vessel vasculitis. The symptoms are provoked by showers of cholesterol-containing microemboli, arising in atheromatous plaques in major arteries. The diagnosis should be suspected when these clinical features occur in patients with widespread atheromatous disease, who have undergone interventions such as surgery or arteriography. It may also be precipitated by anticoagulants and thrombolytic agents. On clinical examination, signs of large-vessel disease and microvascular occlusion in the lower limbs (ischaemic toes, livedo reticularis) are common but not invariable (Fig. 17.18). There is no specific treatment.

![Fig. 17.18 The foot of a patient who suffered extensive atheroembolism following coronary artery stenting.](image)

**Small-vessel vasculitis**

Renal disease caused by small-vessel vasculitis usually affects the glomeruli, as described in the next section and on page 519.
Accelerated phase hypertension

Accelerated phase hypertension (p. 609) is deemed to exist when it is associated with acute damage to renal and other arterioles. It is often symptomatic, with headache, impaired vision and renal impairment (Fig. 17.19). Physical examination reveals severe hypertension with evidence of hypertensive retinopathy. Papilloedema is almost always present. Clinical features of microangiopathy, such as fragmented red cells and anaemia, may be present and, in the absence of a previous history, it may be difficult to distinguish patients with accelerated phase hypertension from those with HUS and hypertension. Most patients respond to effective control of blood pressure, although renal function is permanently lost in 20% of cases.

Fig. 17.19 Glomerular capillary thrombosis in malignant hypertension. Similar changes occur in thrombotic microangiopathy. The adjacent arteriole (arrow) shows gross intimal thickening. From Beutler and Koomans 1977 – see p. 523.
Kidney and urinary tract disease

17.38 Clinical and laboratory features of glomerular injury

- Leakage of cells and macromolecules across the glomerular filtration barrier
  - Proteinuria: characteristic of diseases that affect the podocyte, scarring and deposition of foreign material
  - Haematuria: characteristic of inflammatory and destructive processes
- Impaired renal function and reduced GFR
- Hypertension

(p. 502), metabolic diseases such as diabetes mellitus (p. 830), and deposition of abnormal proteins such as amyloid in the glomeruli. The cell types of the glomerulus that may be the target of injury are shown in Figure 17.20. The response of the glomerulus to injury and hence the predominant clinical features vary according to the nature of the insult (Fig. 17.21). Clinical and laboratory features common to many glomerular diseases are shown in Box 17.38. Most patients with glomerular disease do not present acutely and are asymptomatic until abnormalities are detected on routine screening of blood or urine samples.

Glomerulonephritis

Glomerulonephritis literally means ‘inflammation of glomeruli’. The term is used to describe all types of glomerular disease, even though some of these (such as minimal change nephropathy) are not associated with inflammation.

Most types of glomerulonephritis are immunologically mediated and several respond to immunosuppressive drugs. Deposition of antibody occurs in many types of glomerulonephritis (Box 17.39) but, frequently, the presumed mechanisms involve cellular immunity, which is more difficult to investigate. Although deposition of circulating immune complexes was previously thought to be a common mechanism, it now seems that most granular deposits of immunoglobulin are formed ‘in situ’ by antibodies which complex with glomerular antigens, or with other extraneous antigens derived from viruses and bacteria that have become deposited in the glomeruli (see Fig. 17.20).

Glomerulonephritis is generally classified in terms of the histopathological appearances, as summarised in Box 17.39 and Figure 17.22. The most common subtypes are discussed in more detail below.

**Minimal change nephropathy**

Minimal change disease occurs at all ages but accounts for nephrotic syndrome (see Box 17.11, p. 475) in most children and about one-quarter of adults. It is caused by reversible dysfunction of podocytes. The presentation is with proteinuria or nephrotic syndrome, which typically remits with high-dose corticosteroid therapy (1 mg/kg prednisolone for 6 weeks). Some patients who respond incompletely or relapse frequently need maintenance

![Histopathology of glomerular disease](image)
<table>
<thead>
<tr>
<th>Histology</th>
<th>Immune deposits</th>
<th>Pathogenesis</th>
<th>Associations</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal change</td>
<td>None</td>
<td>Unknown</td>
<td>Atopy, HLA-DR7, Drugs</td>
<td>Acute and often severe nephrotic syndrome, Good response to corticosteroids, Dominant cause of idiopathic nephrotic syndrome in childhood</td>
</tr>
<tr>
<td>Focal segmental glomerulosclerosis (FSGS)</td>
<td>Non-specific trapping in focal scars</td>
<td>Unknown; in some, circulating factors increase glomerular permeability Injury to podocytes may be a common feature</td>
<td>Healing of previous local glomerular injury, HIV infection, Heroin misuse, Morbid obesity</td>
<td>Primary FSGS presents as idiopathic nephrotic syndrome but is less responsive to treatment than minimal change; may progress to renal impairment, can recur after transplantation Secondary FSGS presents with variable proteinuria and outcome</td>
</tr>
<tr>
<td>Focal segmental glomerulonephritis</td>
<td></td>
<td>Small-vessel vasculitis</td>
<td>Primary or secondary small-vessel vasculitis</td>
<td>Often occurs in systemic disease. Responds to treatment with corticosteroids and immunosuppressants Check ANCA, ANA</td>
</tr>
<tr>
<td>Membranous glomerulonephritis</td>
<td>Granular subepithelial IgG</td>
<td>Antibodies to a podocyte surface antigen, with complement-dependent podocyte injury</td>
<td>HLA-DR3 (for idiopathic) Drugs, Mercury, heavy metals, Hepatitis B virus, Malignancy</td>
<td>Usually idiopathic; common cause of adult idiopathic nephrotic syndrome One-third progress; may respond to corticosteroids and immunosuppressants</td>
</tr>
<tr>
<td>IgA nephropathy</td>
<td>Mesangial IgA</td>
<td>Unknown</td>
<td>Usually idiopathic Liver disease</td>
<td>Common disease with a range of presentations, usually including haematuria and hypertension</td>
</tr>
<tr>
<td>Mesangiocapillary glomerulonephritis</td>
<td>Immunoglobulins</td>
<td>Deposition of circulating immune complexes or ‘planted’ antigens</td>
<td>Infections, autoimmunity, or monoclonal immunoglobulin-related</td>
<td>Most common pattern found in association with subacute bacterial infection, but also with Cryoglobulinaemia ± hepatitis C virus, and others In dense deposit disease, intramembranous deposits No proven treatments</td>
</tr>
<tr>
<td>Complement type</td>
<td>Complement components</td>
<td>Complement abnormalities, inherited or acquired, Dense deposit disease is associated with abnormal activation of the alternative complement pathway</td>
<td>Complement gene mutations, C3 nephritic factor and partial lipodystrophy</td>
<td></td>
</tr>
<tr>
<td>Post-infection</td>
<td>Subendothelial</td>
<td>Immune response to streptococcal infection with presumed cross-reactive epitopes</td>
<td>Streptococcal and other infections</td>
<td>Now rare in developed countries Presents with severe sodium and fluid retention, hypertension, haematuria, oliguria Usually resolves spontaneously</td>
</tr>
</tbody>
</table>
corticosteroids, cytotoxic therapy or other agents. Minimal change disease does not progress to CKD but can present with problems related to the nephrotic syndrome and complications of treatment.

**Focal segmental glomerulosclerosis**

Primary focal segmental glomerulosclerosis (FSGS) (Fig. 17.22B) can occur in all age groups. In some patients, FSGS can have specific causes, such as HIV infection, podocyte toxins and massive obesity, but in most cases the underlying cause is unknown (primary FSGS). Patients with primary FSGS present with massive proteinuria and idiopathic nephrotic syndrome. Histological analysis shows sclerosis affecting segments of the glomeruli, which may also show positive staining for deposits of C3 and IgM on immunofluorescence. Since FSGS is a focal process, abnormal glomeruli may not be seen on renal biopsy if only a few are sampled, leading to an initial diagnosis of minimal change nephropathy. Juxtamedullary glomeruli are more likely to be affected in early disease. Although nephrotic syndrome is typical, some patients present with the histological features of FSGS but less proteinuria. In these patients, the focal scarring may reflect healing of previous focal glomerular injury, such as HUS, cholesterol embolism or vasculitis. These examples of secondary FSGS have different course and treatments.

Primary FSGS can respond to high-dose corticosteroid therapy (0.5–2.0 mg/kg/day) but most patients show little or no response. Immunosuppressive drugs, such as cyclosporin, cyclophosphamide and mycophenolate mofetil, have also been used but their efficacy is uncertain. Progression to CKD is common in patients who do not respond to steroids and the disease frequently recurs after renal transplantation, with an almost immediate return of proteinuria following transplant in some cases.

**Membranous glomerulonephritis**

Membranous glomerulonephritis, also known as membranous nephropathy, is the most common cause of nephrotic syndrome in adults. It is caused by antibodies (usually autoantibodies) directed at antigen(s) expressed on the surface of podocytes. Recent studies suggest that one such antigen is the M-type phospholipase A2 receptor 1. A proportion of cases are associated with other causes, such as heavy metal poisoning, drugs, infections and tumours (see Box 17.39 and Fig. 17.22D and F) but most are idiopathic. Approximately one-third of patients with idiopathic membranous glomerulonephritis undergo spontaneous remission; one-third remain in a nephrotic state, and one-third go on to develop CKD. Short-term treatment with high doses of corticosteroids and cyclophosphamide may improve both the nephrotic syndrome and the long-term prognosis. However, because of the toxicity of these regimens, many nephrologists reserve such treatment for those with severe nephrotic syndrome or deteriorating renal function.

**IgA nephropathy**

This is one of the most common types of glomerulonephritis and can present in many ways (Fig. 17.22G). Haematuria is the earliest sign and is almost universal, and hypertension is also very common. Proteinuria can also occur but is usually a later feature. In many cases, there is slowly progressive loss of renal function leading to ESRD. Clinical presentations are protean and vary with age. A particular hallmark of IgA nephropathy in young adults is the occurrence of acute self-limiting exacerbations, often with gross haematuria, in association with minor respiratory infections. This may be so acute as to resemble acute post-infectious glomerulonephritis, with fluid retention, hypertension and oliguria with dark or red urine. Characteristically, the latency from clinical infection to nephritis is short: a few days or less. Asymptomatic presentations dominate in older adults, with haematuria, hypertension and loss of GFR. Occasionally, IgA nephropathy progresses rapidly and crescent formation may be seen. The response to immunosuppressive therapy is usually poor. The management of less acute disease is largely directed towards the control of blood pressure in an attempt to prevent or retard progressive renal disease. There is some evidence

---

**Histology Immune deposits Pathogenesis Associations Comments**

| Goodpasture’s disease | Usually crescentic nephritis | Linear IgG along GBM | Autoantibodies to α3 chain of type IV collagen in the GBM | HLA-DR15 (previously known as DR2) | Associated with lung haemorrhage but either may occur alone. Treat with corticosteroids, cyclophosphamide and plasma exchange |
|-----------------------|-----------------------------|-------------------|--------------------------------------------------------|-----------------------------------|************************************************************************************************************************|
| Lupus nephritis       | Almost any histological type | Always positive and often profuse | Some anti-DNA antibodies also bind to glomerular targets | Complement deficiencies Complement consumption | Variable presentation, sometimes as renal disease alone without systemic features Responds to cytotoxic therapy in addition to prednisolone |

(ANCA = antineutrophil cytoplasmic antibodies; ANA = antinuclear antibody; HLA = human leucocyte antigen)
for additional benefit from several months of high-dose corticosteroid treatment in high-risk disease, but no strong evidence for other immunosuppressive agents. Other therapies are under investigation.

Hench–Schönlein purpura
This condition most commonly occurs in children but can also be observed in adults. It is characterised by a systemic vasculitis that often arises in response to an infectious trigger. The presentation is with a characteristic petechial rash typically affecting buttocks and lower legs, and abdominal pain due to the occurrence of vasculitis involving the gastrointestinal tract. The presence of glomerulonephritis is usually indicated by the occurrence of haematuria. When Hench–Schönlein purpura occurs in older children or adults, the glomerulonephritis is usually more prominent and less likely to resolve completely. Renal biopsy shows mesangial IgA deposition and appearances that are indistinguishable from acute IgA nephropathy. Treatment is supportive in nature; in most patients, the prognosis is good, with spontaneous resolution, but some, particularly adults, progress to develop ESRD.

Mesangiocapillary glomerulonephritis
Mesangiocapillary glomerulonephritis (MCGN), also known as membranoproliferative glomerulonephritis (MPGN), is characterised by an increase in mesangial cellularity with thickening of glomerular capillary walls and subendothelial deposition of immune complexes and/or components of the complement pathway. The typical presentation is with proteinuria and haematuria. Several underlying causes have been identified, as summarised in Box 17.39. It can be classified into two main subtypes. The first is characterised by deposition of immunoglobulins within the glomeruli. This subtype is associated with chronic infections, autoimmune diseases and monoclonal gammopathy. The second is characterised by deposition of complement in the glomeruli and is associated with inherited or acquired abnormalities in the complement pathway. Within this category is so-called ‘dense deposit disease’, which is typified by deposition of electron-dense deposits within the GBM. A third subtype is recognised, in which neither immunoglobulins nor complement are deposited in the glomeruli. This is associated with healing following thrombotic microangiopathies, such as HUS and TTP.

Treatment of MCGN associated with immunoglobulin deposits consists of identifying and treating the underlying disease, if possible, and the use of immunosuppressive drugs such as mycophenolate mofetil or cyclophosphamide. There is no specific treatment for MCGN associated with deposition of complement in the glomeruli or for dense deposit disease.

Infection-related glomerulonephritis
Glomerulonephritis may occur in connection with infections of various types, including subacute bacterial endocarditis. The most common histological pattern in bacterial infection is mesangiocapillary glomerulonephritis, often associated with extensive immunoglobulin deposition in the glomeruli with evidence of complement consumption (low serum C3, Box 17.40). In the developed world, hospital-acquired infections with various organisms are a common cause of these syndromes.

Worldwide, glomerulonephritis more commonly follows hepatitis B, hepatitis C, schistosomiasis, leishmaniasis, malaria and other chronic infections. Infection with HIV may be associated with FSGS (see above), particularly in people of African descent.

Post-streptococcal glomerulonephritis
This is a specific subtype of post-infectious glomerulonephritis. It is much more common in children than adults but is now rare in the developed world. The latency is usually about 10 days after a throat infection or longer after skin infection, suggesting an immune mechanism rather than direct infection. An acute nephritis of varying severity occurs. Sodium retention, hypertension and oedema are particularly pronounced. There is also reduction of GFR, proteinuria, haematuria and reduced urine volume. Characteristically, this gives the urine a red or smoky appearance. As in other causes of post-infectious glomerulonephritis, serum concentrations of C3 and C4 are typically reduced, reflecting complement consumption (see Box 17.40), and evidence of streptococcal infection may be found. Renal function begins to improve spontaneously within 10–14 days, and management by fluid and sodium restriction with diuretic and hypotensive agents is usually adequate. Remarkably, the renal lesion in almost all children and many adults seems to resolve completely, despite the severity of the glomerular inflammation and proliferation seen histologically.

Rapidly progressive glomerulonephritis
Rapidly progressive glomerulonephritis (also known as crescentic glomerulonephritis) is characterised by rapid loss of renal function over days to weeks. Renal biopsy shows crescentic lesions, often associated with necrotising lesions within the glomerulus, termed focal segmental (necrotising) glomerulonephritis. It is typically seen in Goodpasture’s disease, where the underlying cause is the development of antibodies to the glomerular basement membrane (anti-GBM antibodies), and in small-vessel vasculitides (pp. 519 and 1115). It can also be observed in SLE (pp. 520 and 1109) and occasionally IgA and other nephropathies. Rapid-onset disease may be associated with relatively little proteinuria (see Fig. 17.21). Management depends on the underlying cause but immunosuppressive drugs are often required. Patients with anti-GBM disease should be treated with plasma exchange combined with corticosteroids and immunosuppressants. Patients

**17.40 Causes of glomerulonephritis associated with low serum complement**
- Post-infection glomerulonephritis
- Subacute bacterial infection: especially endocarditis
- Systemic lupus erythematosus
- Cryoglobulinaemia
- Mesangiocapillary glomerulonephritis, usually complement type
with renal involvement secondary to ANCA-associated vasculitis and SLE should also be treated with corticosteroids and immunosuppressants, as described on pages 1118 and 1112.

**Inherited glomerular diseases**

**Alport’s syndrome**

A number of uncommon diseases may involve the glomerulus in childhood, but the most important one affecting adults is Alport’s syndrome. Most cases arise from a mutation or deletion of the COL4A5 gene on the X chromosome, which encodes type IV collagen, resulting in inheritance as an X-linked recessive disorder (p. 53). Mutations in COL4A3 or COL4A4 genes are less common and cause autosomal recessive disease. The accumulation of abnormal collagen results in a progressive degeneration of the GBM (Fig. 17.23). Affected patients progress from haematuria to ESRD in their late teens or twenties. Female carriers of COL4A5 mutations usually have haematuria but less commonly develop significant renal disease. Some other basement membranes containing the same collagen isoforms are similarly involved, notably in the cochlea, so that Alport’s syndrome is associated with sensorineural deafness and ocular abnormalities.

ACE inhibitors may slow but not prevent loss of kidney function. Patients with Alport’s syndrome are good candidates for RRT, as they are young and usually otherwise healthy. They can develop an immune response to the normal collagen antigens present in the GBM of the donor kidney and, in a small minority, anti-GBM disease develops and destroys the allograft.

**Thin glomerular basement membrane disease**

In thin glomerular basement membrane disease there is glomerular bleeding, usually only at the microscopic or dipstick level, without associated hypertension, proteinuria or a reduction in GFR. The glomeruli appear normal by light microscopy but, on electron microscopy, the GBM is abnormally thin. The condition may be familial and some patients are carriers of Alport mutations. This does not appear to account for all cases, and in many patients the cause is unclear. Monitoring of these patients is advisable, as proteinuria may develop in some and there seems to be an increased rate of progressive CKD in the long term.

**Acute interstitial nephritis**

Acute interstitial nephritis (AIN) is characterised by acute inflammation affecting the tubulo-interstitium of the kidney. It is commonly drug-induced but can be caused by other factors, such as renal toxins, and can complicate a variety of systemic diseases and infections (Box 17.41).

**Clinical features**

The clinical presentation is typically with renal impairment but, in some patients with drug-induced acute interstitial nephritis, there may be signs of a generalised drug hypersensitivity reaction with fever, rash and eosinophilia. Proteinuria is generally modest (PCR < 100 mg/mmol) and tubular in type (p. 476). The urine may contain red and white blood cells but is sterile on...
Chronic interstitial nephritis

Chronic interstitial nephritis (CIN) is characterised by renal dysfunction with fibrosis and infiltration of the renal parenchyma by lymphocytes, plasma cells and macrophages, in association with tubular damage.

**Pathophysiology**

The disease may follow on from acute interstitial nephritis that does not resolve, or may be associated with ingestion of various toxins and drugs, or with metabolic and chronic inflammatory diseases, as summarised in Box 17.42. In many patients, CIN presents at a late stage and no underlying cause can be identified. Toxins that have been associated with CIN include those contained within the plant *Aristolochia clematitis*. These are probably responsible for the severe nephrotoxicity that can be associated with treatment with herbal medicines in Asia. There is some evidence that these toxins are also responsible for Balkan nephropathy, which affects isolated rural communities in Bosnia, Bulgaria, Croatia, Romania and Serbia. The nephropathy is commonly linked with tumours of the collecting system and is probably due to the mutagenic effects of the plant toxin on the urothelial epithelium. Ingestion of mushrooms within the *Cortinarius* genus can cause a devastating and irreversible renal tubular toxicity. The typical scenario is when a...
patient mistakes a poisonous mushroom for an edible type. It is encountered occasionally in Scandinavia and Scotland.

**Clinical features**

Most patients with CIN present in adult life with CKD, hypertension and small kidneys. The CKD is often moderate (stage 3) but, because of tubular dysfunction, electrolyte abnormalities are typically more severe, resulting in hyperkalaemia and acidosis. Urinalysis abnormalities are non-specific. A minority of patients present with salt-losing nephropathy, characterised by hypotension, polyuria and features of sodium and water depletion. Patients with CIN have an impairment of urine-concentrating ability and sodium conservation, which puts them at risk of AKI due to salt and water depletion during an acute illness. Renal tubular acidosis (p. 446) may complicate CIN but is seen most often in myeloma, sarcoidosis, cystinosis, amyloidosis and Sjögren’s syndrome.

**Management**

Management is supportive in nature, with correction of acidosis and hyperkalaemia; replacement of fluid and electrolytes, as required; and RRT if irreversible renal damage has occurred.

---

**Reflux nephropathy**

This condition, which was previously known as chronic pyelonephritis, is a specific type of chronic interstitial nephritis (see previous section) associated with vesico-ureteric reflux (VUR) in early life and with the appearance of scars in the kidney, as demonstrated by various imaging techniques. About 12% of patients in Europe requiring treatment for ESRD may have this disorder but diagnostic criteria are imprecise.

**Pathophysiology**

Reflux nephropathy is thought to be due to chronic reflux of urine from the bladder into the ureters, in association with recurrent urinary tract infection (UTI) in childhood. It was previously assumed that ascending infection was necessary for progressive renal damage in patients with VUR, but there is evidence to suggest that renal scars can occur, even in the absence of infection. Furthermore, epidemiological surveys and controlled trials have found that efforts to correct VUR by using surgical or other means are ineffective in halting progression of the disease.

Susceptibility to VUR has a genetic component and may be associated with renal dysplasia and other congenital abnormalities of the urinary tract. It can be connected with outflow obstruction, usually caused by urethral valves, but usually occurs with an apparently normal bladder.

**Clinical features**

Usually, the renal scarring and dilatation are asymptomatic and the patient may present at any age with hypertension (sometimes severe), proteinuria or features of CKD. There may be no history of overt UTI. However, symptoms arising from the urinary tract may be present and include frequency of micturition, dysuria and aching lumbar pain. VUR may occur in children but diminishes as the child grows, and usually has disappeared by adulthood. Urinalysis often shows the presence of leukocytes and moderate proteinuria (usually < 1 g/24 hrs) but these are not invariable. The risk of renal stone formation is increased. A number of women first present with hypertension and/or proteinuria in pregnancy. Children and adults with small or unilateral renal scars have a good prognosis, provided renal growth is normal. With significant unilateral scars there is usually compensatory hypertrophy of the contralateral kidney. In patients with more severe bilateral disease, prognosis is related to the severity of renal dysfunction, hypertension and proteinuria. If the serum creatinine is normal and hypertension and proteinuria are absent, then the long-term prognosis is usually good.

**Investigations**

Renal scarring can be detected by ultrasound but it has poor sensitivity and is only capable of detecting major defects and excluding significant obstruction. Radionuclide DMSA scans are more sensitive (see Fig. 17.6, p. 471), and longitudinal imaging by MRI or CT may be useful in assessing progression. Abnormalities may be unilateral or bilateral and of any grade of severity. Gross scarring of the kidneys, commonly at the poles, is seen, with reduced kidney size and narrowing of the cortex.
Cystic diseases of the kidney

Papillary necrosis

The renal papillae lie within a hypertonic environment in the renal medulla, at the end of the vasa recta. They are susceptible to ischaemic damage because of this and can undergo necrosis when their vascular supply is impaired as the result of diabetes mellitus and sickle-cell disease or with long-term ingestion of NSAID. The condition may occasionally occur in other diseases. There is an association with pyelonephritis but it is difficult to determine whether this is a cause of papillary necrosis or a complication. The clinical presentation is variable. Some patients are asymptomatic and clinically silent, whereas others present with renal colic and renal impairment as necrosed papillae slough off and cause ureteric obstruction. Urinalysis may be normal but, more frequently haematuria and sterile pyuria are present. Significant proteinuria is unusual, unless there is renal failure. The imaging method of choice to make the diagnosis is pyelography. Management is based on relieving obstruction, where present, and withdrawal of the offending drugs.

Sickle-cell nephropathy

The longer survival of patients with sickle-cell disease (p. 1033) means that a high proportion now live to develop chronic complications of microvascular occlusion. In the kidney, these changes are most pronounced in the medulla, where the vasa recta are the site of sickling because of hypoxia and hypertonicity. Loss of urinary concentrating ability and polyuria are the earliest changes; distal renal tubular acidosis and impaired potassium excretion are typical. Papillary necrosis may also occur (see above). A minority of patients develop ESRD. This is managed according to the usual principles, but response to recombinant erythropoietin is poor due to the haemoglobinopathy. Patients with sickle trait have an increased incidence of unexplained microscopic haematuria.

Cystic diseases of the kidney

It is common to encounter patients with a single renal cyst or even multiple ones as an incidental finding, especially in those aged 50 years and over. Usually, these cysts are of no clinical consequence and are asymptomatic, but occasionally they can cause pain or haematuria. In addition, several specific diseases are recognised as being caused by the formation of multiple renal cysts. These are discussed in more detail below.

Adult polycystic kidney disease

Adult polycystic kidney disease (PKD) is a common condition, with a prevalence of approximately 1:1000, and is inherited as an autosomal dominant trait. Small cysts lined by tubular epithelium develop from infancy or childhood and enlarge slowly and irregularly. The surrounding normal kidney tissue is compressed and progressively damaged. Mutations in the PKD1 gene account for 85% of cases and PKD2 for about 15%. ESRD occurs in approximately 50% of patients with PKD1.

EBM 17.43 Prophylactic antibiotics and vesico-ureteric reflux

‘Prophylactic antibiotics reduce recurrences of UTI but there is no evidence that they protect against further renal scarring or dysfunction.’


For further information: www.clinicevidence.org

Fig. 17.25 Vesico-ureteric reflux (grade IV) shown by micturating cystogram. The bladder has been filled with contrast medium through a urinary catheter. After micturition, there was gross VUR into widely distended ureters and pelvicalyceal systems.
Kidney and urinary tract disease

1. Vague discomfort in loin or abdomen due to increasing mass of renal tissue
2. Acute loin pain or renal colic due to haemorrhage into a cyst
3. Hypertension
4. Haematuria (with little or no proteinuria)
5. Urinary tract or cyst infections
6. Renal failure

Adult polycystic kidney disease: common clinical features

Mutations, with a mean age of onset of 52 years, but in a minority of patients with PKD2 mutations, with a mean age of onset of 69 years. It has been estimated that between 5% and 10% of patients on RRT have adult PKD.

Clinical features

Common clinical features are shown in Box 17.44. Affected subjects are usually asymptomatic until later life but hypertension usually occurs from the age of 20 onwards. One or both kidneys may be palpable and the surface may feel nodular. About 30% of patients with PKD also have hepatic cysts (see Fig. 23.38, p. 970) but disturbance of liver function is rare. Sometimes (almost always in women) this causes massive and symptomatic hepatomegaly, usually concurrent with renal enlargement but occasionally with only minor renal involvement. Berry aneurysms of cerebral vessels are an associated feature and about 10% of patients have a subarachnoid haemorrhage. This feature appears to be largely restricted to certain families (and presumably specific mutations). Mitral and aortic regurgitation is frequent but rarely severe, and colonic diverticula and abdominal wall hernias may occur.

Investigations

The diagnosis is usually based on family history, clinical findings and ultrasound examination. Ultrasound examination demonstrates cysts in approximately 95% of affected patients over the age of 20 and is the screening method of choice but may not detect small developing cysts in younger subjects. Cysts may also be identified by other imaging modalities, such as MRI (Fig. 17.26). It is now possible to make a molecular diagnosis by mutation screening of PDK1 or PDK2 but this is seldom used in routine clinical practice. Screening for intracranial aneurysms is not generally indicated but can be done by MR angiography in families with a history of subarachnoid haemorrhage. The yield of screening is low, however, and the risk–benefit ratio of intervention in asymptomatic aneurysms in this disease is not clear.

Management

Good control of blood pressure is important because cardiovascular morbidity and mortality are so common in renal disease, but there is no evidence that control of moderate hypertension retards the development of renal failure in PKD, in contrast to the evidence for glomerular diseases. There is some evidence that the vasopressin V2 receptor antagonist, tolvaptan, can slow cyst formation in some patients but its place in treatment has yet to be established.

Patients with PKD are usually good candidates for dialysis and transplantation. Sometimes kidneys are so large that one or both have to be removed to make space for a renal transplant. Otherwise, unless they are a source of pain or infection, they are usually left in situ.

Medullary sponge kidney disease

Medullary sponge kidney is characterised by cysts confined to papillary collecting ducts. The disease is not inherited and its cause is unknown. Patients usually present as adults with renal stones. These are often recurrent, and preventive measures (p. 509) need to be implemented if so, but the prognosis is generally good. The diagnosis is made by ultrasound or IVU (Fig. 17.27). Contrast medium is seen to fill dilated or cystic tubules, which are sometimes calcified.

Medullary cystic kidney disease

This is a heterogeneous group of inherited disorders, known as nephronophthisis in children. Small cortical cysts are associated with progressive destruction of the nephron. The childhood variants are characterised by thirst and polyuria due to nephrogenic diabetes insipidus, often with a family history of similar disease. Sometimes, affected patients are ‘salt-losing’, which
Renal stone disease

The clinical presentation is highly variable. Most patients with renal stone disease are asymptomatic, whereas others present with pain, haematuria, UTI or urinary tract obstruction. A common presentation is with acute pain that may worsen over time, leading to nausea, vomiting and sometimes fever. This pain is typically located in the flank, lower back or abdomen, and may radiate to the groin or testis in men. Hematuria is common, and the presence of blood in the urine can be visible to the naked eye or detected by a urine dipstick. Other symptoms may include frequency, urgency, dysuria and suprapubic pain. Older adults with renal stone disease may present with these symptoms, or they may present with acute urinary tract infection or sepsis caused by the infected stone.

Common complications of renal stone disease include:

- **Renal colic**: Severe pain that can be intermittent or constant, accompanied by nausea, vomiting, and sometimes fever. It is due to the passage of the stone through the ureter and can be quite severe.
- **Hematuria**: Presence of blood in the urine, which can vary from a minor discoloration to a deep red or maroon color.
- **Acute pyelonephritis**: Inflammation of the kidney due to a urinary tract infection, which can cause fever, chills, and pain in the flank or lower back.
- **Obstructive uropathy**: Blockage of the urinary tract, which can lead to renal failure and permanent damage to the kidney.

**Screening for renal stone disease**

Screening for renal stone disease is usually not routine, as it is asymptomatic in many patients. However, individuals with a history of kidney stones, family history of kidney stones, or those with certain medical conditions (e.g., hypercalciuria, sarcoidosis) may be screened. Screening typically involves a medical history, physical examination, and laboratory tests such as urine analysis, blood tests, and imaging studies like ultrasound or CT scan. Identification of risk factors and early intervention can help prevent further complications and reduce the risk of recurrence.
loin pain radiating to the anterior abdominal wall, together with haematuria: a symptom complex termed renal or ureteric colic. This is most commonly caused by ureteric obstruction by a calculus but the same symptoms can occur in association with a sloughed renal papilla, tumour or blood clot. The patient is suddenly aware of pain in the loin, which radiates round the flank to the groin and often into the testis or labium, in the sensory distribution of the first lumbar nerve. The pain steadily increases in intensity to reach a peak in a few minutes. The patient is restless and generally tries unsuccessfully to obtain relief by changing position or pacing the room. There is pallor, sweating and often vomiting. Frequency, dysuria and haematuria may occur. The intense pain usually subsides within 2 hours but may continue unabated for hours or days. It is usually constant during attacks, although slight fluctuations in severity may be seen. Subsequent to an attack of renal colic, intermittent dull pain in the loin or back may persist for several hours.

**Investigations**

Patients with symptoms of renal colic should be investigated to determine whether or not a stone is present, to identify its location and to assess whether it is causing obstruction. About 90% of stones contain calcium and these can be visualised on plain abdominal X-ray but CTKUB is the gold standard for diagnosing a stone within the kidney or ureter, as 99% are visible using this method. Alternatively, an IVU can be performed. The advantage of CTKUB over IVU is that it is more sensitive and can identify non-radio-opaque stones, such as those containing uric acid and cystine. When the stone is in the ureter, an IVU shows delayed excretion of contrast from the kidney, and a ureter that is dilated down as far as the stone (Fig. 17.29). Ultrasound can show stones within the kidney and dilatation of the renal pelvis and ureter if the stone is obstructing urine flow; it is useful in unstable patients or young women, in whom exposure to ionising radiation is undesirable.

A minimum set of investigations (Box 17.47) should be performed in patients with a first renal stone. The yield of more detailed investigation is low, and hence usually reserved for young patients, those with recurrent or multiple stones, or those with complicated or unexpected presentations. Chemical analysis of stones is often helpful in defining the underlying cause. Since most stones pass spontaneously through the urinary tract, urine should be sieved for a few days after an episode of colic in order to collect the calculus for analysis.

**Fig. 17.28** Radio-opaque bilateral staghorn calculi visible during intravenous urography. The intravenous pyelogram demonstrates that, while some dye is being excreted by the right kidney, there is little function on the left.

**Fig. 17.29** Unilateral ureteric obstruction. Intravenous urogram of a patient with a stone (not visible) at the lower end of the right ureter. This film, taken 2 hours post-contrast injection, demonstrates persistence of contrast medium in the right kidney, pelvicalyceal system and ureter, whereas only a small amount remains visible in the normal left pelvicalyceal system.

**17.47 Investigations for renal stones**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Test</th>
<th>First stone</th>
<th>Recurrent stone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone</td>
<td>Chemical composition¹</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Blood</td>
<td>Calcium</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Phosphate</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uric acid</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urea and electrolytes</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Bicarbonate</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Parathyroid hormone ²</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Urine</td>
<td>Dipstick test for protein, blood, glucose</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>Amino acids</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>24-hr urine</td>
<td>Urea</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Creatinine clearance</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sodium</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calcium</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxalate</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uric acid</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

¹The most valuable test if a stone can be obtained.
²Only if serum calcium or urinary calcium excretion high.
Renal stone disease

Management

The immediate treatment of renal colic is with analgesia and antiemetics. Renal colic is often unbearably painful and demands powerful analgesia; diclofenac orally or as a suppository (100 mg) is often very effective, followed by morphine (10–20 mg) or pethidine (100 mg) intramuscularly. Around 90% of stones of less than 4 mm diameter pass spontaneously, but only 10% of stones bigger than 6 mm, and these may require endoscopic surgical intervention (see below). Patients with renal stones are at high risk of infection; if surgery is contemplated, it should be covered with appropriate antibiotics. Immediate action is required if infection occurs in the stagnant urine proximal to the stone (pyonephrosis), and in patients with a solitary kidney who develop anuria in association with a stone in the ureter.

Stones that do not pass spontaneously through the urinary tract may need to be removed surgically or fragmented by extracorporeal shock wave lithotripsy (ESWL), in which shock waves generated outside the body are focused on the stone, breaking it into small pieces that can pass easily down the ureter. The indications for intervention to remove stones from the renal tract are summarised in Box 17.48. Procedures vary, depending on the site (Fig. 17.30).

Measures to prevent further stone formation are guided by the investigations in Box 17.47. Some general principles apply to almost every patient with calcium-containing stones (Box 17.49). More specific measures

---

**Box 17.48 Indications for intervention to remove stones from the urinary tract**

<table>
<thead>
<tr>
<th>Clinical presentation</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstruction and/or anuria</td>
<td>Emergency PCNL or stent</td>
</tr>
<tr>
<td>Pyonephrosis associated with stone</td>
<td>Emergency PCNL or stent</td>
</tr>
<tr>
<td>Stone in a patient with solitary kidney</td>
<td>Urgent PCNL, stent, ESWL or ureteroscopy*</td>
</tr>
<tr>
<td>Severe pain and persistence of stone in renal tract</td>
<td>Urgent PCNL, stent, ESWL or ureteroscopy*</td>
</tr>
<tr>
<td>Pain and persistence of stone in renal tract</td>
<td>Elective PCNL, ESWL or ureteroscopy*</td>
</tr>
</tbody>
</table>

*Procedure depends on site of stone – see Fig. 17.30. (ESWL = extracorporeal shock wave lithotripsy; PCNL = percutaneous nephrolithotomy)

**Box 17.49 Measures to prevent calcium stone formation**

**Diet**

- **Fluid**
  - At least 2 L output per day (intake 3–4 L): check with 24-hr urine collections
  - Intake distributed throughout the day (especially before bed)

- **Sodium**
  - Restrict intake

- **Protein**
  - Moderate, not high

- **Calcium**
  - Maintain good calcium intake (calcium forms an insoluble salt with dietary oxalate, lowering oxalate absorption and excretion)
  - Avoid calcium supplements separate from meals (increase calcium excretion without reducing oxalate excretion)

- **Oxalate**
  - Avoid foods that are rich in oxalate (spinach, rhubarb)

**Drugs**

- **Thiazide diuretics**
  - Reduce calcium excretion
  - Valuable in recurrent stone-formers and hypercalciuria

- **Allopurinol**
  - If urate excretion high (unproven except for urate stones)

- **Avoid**
  - Vitamin D supplements (increase calcium absorption and excretion)
  - Vitamin C supplementation (increases oxalate excretion)

---

**Fig. 17.30 Options for removal of urinary stones.**

A patient undergoing extracorporeal shock wave lithotripsy (ESWL). The procedures that are used for removal of stones in the urinary tract, shown in relation to the site of the stone. Very rarely, open or laparoscopic surgery may be necessary for removal of stones in the upper renal tract, if other methods fail. (PCNL = percutaneous nephrolithotomy)
apply to some types. Urate stones can be prevented by allopurinol, but its role in patients with calcium stones and high urate excretion is uncertain. Stones formed in cystinuria can be reduced by penicillamine therapy. It may also be helpful to attempt to alkalise the urine with sodium bicarbonate, as a high pH discourages urate and cystine stone formation.

**ISOLATED DEFECTS OF TUBULAR FUNCTION**

An increasing number of disorders have been identified that are caused by specific defects in transporter molecules expressed in renal tubular cells. Only the most common are mentioned here. Renal glycosuria is a benign autosomal recessive defect of tubular reabsorption of glucose, caused by mutations of the sodium/glucose co-transporter SGLT2. Glucose appears in the urine in the presence of a normal blood glucose concentration. Cystinuria is a rare condition, in which reabsorption of filtered cystine, ornithine, arginine and lysine is defective. It is caused by mutations in the SLC3A1 amino acid transporter gene. The high concentration of cystine in urine leads to cystine stone formation (p. 507).

Other uncommon tubular disorders include hereditary hypophosphataemic rickets (pp. 127 and 1128), in which reabsorption of filtered phosphate is reduced; nephrogenic diabetes insipidus (p. 794), in which the tubules are resistant to the effects of vasopressin; and Bartter’s and Gitelman’s syndromes, in which there is sodium-wasting and hypokalaemia (p. 440).

The term ‘Fanconi syndrome’ is used to describe generalised proximal tubular dysfunction. The condition typically presents with low blood phosphate and uric acid, glycosuria, amino aciduria and proximal renal tubular acidosis. In addition to the causes of interstitial nephritis described above, some congenital metabolic disorders are associated with Fanconi syndrome, notably Wilson’s disease, cystinosis and hereditary fructose intolerance.

Renal tubular acidosis describes the common endpoint of a variety of diseases affecting distal (classical or type 1) or proximal (type 2) renal tubular function. These syndromes are described on page 446.

**DISEASES OF THE COLLECTING SYSTEM AND URETERS**

### Congenital abnormalities

Various congenital anomalies of the urinary tract can occur (Fig. 17.31); they affect more than 10% of infants. If not immediately lethal, they can lead to complications in later life, including obstructive nephropathy and CKD.

### Single kidneys

About 1 in 500 infants is born with only one kidney. Although this is usually compatible with normal life, it may be associated with other abnormalities.

**Ureterocele**

A ureterocele occurs behind a pin-hole ureteric orifice when the intramural part of the ureter dilates and bulges into the bladder. It can become very large and cause lower urinary tract obstruction. Incision of the pin-hole opening relieves the obstruction.

**Ectopic ureters and duplex kidneys**

Ectopic ureters occur with congenital duplication of one or both kidneys (duplex kidneys). Developmentally, the ureter has two main branches and, if this arrangement persists, the two ureters of the duplex kidneys may drain separately into the bladder. The lower pole moiety enters the bladder superiorly and laterally, while the upper pole moiety enters the bladder inferomedially to the lower pole moiety ureter or, more rarely, the vagina or seminal vesicle. The lower pole moiety has an ineffective valve mechanism, so that urine passes up the ureter on voiding (vesico-ureteric reflux, p. 504), whereas the upper pole moiety is often associated with a ureterocele.

**Obstructive megaureter**

In primary obstructive megaureter, there is dilatation of the ureter in all but its terminal segment without obvious cause and without vesico-ureteric reflux. Radiographic and pressure/flow studies may be needed to determine whether there is obstruction to urine flow. Narrowing of the ureter and re-implantation may be necessary.

**Pelvi-ureteric junction obstruction**

This causes idiopathic hydronephrosis and results from a functional obstruction at the junction of the ureter and renal pelvis. The abnormality is likely to be congenital and is often bilateral. It can be seen in very young
children but gross hydronephrosis may present at any age. The common presentation is ill-defined renal pain or ache, exacerbated by drinking large volumes of liquid (Dietl’s crisis). Rarely, it is asymptomatic. The diagnosis is often suspected after ultrasound or IVU, and can be confirmed with a 99mTc-MAG3 renogram followed by diuretic. Treatment is surgical excision of the PUJ and re-anastomosis (pyeloplasty), which can now be performed laparoscopically. Less invasive alternatives are also possible, including balloon dilatation and endoscopic pyelotomy, but are generally less effective.

**Retroperitoneal fibrosis**

Fibrosis of the retroperitoneal connective tissues may encircle and compress the ureter(s), causing obstruction. The fibrosis is most commonly idiopathic, but can represent a reaction to infection, radiation or aortic aneurysm, or be caused by cancer or a drug reaction – methysergide, for example. Patients usually present with ill-defined symptoms of ureteric obstruction. Typically, there is an acute-phase response (high CRP and ESR). Imaging with UVI or CT shows ureteric obstruction with medial deviation of the ureters. Idiopathic retroperitoneal fibrosis responds well to corticosteroids and may respond more slowly to tamoxifen, but ureteric stenting is often necessary to relieve obstruction. Failure to improve indicates the need for surgery (ureterolysis) both to relieve obstruction and to exclude malignancy.

**INFECTIONS OF THE URINARY TRACT**

In health, bacterial colonisation is confined to the lower end of the urethra and the remainder of the urinary tract is sterile (see Ch. 6). The urinary tract can become infected with various bacteria but the most common is *E. coli* derived from the gastrointestinal tract. The most common presenting problem is cystitis with urethritis (generally referred to as urinary tract infection) but this is only part of a spectrum of severity (Box 17.50).

**Urinary tract infection**

Urinary tract infection (UTI) is the term used to describe acute urethritis and cystitis caused by a microorganism. It is a common disorder accounting for 1–3% of consultations in general medical practice. The prevalence of UTI in women is about 3% at the age of 20, increasing by about 1% in each subsequent decade. In males, UTI is uncommon, except in the first year of life and in men over 60, when it may complicate bladder outflow obstruction.

**Pathophysiology**

Urine is an excellent culture medium for bacteria; in addition, the urothelium of susceptible persons may have more receptors, to which virulent strains of *E. coli* become adherent. In women, the ascent of organisms into the bladder is easier than in men; the urethra is shorter and the absence of bactericidal prostatic secre­tions may be relevant. Sexual intercourse may cause minor urethral trauma and transfer bacteria from the perineum into the bladder. Instrumentation of the bladder may also introduce organisms. Multiplication of organisms then depends on a number of factors, including the size of the inoculum and virulence of the bacteria. Conditions that predispose to UTI are shown in Box 17.51.

**Clinical features**

Typical features of cystitis and urethritis include:
- abrupt onset of frequency of micturition and urgency
- scalding pain in the urethra during micturition (dysuria)
- suprapubic pain during and after voiding
- intense desire to pass more urine after micturition, due to spasm of the inflamed bladder wall (strangury)
- urine that may appear cloudy and have an unpleasant odour
- microscopic or visible haematuria.

Systemic symptoms are usually slight or absent. However, infection in the lower urinary tract can spread (see Box 17.50); acute pyelonephritis is suggested by prominent systemic symptoms with fever, rigors, vomiting, hypotension and loin pain, guarding or tenderness, and may be an indication for hospitalisation. Only about 30% of patients with acute pyelonephritis have associated symptoms of cystitis or urethritis. Prostatitis is suggested by perineal or suprapubic pain, pain on ejaculation and prostatic tenderness on rectal examination.

The differential diagnosis of lower urinary tract symptoms includes urethritis due to sexually transmitted disease, notably chlamydia (p. 422) or Reiter’s syndrome (p. 1107). Some patients, usually female, have symptoms suggestive of urethritis and cystitis but no bacteria are cultured from the urine (the ‘urethral syndrome’). Possible explanations include infection with
organisms not readily cultured by ordinary methods (such as Chlamydia and certain anaerobes), intermittent or low-count bacteriuria, reaction to toiletries or disinfectants, symptoms related to sexual intercourse, or post-menopausal atrophic vaginitis.

The differential diagnosis of acute pyelonephritis includes pyelonephrosis, acute appendicitis, diverticulitis, cholecystitis, salpingitis, ruptured ovarian cyst or ectopic pregnancy. In pyelonephrosis due to upper urinary tract obstruction, patients may become extremely ill, with fever, leucocytosis and positive blood cultures. With a perinephric abscess, there is marked pain and tenderness, and often bulging of the loin on the affected side. Urinary symptoms may be absent in this situation and urine testing negative, containing neither pus cells nor organisms.

**Investigations**

An approach to investigation is shown in Box 17.52. In an otherwise healthy woman with a single lower urinary tract infection, urine culture prior to treatment is not mandatory. Investigation is necessary, however, in patients with recurrent infection or after failure of initial treatment, during pregnancy, or in patients susceptible to serious infection, such as the immunocompromised, those with diabetes or an indwelling catheter, and older people (Box 17.53). The diagnosis can be made from the combination of typical clinical features and abnormalities on urinalysis. Most urinary pathogens can reduce nitrate to nitrite, and neutrophils and nitrates can usually be detected in symptomatic infections by urine dipstick tests for leucocyte esterase and nitrite, respectively. The absence of both nitrates and leucocyte esterase in the urine makes UTI unlikely. Interpretation of bacterial counts in the urine, and of what is a ‘significant’ culture result, is based on probabilities. Urine taken by suprapubic aspiration should be sterile, so the presence of any organisms is significant. If the patient has symptoms and there are neutrophils in the urine, a small number of organisms is significant. In asymptomatic patients, more than 10^9 organisms/mL is usually regarded as significant (asymptomatic bacteriuria, see below).

![17.52 Investigation of patients with urinary tract infection](image)

**17.52 Investigation of patients with urinary tract infection**

**All patients**
- Dipstick* estimation of nitrite, leucocyte esterase and glucose
- Microscopy/cytometry of urine for white blood cells, organisms
- Urine culture

**Infants, children, and anyone with fever or complicated infection**
- Full blood count; urea, electrolytes, creatinine
- Blood cultures

**Pyelonephritis; males; children; women with recurrent infections**
- Renal tract ultrasound or CT
- Pelvic examination in women, rectal examination in men
- Continuing haematuria or other suspicion of bladder lesion
- Cystoscopy

*May substitute for microscopy and culture in simple uncomplicated infection.

Typical organisms causing UTI in the community include E. coli derived from the gastrointestinal tract (about 75% of infections), Proteus spp., Pseudomonas spp., streptococci and Staphylococcus epidermidis. In hospital, E. coli still predominates, but Klebsiella or streptococci are more common. Certain strains of E. coli have a particular propensity to invade the urinary tract.

Investigations to detect underlying predisposing factors for UTI are used selectively, most commonly in children, men or patients with recurrent infections (see Box 17.52).

**Management**

Antibiotics are recommended in all cases of proven UTI (Box 17.54). If urine culture has been performed, treatment may be started while awaiting the result. For infection of the lower urinary tract, treatment for 3 days is the norm and is less likely to induce significant alterations in bowel flora than more prolonged therapy. Trimethoprim is the usual choice for initial treatment; however, between 10% and 40% of organisms causing UTI are resistant to trimethoprim, the lower rates being seen in community-based practice. Nitrofurantoin, quinolone antibiotics such as ciprofloxacin and norfloxacin, and cefalexin are also generally effective. Co-amoxiclav or amoxicillin should only be used when the organism is known to be sensitive. Penicillins and cephalosporins are safe to use in pregnancy but trimethoprim, sulphonamides, quinolones and tetracyclines should be avoided.

In more severe infection, antibiotics should be continued for 7–14 days. Seriously ill patients may require intravenous therapy with a cephalosporin, quinolone or gentamicin for a few days (see Box 17.54), later switching to an oral agent.

A fluid intake of at least 2 L/day is usually recommended, although this is not based on evidence and may make symptoms of dysuria worse.

**Persistent or recurrent UTI**

If the causative organism persists on repeat culture despite treatment, or if there is re-infection with any organism after an interval, then an underlying cause is more likely to be present (see Box 17.51) and more detailed investigation is justified (see Box 17.52). In women, recurrent infections are common and
Infections of the urinary tract

Infections of the urinary tract are common, particularly in women aged over 65. There is no evidence that this condition causes renal scarring in adults who are not pregnant and have a normal urinary tract and, in general, treatment is not indicated. However, up to 30% will develop symptomatic infection within 1 year. Treatment is required in infants, pregnant women and those with urinary tract abnormalities.

Catheter-related bacteriuria

In patients with a urethral catheter, bacteriuria increases the risk of Gram-negative bacteraemia fivefold. Bacteriuria is common, however, and almost universal during long-term catheterisation. Treatment is usually avoided in asymptomatic patients, as this may promote antibiotic resistance. Careful sterile insertion technique is important, and the catheter should be removed as soon as it is not required.

Acute pyelonephritis

The kidneys are infected in a minority of patients with UTI. Acute renal infection (pyelonephritis) presents as a classic triad of loin pain, fever and tenderness over the kidneys. The renal pelvis is inflamed and small abscesses are often evident in the renal parenchyma (see Fig. 17.24C, p. 503)

Asymptomatic bacteriuria

This is defined as more than 10⁵ organisms/mL in the urine of apparently healthy asymptomatic patients. Approximately 1% of children under the age of one, 1% of schoolgirls, 0.03% of schoolboys and men, 3% of non-pregnant adult women and 5% of pregnant women have asymptomatic bacteriuria. It is increasingly common in those aged over 65. There is no evidence that this condition causes renal scarring in adults who are not pregnant and have a normal urinary tract and, in general, treatment is not indicated. However, up to 30% will develop symptomatic infection within 1 year. Treatment is required in infants, pregnant women and those with urinary tract abnormalities.

Catheter-related bacteriuria

In patients with a urethral catheter, bacteriuria increases the risk of Gram-negative bacteraemia fivefold. Bacteriuria is common, however, and almost universal during long-term catheterisation. Treatment is usually avoided in asymptomatic patients, as this may promote antibiotic resistance. Careful sterile insertion technique is important, and the catheter should be removed as soon as it is not required.

Acute pyelonephritis

The kidneys are infected in a minority of patients with UTI. Acute renal infection (pyelonephritis) presents as a classic triad of loin pain, fever and tenderness over the kidneys. The renal pelvis is inflamed and small abscesses are often evident in the renal parenchyma (see Fig. 17.24C, p. 503).

Racial infection is almost always caused by organisms ascending from the bladder, and the bacterial profile is the same as for lower urinary tract infection (p. 512). Rarely, bacteraemia may give rise to renal or perinephric abscesses, most commonly due to staphylococci. Predisposing factors, such as cysts or renal scarring, facilitate infection.

### 17.54 Antibiotic regimens for urinary tract infection in adults

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Drug</th>
<th>Regimen</th>
<th>Duration</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cystitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice</td>
<td>Trimethoprim</td>
<td>200 mg twice daily</td>
<td>3 days</td>
<td>7–10 days in men</td>
</tr>
<tr>
<td>Second choices</td>
<td>Amoxicillin</td>
<td>250 mg 3 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nitrofurantoin</td>
<td>50 mg 4 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cefalexin</td>
<td>250 mg 4 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ciprofloxacin</td>
<td>100 mg twice daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In pregnancy</td>
<td>Co-amoxiclav</td>
<td>250/125 mg 3 times daily</td>
<td>7 days</td>
<td>Avoid trimethoprim and quinolones during pregnancy</td>
</tr>
<tr>
<td></td>
<td>Cefalexin</td>
<td>250 mg 4 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amoxicillin</td>
<td>250 mg 3 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prophylactic therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice</td>
<td>Trimethoprim</td>
<td>100 mg at night</td>
<td>Continuous</td>
<td></td>
</tr>
<tr>
<td>Second choices</td>
<td>Nitrofurantoin</td>
<td>50 mg at night</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Co-amoxiclav</td>
<td>250/125 mg at night</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pyelonephritis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice</td>
<td>Co-amoxiclav</td>
<td>500/125 mg 3 times daily</td>
<td>14 days</td>
<td>Admit to hospital if no response within 24 hrs</td>
</tr>
<tr>
<td></td>
<td>Ciprofloxacin</td>
<td>500 mg twice daily</td>
<td>7 days</td>
<td>Switch to appropriate oral agent as soon as possible</td>
</tr>
<tr>
<td>Second choice</td>
<td>Gentamicin</td>
<td>Adjust dose according to renal function and serum levels</td>
<td>14 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cefuroxime</td>
<td>150–1500 mg 3 times daily</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Epididymo-orchitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice</td>
<td>Ciprofloxacin</td>
<td>500 mg twice daily</td>
<td>14 days</td>
<td>Refer to genito-urinary department to exclude N. gonorrhoeae</td>
</tr>
<tr>
<td><strong>Acute prostatitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice</td>
<td>Trimethoprim</td>
<td>200 mg twice daily</td>
<td>28 days</td>
<td></td>
</tr>
<tr>
<td>Second choice</td>
<td>Ciprofloxacin</td>
<td>500 mg twice daily</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1In all cases, the choice of drug should take locally determined antibiotic resistance patterns into account. 2See Hartford nomogram (p. 157).

### 17.55 Prophylactic measures to be adopted by women with recurrent urinary infections

- Fluid intake of at least 2 L/day
- Regular complete emptying of bladder
- Good personal hygiene
- Emptying of bladder before and after sexual intercourse
- Cranberry juice may be effective
KIDNEY AND URINARY TRACT DISEASE

Rarely, acute pyelonephritis is associated with papillary necrosis. Fragments of renal papillary tissue are passed per urethra and can be identified histologically. They may cause ureteric obstruction and, if this occurs bilaterally or in a single kidney, it may lead to AKI. Predisposing factors include diabetes mellitus, chronic urinary obstruction, analgesic nephropathy and sickle-cell disease. A necrotising form of pyelonephritis with gas formation, ‘emphysematous pyelonephritis’, is occasionally seen in patients with diabetes mellitus. Xanthogranulomatous pyelonephritis is a chronic infection that can resemble renal cell cancer. It is usually associated with obstruction, is characterised by accumulation of foamy macrophages and generally requires nephrectomy. Infection of cysts in polycystic kidney disease (p. 505) calls for prolonged antibiotic treatment.

Appropriate investigations are shown in Box 17.52 and management is described above and in Box 17.54. Intravenous rehydration may be needed in severe cases. If complicated infection is suspected or response to treatment is not prompt, urine should be recultured and renal tract ultrasound performed to exclude urinary tract obstruction or a perinephric collection. If obstruction is present, drainage by a percutaneous nephrostomy or ureteric stent should be considered.

Tuberculosis

Tuberculosis of the kidney and renal tract is secondary to tuberculosis elsewhere (p. 688) and is the result of blood-borne infection. Initially, lesions develop in the renal cortex; these may ulcerate into the renal pelvis and involve the ureters, bladder, epididymis, seminal vesicles and prostate. Calcification in the kidney and structure formation in the ureter are typical.

Clinical features may include symptoms of bladder involvement (frequency, dysuria); haematuria (sometimes macroscopic); malaise, fever, night sweats, lassitude and weight loss; loin pain; associated genital disease; and chronic renal failure as a result of urinary tract obstruction or destruction of kidney tissue. Neutrophils are present in the urine but routine urine culture may be negative (‘sterile pyuria’). Special techniques of microscopy and culture may be required to identify tubercle bacilli and are most usefully performed on early morning urine specimens. Bladder involvement should be assessed by cystoscopy. Radiology of the urinary tract and a chest X-ray to look for pulmonary tuberculosis are mandatory. Anti-tuberculous chemotherapy follows standard regimes (p. 693). Surgery to relieve urinary tract obstruction or to remove a very severely infected kidney may be required.

Benign Prostatic Enlargement

Benign prostatic enlargement (BPE) is extremely common. It has been estimated that about half of all men aged 80 years and over will have lower urinary tract symptoms associated with BPE. Benign prostatic hyperplasia (BPH) is the histological abnormality that underlies BPE.

Pathophysiology

The prostate gland increases in volume by 2.4 cm³ per year on average from 40 years of age. The process begins in the peri-urethral (transitional) zone and involves both glandular and stromal tissue to a variable degree. The cause is unknown, although BPE does not occur in patients with hypogonadism, suggesting that hormonal factors may be important.

Clinical features

The primary symptoms of BPE arise because of difficulty in voiding urine due to obstruction of the urethra by the prostate; they consist of hesitancy, poor urinary flow and a sensation of incomplete emptying. Other symptoms include urinary frequency, urgency of micturition and urge incontinence, although these are not specific to BPE. Some patients present suddenly with acute urinary retention, when they are unable to micturate and develop a painful distended bladder. This is often precipitated by excessive alcohol intake, constipation and prostatic infection. Severity of symptoms can be ascertained by using the International Prostate Symptom Score questionnaire (IPSS, Box 17.56), which serves as a valuable starting point for assessment of the patient. Once a baseline value is established, any improvement or deterioration may be monitored on subsequent visits. The IPSS may be combined with a quality of life score, in which patients are asked the following question ‘If you were to spend the rest of your life with your

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Question</th>
<th>Example score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straining</td>
<td>How often have you had to push or strain to begin urination?</td>
<td>1</td>
</tr>
<tr>
<td>Urgency</td>
<td>How often have you found it difficult to postpone urination?</td>
<td>2</td>
</tr>
<tr>
<td>Hesitancy</td>
<td>How often have you found that you stopped and started again several times when you urinated?</td>
<td>1</td>
</tr>
<tr>
<td>Incomplete emptying</td>
<td>How often have you had a sensation of not emptying your bladder completely after you finished urinating?</td>
<td>3</td>
</tr>
<tr>
<td>Frequency</td>
<td>How often have you had to urinate again less than 2 hours after you finished urinating?</td>
<td>1</td>
</tr>
<tr>
<td>Weak stream</td>
<td>How often have you had a weak urinary stream?</td>
<td>2</td>
</tr>
<tr>
<td>Nocturia</td>
<td>How many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?</td>
<td>1</td>
</tr>
</tbody>
</table>

Total score: 11

0 = not at all; 1 = less than one-fifth of the time; 2 = less than half the time; 3 = about half of the time; 4 = more than half of the time; 5 = almost always.

A score of 0–1 indicates mild symptoms, 8–19 moderate symptoms and 20–35 severe symptoms. In the example shown, the patient had moderate symptoms.
urinary condition the way it is now, how would you feel about that? Responses range from 0 (delighted) to 6 (terrible).

Patients may also present with chronic urinary retention. Here, the bladder slowly distends due to inadequate emptying over a long period of time. Patients with chronic retention can also develop acute retention: so-called acute-on-chronic retention. This condition is characterised by pain-free bladder distension, which may result in hydroureter, hydrenephrosis and renal failure (high-pressure chronic retention). On digital rectal examination, patients with BPE have evidence of prostatic enlargement with a smooth prostate gland. Abdominal examination may also reveal evidence of bladder enlargement in patients with urinary retention.

**Investigations**

The diagnosis of BPE is a clinical one but flow rates can be accurately measured with a flow meter, and prostate volume by transrectal ultrasound scan (TRUS). Objective assessment of obstruction is possible by urodynamics but this is seldom required. If symptoms or signs, such as a palpable bladder, nocturnal enuresis, recurrent urinary tract infections or a history of renal stones, are present, renal function should be assessed; if it is abnormal, screening should be conducted for evidence of obstructive nephropathy by ultrasound examination.

**Management**

Patients who present with acute retention require urgent treatment and should undergo immediate catheterisation to relieve the obstruction. Those with mild to moderate symptoms can be treated by medication (Boxes 17.57 and 17.58). The first-line treatments are $\alpha_1$-adrenoceptor blockers such as tamsulosin, which reduce the tone of smooth muscle cells in the prostate and bladder neck, thereby reducing the obstruction. The 5α-reductase inhibitors, finasteride and dutasteride, inhibit conversion of testosterone to the more potent dihydrotestosterone in the prostate and so cause the prostate to shrink. This class of drugs is indicated in patients with an estimated prostate size of more than 30 g or a prostate specific antigen (PSA) of more than 1.4 μg/L. Patients who fail to respond to a single drug may be treated with a combination of $\alpha$-blockers and 5α-reductase inhibitors, since this is more efficacious than either agent alone. Symptoms that are resistant to medical therapy require surgical treatment to remove some of the prostate tissue that is causing urethral obstruction. This is usually achieved by transurethral resection of the prostate (TURP) but enucleation of the prostate by holmium laser is equally effective and has potentially fewer complications. Open surgery is rarely needed, unless the gland is very large.

### Prostatitis

This condition is due to inflammation of the prostate gland. Acute or chronic bacterial prostatitis can be caused by infection with the same bacteria that are associated with UTI (p. 511) but prostatitis can also be ‘non-bacterial’, in which case no organism can be cultured from the urine. This is also known as chronic pelvic pain syndrome. Clinical features of prostatitis include frequency, dysuria, painful ejaculation, perineal or groin pain, difficulty passing urine and, in acute disease, considerable systemic disturbance. The prostate is enlarged and tender. Bacterial prostatitis is confirmed by a positive culture from urine or from urethral discharge obtained after prostatic massage, and the treatment of choice is trimethoprim or a quinolone antibiotic. A 4–6-week course is required (see Box 17.54). Treatment of chronic pelvic pain syndrome is challenging but some patients respond to a combination of $\alpha$-blockers, NSAIDs and amitriptyline.

### TUMOURS OF THE KIDNEY AND URINARY TRACT

Several malignant tumours can affect the kidney and urinary tract, including renal adenocarcinoma, bladder carcinoma, prostate carcinoma, and tumours of the testis and penis. The urogenital tract can also be affected by benign tumours and secondary tumour deposits, which can cause obstructive uropathy.

### Renal adenocarcinoma

Renal adenocarcinoma is by far the most common malignant tumour of the kidney in adults, making up 2.5% of all adult cancers, with a prevalence of 16 cases per 100000 population. It is twice as common in males as in females. The peak incidence is between 65 and 75 years of age and it is uncommon before 40. The tumour arises from renal tubular cells. Haemorrhage and necrosis give the cut surface a characteristic mixed golden-yellow and red appearance (Fig. 17.32B). Microscopically, clear cell carcinomas are the most common histological subtype (85%), with papillary, chromophobe and collecting duct tumours making up the remaining 15%. With clear cell...
Kidney and urinary tract disease

Open surgery. Partial nephrectomy is recommended for tumours of 4 cm or less, as there is a lower incidence of cardiac and renal morbidity. Patients at high operative risk, who have small tumours, may also be treated percutaneously by cryoablation or radiofrequency ablation. Surgery may also play a role in the treatment of solitary metastases, since these can remain single for long periods and excision may be worthwhile.

Renal adenocarcinoma is resistant to most chemotherapeutic agents. For many years, cytokine therapy with interferon and interleukin-2 was used in metastatic renal cancer but, in recent years, two new classes of targeted drugs have been introduced and are now the mainstay of therapy. These are the tyrosine kinase inhibitors, sunitinib and pazopanib, and the mammalian target of rapamycin (mTOR) inhibitors, temsirolimus and everolimus.

In previous years, patients who presented with distant metastases were treated with cytoreductive nephrectomy, in which nephrectomy was coupled with systemic cytokine treatment, since this was shown to improve survival as compared with either treatment in isolation. It is, at present, unclear whether this survival benefit still prevails with the newer agents mentioned above.

Renal adenocarcinoma is resistant to most chemotherapeutic agents. For many years, cytokine therapy with interferon and interleukin-2 was used in metastatic renal cancer but, in recent years, two new classes of targeted drugs have been introduced and are now the mainstay of therapy. These are the tyrosine kinase inhibitors, sunitinib and pazopanib, and the mammalian target of rapamycin (mTOR) inhibitors, temsirolimus and everolimus.

In previous years, patients who presented with distant metastases were treated with cytoreductive nephrectomy, in which nephrectomy was coupled with systemic cytokine treatment, since this was shown to improve survival as compared with either treatment in isolation. It is, at present, unclear whether this survival benefit still prevails with the newer agents mentioned above.

Studies that antedate the introduction of these new agents show that, if the tumour is confined to the kidney, 5-year survival is 75%, but this falls to 5% when there are distant metastases.

Clinical features

In 50% of patients, asymptomatic renal tumours are identified as an incidental finding during imaging investigations carried out for other reasons. Amongst symptomatic patients, about 60% present with haematuria, 40% with loin pain and 25% with a mass; about 10% present with a triad of pain, haematuria and a mass. A remarkable range of systemic effects may be present, including fever, raised ESR, polycythaemia, disorders of coagulation, hypercalcaemia, and abnormalities of plasma proteins and liver function tests. The patient may present with pyrexia of unknown origin or, rarely, with neuropathy. Some of these systemic effects are caused by secretion of products by the tumour, such as renin, erythropoietin, PTH-related protein (PTHrP) and gonadotrophins. The effects disappear when the tumour is removed but may re-appear when metastases develop, and so can be used as markers of tumour activity.

Investigations

Ultrasound is the investigation of first choice and allows differentiation between solid tumour and simple renal cysts. If the results are suggestive of a tumour, contrast-enhanced CT of the abdomen and chest should be performed for staging (see Fig. 17.32A). For tumours more than 3 cm in diameter with no evidence of metastatic spread and when the nature of the lesion is uncertain, ultrasound or CT-guided biopsy may be used to avoid nephrectomy for benign disease.

Management

Radical nephrectomy that includes the perirenal fascial envelope and ipsilateral para-aortic lymph nodes is the treatment of choice. Nephrectomy is commonly performed laparoscopically, with equivalent outcomes to open surgery. Partial nephrectomy is recommended for tumours of 4 cm or less, as there is a lower incidence of cardiac and renal morbidity. Patients at high operative risk, who have small tumours, may also be treated percutaneously by cryoablation or radiofrequency ablation. Surgery may also play a role in the treatment of solitary metastases, since these can remain single for long periods and excision may be worthwhile.

Renal adenocarcinoma is resistant to most chemotherapeutic agents. For many years, cytokine therapy with interferon and interleukin-2 was used in metastatic renal cancer but, in recent years, two new classes of targeted drugs have been introduced and are now the mainstay of therapy. These are the tyrosine kinase inhibitors, sunitinib and pazopanib, and the mammalian target of rapamycin (mTOR) inhibitors, temsirolimus and everolimus.

In previous years, patients who presented with distant metastases were treated with cytoreductive nephrectomy, in which nephrectomy was coupled with systemic cytokine treatment, since this was shown to improve survival as compared with either treatment in isolation. It is, at present, unclear whether this survival benefit still prevails with the newer agents mentioned above.

Studies that antedate the introduction of these new agents show that, if the tumour is confined to the kidney, 5-year survival is 75%, but this falls to 5% when there are distant metastases.

Urothelial tumours

Tumours arising from the transitional epithelium of the renal tract can affect the renal pelvis, ureter, bladder or urethra. They are rare under the age of 40, affect males about 3-4 times as commonly as females, and account for about 3% of all malignant tumours. The bladder is by far the most frequently affected site. Although almost all tumours are transitional cell carcinomas, squamous carcinoma may occur in urothelium that has undergone metaplasia, usually following chronic inflammation or irritation due to stones or schistosomiasis. The
appearance of a transitional cell tumour ranges from a delicate papillary structure with relatively good prognosis to a solid ulcerating mass in more aggressive disease (Fig. 17.33).

Pathophysiology
Risk factors include cigarette smoking and exposure to industrial carcinogens like aromatic amines, aniline dyes and aldehydes. Chronic inflammatory processes, such as schistosomiasis and chronic bladder stones, predispose to squamous carcinomas by causing squamous metaplasia.

Clinical features
More than 80% of patients present with painless, visible haematuria. It should be assumed that such bleeding is from a tumour until proven otherwise (p. 474). Tumours of the ureter or bladder may also cause symptoms of obstruction, depending on the site of involvement, and tumours of the bladder present with dysuria or storage symptoms. Physical examination is usually unremarkable, except in patients with very advanced disease, when bimanual examination may reveal a palpable mass.

Investigations
Cystoscopy (usually flexible cystoscopy under a local anaesthetic) is mandatory to evaluate the bladder in cases of haematuria or suspected bladder cancer. Imaging of the upper urinary tract (CT urogram is the gold standard but IVU and renal ultrasound are also acceptable) is also important to rule out abnormalities of the kidney, ureters and renal pelvis in patients with haematuria. If a suspicious defect is seen on CTU or IVU in the ureter or renal pelvis, a retrograde ureteropyelogram, ureteroscopy and biopsy are required. If evidence of a solid invasive urothelial tumour is found, CT of the abdomen, pelvis and chest should be performed as a staging procedure.

Management
Most bladder tumours are low-grade superficial lesions that can be successfully treated endoscopically by transurethral resection of the tumour. Intravesical chemotherapy with mitomycin C is usually administered as a one-off treatment post resection to prevent tumour recurrence, or may be given as a prolonged course to treat multiple low-grade bladder tumours. Patients with carcinoma in situ have a high risk of progression to invasive cancer. These patients often respond well to intravesical bacille Calmette–Guérin (BCG) treatment but more radical treatment may also be needed if this is unsuccessful. Following initial treatment and endoscopic clearance of bladder tumours, regular check cystoscopies are required to look for evidence of recurrence. Patients with recurrences of superficial disease can usually be treated by further resection and diathermy, but if this is unsuccessful, a cystectomy may be needed.

The management of invasive bladder tumours involves radical cystectomy with urinary diversion into an incontinent ileal conduit or a continent catheterisable bowel pouch; the latter is usually reserved for patients under the age of 70 years.

The prognosis of bladder tumours depends on tumour stage and grade. About 5% of patients with low-grade superficial bladder cancer progress to develop invasion of the bladder muscle, compared with about 50% of those with high-grade superficial bladder cancers. Overall, the 5-year survival for patients with muscle-invasive bladder cancer of either grade is about 50–70%.

Transitional cell carcinoma of the renal pelvis and ureter is usually treated by open or laparoscopic nephroureterectomy, but if the tumour is solitary and low-grade, it can be treated endoscopically.

Prostate cancer
Prostate cancer is the most common malignancy in men in the UK, with a prevalence of 105 per 100000 population. It is also common in northern Europe and the USA (particularly in the black population) but is rare in China and Japan. It rarely occurs before the age of 50 and has a mean age at presentation of 70 years.

Pathophysiology
Prostate cancers tend to arise within the peripheral zone of the prostate and almost all are adenocarcinomas. Metastatic spread to pelvic lymph nodes occurs early and metastases to bone, mainly the lumbar spine and pelvis, are common. Genetic factors are known to play an important role in pathogenesis, and multiple predisposing loci have been found to predispose to the disease in genome-wide association studies. A family history of prostate cancer greatly increases a man’s chances of developing the disease. It seems likely that the differences in incidence between different ethnic groups may be explained by genetic factors.

Clinical features
Most patients either are asymptomatic or present with lower urinary tract symptoms indistinguishable from benign prostatic enlargement. On digital rectal examination (DRE), the prostate may feel nodular and stony-hard, and the median sulcus may be lost, but up to 45% of tumours are impalpable. Symptoms and signs due to

Fig. 17.33 Transitional cell carcinoma of the bladder. Stages are shown from carcinoma in situ (Cis) to invasive tumour progressing beyond the bladder and prostate (T4b).
metastases are much less common at the initial presentation but may include back pain, weight loss, anaemia and obstruction of the ureters.

**Investigations**

Measurement of PSA levels in a peripheral blood sample, together with DRE, is the cornerstone of diagnosis. Prior to a PSA test, men should be given careful counselling about the limitations of the test: namely, a normal level does not exclude prostate cancer; a high value does not confirm the diagnosis but will open a discussion about biopsy. The need for radical treatment of localised prostate cancer is still not established; radical treatments have significant potential morbidity and mortality; and early identification and treatment of prostate cancer may save lives. Current evidence suggests that population-based screening for prostate cancer with PSA is of limited value, due in part to the fact that over 1000 patients would need to be screened to cure one man of prostate cancer. Individuals suspected of having prostate cancer, based on an elevated PSA and/or abnormal DRE, should undergo transrectal ultrasound-guided prostate biopsies. About 40% of patients with a serum PSA of 4.0–10 µg/L or more will have prostate cancer on biopsy, although 25% patients with a PSA of less than 4 ng/mL may also have prostate cancer. Occasionally, a small focus of tumour is found incidentally in patients undergoing transurethral resection of the prostate for benign hyperplasia. If the diagnosis of prostate cancer is confirmed, staging should be performed by pelvic MRI to assess the presence and extent of local involvement. An isotope bone scan should be carried out if distant metastases are suspected (rare if the PSA is below 20 ng/mL); very high levels of serum PSA (> 100 ng/mL) almost always indicate distant bone metastases. Following diagnosis, serial assessment of PSA levels is useful for monitoring response to treatment and disease progression.

**Management**

Tumour confined to the prostate is potentially curable by radical prostatectomy, radical radiotherapy or brachytherapy (implantation of small radioactive particles into the prostate). These options should only be considered in patients with more than 10 years’ life expectancy. Patients who are found to have small foci of tumour do not require specific treatment but should be followed up periodically with PSA, DRE and a schedule of biopsies; this is known as active surveillance. Prostatic cancer, like breast cancer, is sensitive to steroid hormones; metastatic prostate cancer is treated by androgen depletion, involving either surgery (orchidectomy) or, more commonly, androgen-suppressing drugs (Box 17.59). Androgen receptor blockers, such as bicalutamide or cyproterone acetate, may also prevent tumour cell growth. Gonadotrophin-releasing hormone (GnRH) analogues, such as goserelin, continuously occupy pituitary receptors, preventing them from responding to the GnRH pulses that normally stimulate luteinising hormone (LH) and follicle-stimulating hormone (FSH) release. This initially causes an increase in testosterone before producing a prolonged reduction, and for this reason the initial dose must be covered with an androgen receptor blocker to prevent a tumour flare.

**EBM 17.59 Hormone manipulation in prostate cancer**

> ‘Reducing circulating testosterone levels (either by castration or by medication) results in a 70% initial response rate. Additional androgen blockade produces a small increase in survival but with poorer quality of life.’


For further information: [www.cochrane.org/cochrane-reviews](http://www.cochrane.org/cochrane-reviews)

A small proportion of patients fail to respond to endocrine treatment. A larger number respond for a year or two but then the disease progresses. Chemotherapy with docetaxel can then be effective and provide a modest (around 3 months) survival advantage. Radiotherapy is useful for localised bone pain but the basis of treatment remains pain control by analgesia (p. 286). Provided that patients do not die of another cause, the 10-year survival rate of patients with tumours localised to the prostate is 95%, but if metastases are present, this falls to 10%. Life expectancy is not reduced in patients with small foci of tumour.

**Testicular tumours**

Testicular tumours are uncommon, with a prevalence of 5 cases per 100000 population. They occur mainly in young men aged between 20 and 40 years. They often secrete α-fetoprotein (AFP) and β-human chorionic gonadotrophin (β-hCG), which are useful biochemical markers for both diagnosis and prognosis. Seminoma and teratoma account for 85% of all tumours of the testis. Leydig cell tumours are less common.

Seminomas arise from seminiferous tubules and represent a relatively low-grade malignancy. Metastases can occur through lymphatic spread, however, and typically involve the lungs. Teratomas arise from primitive germinal cells and tend to occur at a younger age than seminomas. They may contain cartilage, bone, muscle, fat and a variety of other tissues, and are classified according to the degree of differentiation. Well-differentiated tumours are the least aggressive; at the other extreme, trophoblastic teratoma is highly malignant. Occasionally, teratoma and seminoma occur together.

Leydig cell tumours are usually small and benign but secrete oestrogens, leading to presentation with gynaecomastia (p. 762).

**Clinical features and investigations**

The common presentation is incidental discovery of a painless testicular lump, although some patients complain of a testicular ache.

All suspicious scrotal lumps should be imaged by ultrasound. Serum levels of AFP and β-hCG are elevated in extensive disease. Oestradiol may be elevated, suppressing luteinising hormone (LH), follicle-stimulating hormone (FSH) and testosterone. Accurate staging is based on CT of the lungs, liver and retroperitoneal area.
Management and prognosis

The primary treatment is surgical orchidectomy. Subsequent treatment depends on the histological type and stage. Radiotherapy is the treatment of choice for early-stage seminoma. Teratoma confined to the testes may be managed conservatively, but more advanced cancers are treated with chemotherapy, usually the combination of bleomycin, etoposide and cisplatin. Follow-up is by CT and assessment of AFP and β-hCG. Retroperitoneal lymph node dissection is now only performed for residual or recurrent nodal masses.

The 5-year survival rate for patients with seminoma is 90–95%. For teratomas, the 5-year survival varies between 60% and 95%, depending on tumour type, stage and volume.

Inherited tumour syndromes affecting the renal tract

Some uncommon autosomal dominantly inherited conditions are associated with multiple renal tumours in adult life. In tuberous sclerosis (p. 1302), replacement of renal tissue by multiple angiomyolipomas (tubers) may occasionally cause renal failure in adults. Other organs affected include the skin (adenoma sebaceum on the face) and brain (causing seizures and mental retardation). The von Hippel–Lindau syndrome (p. 1216) is connected with multiple renal cysts, renal adenomas and renal adenocarcinoma. Other organs affected include the central nervous system (haemangioblastomas) and the adrenals (pheochromocytoma).

RENAL INVOLVEMENT IN SYSTEMIC CONDITIONS

The kidneys may be directly involved in a number of multisystem diseases or secondarily affected by diseases of other organs. Involvement may be at a pre-renal, renal (glomerular or interstitial) or post-renal level. Many of the diseases are described in other sections of this chapter or in other chapters of the book.

Diabetes mellitus

Diabetes mellitus is the most common cause of CKD in developed countries. In patients with diabetes, there is a steady advance from microalbuminuria to dipstick-positive proteinuria, and a progression to renal failure, as described on page 830. Few patients require renal biopsy to establish the diagnosis, but atypical features or progression should lead to suspicion that an alternative condition could be present.

Management with ACE inhibitors and other hypotensive agents to slow progression is described on page 831 and has been dramatically effective. In some patients, proteinuria may be eradicated and progression completely halted, even if renal function is abnormal.

Hepatic–renal disease

Severe hepatic dysfunction may cause a haemodynamically mediated type of renal failure, hepatorenal syndrome, described on page 940. Patients with chronic liver disease are also predisposed to develop AKI (acute tubular necrosis) in response to relatively minor insults, including bleeding and infection. Such patients are often difficult to treat by dialysis and have a poor prognosis. Where treatment is justified – for example, if there is a good chance of recovery or of a liver transplant – slow or continuous treatments are less likely to precipitate or exacerbate hepatic encephalopathy. IgA nephropathy (p. 500) is more common in patients with liver disease.

Pulmonary-renal syndrome

The pulmonary–renal syndrome is a dramatic presentation with renal and respiratory failure that is not explained by excess intravascular fluid or by severe pneumonia; it occurs in Goodpasture’s (anti-GBM) disease and small-vessel vasculitis (see below). There are some other uncommon causes of a similar syndrome, including poisoning with the herbicide paraquat.

Malignant diseases

The kidney may be affected in many different ways in patients with malignant disease (Box 17.60). Direct involvement of the kidneys or other parts of the urinary tract can occur, causing obstructive uropathy, and glomerulonephritis may occur, presumably as the result of an immunological reaction to the tumour. Hypercalcaemia can be caused by parathyroid hormone-related protein (PTHrP) production by tumours, whereas treatment of leukaemia and lymphoma can sometimes be complicated by interstitial nephritic and acute kidney injury due to uric acid release from necrotic tumour cells. Finally, in myeloma and other monoclonal gammopathies, renal impairment can occur as the result of the nephrotoxic effects of immunoglobulin light chains released by the tumour.

Systemic vasculitis

Small-vessel vasculitis (p. 1115) commonly affects the kidneys, with rapid and profound impairment of glomerular function. Histologically, there is a focal inflammatory glomerulonephritis, usually with focal necrosis (see Box 17.39, p. 499, and Fig. 17.22, p. 498) and often with crescentic changes. Typically, the patient is systemically unwell with an acute phase response, weight loss and arthralgia. In some patients, pulmonary haemorrhage may occur, which can be life-threatening. In others, it presents as a kidney-limited disorder, with
Kidney and urinary tract disease

**Systemic lupus erythematosus**

Subclinical renal involvement, with low-level haematuria and proteinuria but minimally impaired or normal renal function, is common in systemic lupus erythematosus (SLE). Usually, this is due to glomerular disease, although interstitial nephritis may also occur, particularly in patients with overlap syndromes such as mixed connective tissue disease and Sjögren’s syndrome (p. 499 and Figure 17.21 (p. 497)).

Almost any histological pattern of glomerular disease can be observed in SLE and the clinical presentation ranges from florid, rapidly progressive glomerulonephritis to nephrotic syndrome. The most common presentation is with subacute disease and inflammatory features (haematuria, hypertension, variable renal impairment), accompanied by heavy proteinuria that often reaches nephrotic levels. In severely affected patients, the most common histological pattern is a proliferative glomerulonephritis with substantial deposits of immunoglobulins on immunofluorescence. Randomised controlled trials have shown that the risk of ESRD in lupus nephritis is significantly reduced by high-dose steroids administered in combination with cyclophosphamide, usually given as regular intravenous pulses. More recently, it has been shown that the combination of corticosteroids and mycophenolate mofetil is as effective as corticosteroids and cyclophosphamide, at least in short-term studies.

Many patients with SLE who develop ESRD go into remission, possibly because of immunosuppression related to the ESRD. Patients with ESRD caused by SLE are usually good candidates for dialysis and transplantation. Although it may recur in renal allografts, the immunosuppression required to prevent allograft rejection usually controls the SLE too.

**Pregnancy and renal disease**

Pregnancy has important physiological effects on the renal system. Some diseases are more common in pregnancy, the manifestations of others are modified during pregnancy, and a few diseases, such as pre-eclampsia, are unique to pregnancy.

Physiological adaptations begin in the first few weeks. Peripheral vascular resistance declines, blood volume, cardiac output and GFR increase, and there is usually a reduction in blood pressure and plasma creatinine and urea values in the first trimester. Baseline blood pressure and urine testing from the first antenatal clinic visit are valuable if problems arise later.

**Infections**

Pyelonephritis is more common during pregnancy, perhaps because of dilatation of the urinary collecting system and ureters. It is important to treat asymptomatic bacteriuria during pregnancy, since this can progress to pyelonephritis, which, in turn, can trigger premature labour (Box 17.62 and p. 513).

**Glomerular diseases**

Proteinuria caused by glomerular disease is usually exacerbated, and nephrotic syndrome may develop

---

**EBM 17.61 Role of rituximab in ANCA-associated vasculitis**

‘The RAVE study compared the effectiveness of rituximab and oral cyclophosphamide for inducing remission in 197 patients with ANCA-associated vasculitis with renal and/or pulmonary involvement. Both groups received high-dose glucocorticoids. The rituximab-based regimen was non-inferior to the cyclophosphamide-based regimen at inducing remission and was superior in patients with a history of relapsing disease. Rituximab was equally effective as cyclophosphamide in the treatment of major renal disease and alveolar haemorrhage. There were no significant differences between the treatment groups with respect to glucocorticoid dose or adverse events.’

without any alteration in the underlying disease activity in individuals who had only slight proteinuria before pregnancy. This further increases the risk of venous thromboembolism, the leading cause of maternal deaths in developed countries.

**Autoimmune diseases**

Systemic autoimmune diseases typically are relatively quiescent during pregnancy but tend to relapse in the first few weeks and months following delivery. Pre-existing renal disease increases the fetal and maternal risk involved in pregnancy, to a degree dependent on the level of renal function, proteinuria and hypertension. Patients with such diseases who may become pregnant should be aware of the associated risks. During pregnancy, therapy should not usually be stopped, but blood pressure targets may be modified and agents altered to those of proven safety.

**Pre-eclampsia**

Pre-eclampsia is a systemic disorder that occurs in or near the third trimester of pregnancy (triplets > twins > singleton). The cause is unknown, although a number of risk factors have been identified (Box 17.63).

**Clinical features**

Pre-eclampsia is traditionally defined by the triad of oedema, proteinuria and hypertension. However, oedema is common in late pregnancy, proteinuria is a late sign and, while hypertension is usually present, it may be relative, mild or even absent. Furthermore, all these features occur in pre-existing renal disease exacerbated by pregnancy.

It is important to distinguish pre-eclampsia from pre-existing renal disease, since this affects management. Pre-eclampsia presents progressively, increasing risks to mother and fetus, which can be reversed almost immediately by early delivery. In contrast, continuing the pregnancy for as long as possible in patients with pre-existing renal disease, may permit delivery of a healthier, more mature baby. Proteinuria and hypertension in the first trimester of pregnancy suggest pre-existing renal disease.

**Management**

The only effective management for pre-eclampsia is delivery. The role of antiplatelet therapy (low-dose aspirin) remains controversial. Hypertension is a consequence and not the cause of the disorder, and treatment is only justified to lower it from severe and immediately dangerous levels (> 150–160/100–110 mmHg). Treating lower levels has been shown to confer no benefit and exposes the fetus to additional drugs. If life-threatening complications are not present and the baby is immature, corticosteroids may be given to induce maturation of fetal lungs, and delivery postponed while mother and baby are closely observed. Magnesium sulphate reduces the incidence of eclamptic convulsions.

**Acute kidney injury**

Maternal AKI may occur in almost any of the pre-eclamptic syndromes. Worldwide, a more important cause of AKI is septic abortion, when the uterus becomes infected because of retained products of conception or poor sterility in an often illegally induced abortion. Renal function is usually recoverable, but AKI in pregnancy is particularly prone to progress to cortical necrosis, with incomplete or total failure to recover renal function.

---

### 17.63 Pre-eclampsia and related diseases in pregnancy

**Clinical syndromes**

- Eclampsia: severe hypertension, encephalopathy and fits
- Disseminated intravascular coagulation
- Thrombotic microangiopathy: may also occur post-partum (post-partum haemolytic uraemic syndrome)
- Acute fatty liver of pregnancy
- ‘HELLP’ syndrome: haemolysis, elevated liver enzymes, low platelets (thrombotic microangiopathy with abnormal liver function)

**Risk factors**

- First pregnancy
- First pregnancy with a new partner or long inter-pregnancy interval
- Pre-eclampsia in previous pregnancies
- Age < 20 yrs or > 35 yrs
- Multiple pregnancy (triplets > twins > singleton)
- Pre-existing hypertension
- Pre-existing renal disease

**Clinical signs**

- Hypertension
- Proteinuria
- Oedema

**Investigations**

- Uric acid levels increased (before renal impairment apparent)
- Platelets decreased

- Evidence of the clinical syndromes listed above

- Reduced GFR (late)
- Fetus small for dates and growing slowly
- Fetal distress (late)

### 17.64 Kidney disease in adolescence

**Adherence:** young adults moving from parental supervision may become disengaged. There may also be reduced adherence to prophylactic and therapeutic treatment.

**Adverse events:** there is an increased risk of transplant loss and other adverse events in young adults on RRT.

**Management:** joint clinics should be established with the paediatric team to facilitate transfer to regional specialist clinics.
## Drug-induced renal disease

The kidney is susceptible to damage by drugs because it is the route of excretion of many water-soluble compounds, including drugs and their metabolites. Some may reach high concentrations in the renal cortex as a result of proximal tubular transport mechanisms. Others are concentrated in the medulla by the operation of the countercurrent system. The same applies to certain toxins.

Toxic renal damage may occur by a variety of mechanisms (Box 17.65). Very commonly, drugs contribute to the development of acute tubular necrosis as one of multiple insults. Numerically, reactions to NSAIDs and ACE inhibitors are the most important. Haemodynamic renal impairment, acute tubular necrosis and allergic reactions are usually reversible if recognised early enough. Other types, however, especially those associated with extensive fibrosis, are less likely to be reversible.

### Mechanisms and examples of drug-induced renal disease/dysfunction

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Drug or toxin</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haemodynamic</td>
<td>NSAIDs</td>
<td>Reduced renal blood flow due to inhibition of prostaglandin synthesis</td>
</tr>
<tr>
<td></td>
<td>ACE inhibitors</td>
<td>Reduce efferent glomerular arteriolar tone. Toxic in the presence of renal artery stenosis and other conditions of renal hypoperfusion</td>
</tr>
<tr>
<td></td>
<td>Radiographic contrast media</td>
<td>Multifactorial aetiology may include intense vasoconstriction</td>
</tr>
<tr>
<td>Acute tubular necrosis</td>
<td>Aminoglycosides, amphotericin</td>
<td>In most examples, there is evidence of direct tubular toxicity but haemodynamic and other factors probably contribute</td>
</tr>
<tr>
<td></td>
<td>Paracetamol overdose</td>
<td>May occur with or without serious hepatotoxicity</td>
</tr>
<tr>
<td></td>
<td>Radiographic contrast media</td>
<td>May be secondary to precipitation in tubules. Furosemide is a co-factor</td>
</tr>
<tr>
<td>Loss of tubular/collecting duct function</td>
<td>Lithium</td>
<td>Dose-related, partially reversible loss of concentrating ability. Occurs at lower exposures than cause acute tubular necrosis</td>
</tr>
<tr>
<td></td>
<td>Cisplatin</td>
<td>Aminoglycosides, amphotericin</td>
</tr>
<tr>
<td>Glomerulonephritis</td>
<td>Penicillamine, gold</td>
<td>Membranous nephropathy</td>
</tr>
<tr>
<td>(immune-mediated)</td>
<td>Penicillamine</td>
<td>Crescentic or focal necrotising glomerulonephritis in association with ANCA and systemic small-vessel vasculitis</td>
</tr>
<tr>
<td></td>
<td>NSAIDs</td>
<td>Minimal change nephropathy</td>
</tr>
<tr>
<td>Interstitial nephritis</td>
<td>NSAIDs, penicillins, proton pump inhibitors, many others</td>
<td>Acute interstitial nephritis</td>
</tr>
<tr>
<td>(immune-mediated)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interstitial nephritis</td>
<td>Lithium</td>
<td>As a consequence of acute toxicity. Otherwise controversial</td>
</tr>
<tr>
<td>(toxicity)</td>
<td>Ciclosporin, tacrolimus</td>
<td>The major problem with these drugs</td>
</tr>
<tr>
<td>Interstitial nephritis</td>
<td>Various NSAIDs (p. 505)</td>
<td>Ischaemic damage secondary to NSAID effects on renal blood flow</td>
</tr>
<tr>
<td>(with papillary necrosis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubular obstruction (crystal formation)</td>
<td>Aciclovir</td>
<td>Crystals of the drug form in tubules. Aciclovir is now more common than the original example of sulphonamides</td>
</tr>
<tr>
<td></td>
<td>Chemotherapy</td>
<td></td>
</tr>
<tr>
<td>Nephrocalcinosis</td>
<td>Oral sodium phosphate-containing bowel cleansing agents</td>
<td>Precipitation of calcium phosphate occurring in 1–4% and exacerbated by volume depletion. Usually mild but damage can be irreversible</td>
</tr>
<tr>
<td>Retroperitoneal fibrosis</td>
<td>Ergolinic dopamine agonists (cabergoline), methysergide*, practolol*</td>
<td>Idiopathic is more common (p. 511)</td>
</tr>
</tbody>
</table>

*These drugs are no longer in use in the UK.

### NSAIDs

Impairment of renal function may develop in patients on NSAID, since prostaglandins play an important role in regulating renal blood flow. This is particularly likely in patients with other disorders, such as heart failure, cirrhosis, sepsis and pre-existing renal impairment. In addition, idiosyncratic immune reactions may occur, causing minimal change nephrotic syndrome (p. 498) and acute interstitial nephritis (p. 502). Analgesic nephropathy (p. 504) is now a rare complication of long-term use.

### ACE inhibitors

These abolish the compensatory angiotensin II-mediated vasoconstriction of the glomerular efferent arteriole that takes place in order to maintain glomerular perfusion pressure distal to a renal artery stenosis and in renal hypoperfusion (see Fig. 17.1, p. 465). Monitoring of renal function before and after initiation of therapy is essential.
Prescribing in renal disease
Many drugs and drug metabolites are excreted by the kidney and so the presence of renal impairment alters the required dose and frequency (p. 36).

Further information and acknowledgements
Websites
www.edren.org Renal Unit, Royal Infirmary of Edinburgh; information about individual diseases, protocols for immediate in-hospital management and more.
www.edrep.org/resources Educational resources.
www.nephron.com The links under ‘Physicians’ include useful urology pages, eGFR and other calculators, and other resources.

www.renal.org/ckd UK Renal Association; current UK guidelines on the detection, referral and management of chronic kidney disease.

Figure acknowledgements
Fig. 17.19 Beutler JJ, Koomans HA. Malignant hypertension: still a challenge. Nephrol Dial Transplant 1977; 12:2019–2023; photograph courtesy of Prof PJ Slootweg, University Hospital, Utrecht. By permission of Oxford University Press.
Clinical examination of the cardiovascular system 526

Functional anatomy and physiology 528
  Anatomy 528
  Physiology 530

Investigation of cardiovascular disease 532
  Electrocardiogram 532
  Cardiac biomarkers 535
  Chest X-ray 535
  Echocardiography (echo) 536
  Computed tomographic imaging 537
  Magnetic resonance imaging 537
  Cardiac catheterisation 538
  Electrophysiology study 539
  Radionuclide imaging 539

Presenting problems in cardiovascular disease 539
  Chest pain 539
  Breathlessness (dyspnoea) 543
  Acute circulatory failure (cardiogenic shock) 544
  Heart failure 546
  Syncope and presyncope 554
  Palpitation 556
  Cardiac arrest and sudden cardiac death 557
  Abnormal heart sounds and murmurs 560

Disorders of heart rate, rhythm and conduction 562
  Sinoatrial nodal rhythms 563
  Atrial tachyarrhythmias 564
  ‘Supraventricular’ tachycardias 567
  Ventricular tachyarrhythmias 569
  Atrioventricular and bundle branch block 571
  Anti-arrhythmic drug therapy 573
  Therapeutic procedures 577

Atherosclerosis 579

Coronary artery disease 583
  Stable angina 583
  Acute coronary syndrome 589
  Cardiac risk of non-cardiac surgery 600

Vascular disease 600
  Peripheral arterial disease 600
  Diseases of the aorta 603
  Hypertension 607

Diseases of the heart valves 613
  Rheumatic heart disease 614
  Mitral valve disease 616
  Aortic valve disease 620
  Tricuspid valve disease 624
  Pulmonary valve disease 625
  Infective endocarditis 625
  Valve replacement surgery 629

Congenital heart disease 629

Diseases of the myocardium 636
  Myocarditis 636
  Cardiomyopathy 636
  Specific diseases of heart muscle 638
  Cardiac tumours 639

Diseases of the pericardium 639
CLINICAL EXAMINATION OF THE CARDIOVASCULAR SYSTEM

1. **Hands**
   - Clubbing
   - Splinter haemorrhages and other stigmata of infective endocarditis

2. **Radial pulse**
   - Rate
   - Rhythm

3. **Blood pressure**

4. **Carotid pulses**
   - Volume
   - Character
   - Bruits

5. **Jugular venous pulse**
   - (see opposite)
   - Height
   - Waveform

6. **Face, mouth and eyes**
   - Pallor
   - Central cyanosis
   - Malar flush
   - Dental caries
   - Fundi (retinopathy)
   - Stigmata of hyperlipidaemia and thyroid disease

7. **Precordium**
   - Inspect
   - Palpate
   - (see opposite)

8. **Auscultation**
   - (see opposite)

9. **Back**
   - Lung crepitations
   - Sacral oedema

10. **Abdomen**
    - Hepatomegaly
    - Ascites
    - Aortic aneurysm
    - Bruits

11. **Tendon xanthomas**
    - (hyperlipidaemia)

12. **Femoral pulses**
    - Radio-femoral delay
    - Bruits

13. **Legs**
    - Peripheral pulses
    - Oedema

Observation

**Symptoms and well-being**
- Breathlessness
- Distress etc.

**Body habitus**
- Body mass (obesity, cachexia)
- Marfan’s and other syndromes

**Tissue perfusion**
- Skin temperature
- Sweating
- Urine output

4 Examination of the arterial pulse

- The character of the pulse is determined by stroke volume and arterial compliance, and is best assessed by palpating a major artery, such as the carotid or brachial artery.
- Aortic regurgitation, anaemia, sepsis and other causes of a large stroke volume typically produce a bounding pulse with a high amplitude and wide pulse pressure (panel A).
- Aortic stenosis impedes ventricular emptying and may cause a slow-rising, weak and delayed pulse (panel A).
- Normal sinus rhythm produces a pulse that is regular in time and force. Arrhythmias may cause irregularity. Atrial fibrillation produces a pulse that is irregular in time and volume (panel B).

5 Examination of the jugular venous pulse

The internal jugular vein, superior vena cava and right atrium are in continuity, so the height of the jugular venous pulsation reflects right atrial pressure. When the patient is placed at 45°, with the head supported and turned a few degrees to the left, the jugular venous pulse is visible along the line of the sternocleidomastoid muscle (see opposite).
- The height of the jugular venous pulse is determined by right atrial pressure and is therefore elevated in right heart failure and reduced in hypovolaemia.
- If the jugular venous pulse is not easily seen, it may be highlighted by gentle pressure on the abdomen.
- In normal sinus rhythm, the two venous peaks, the a and v waves, approximate to atrial and ventricular systole respectively.
- The x descent reflects atrial relaxation and apical displacement of the tricuspid valve ring. The y descent reflects atrial emptying early in diastole. These signs are subtle.
- Tricuspid regurgitation produces giant v waves that coincide with ventricular systole.

6 Auscultation of the heart

- Use the diaphragm to examine at the apex, lower left sternal edge (tricuspid area) and upper left (pulmonary area) and right (aortic area) sternal edges.
- Use the bell to examine low-pitched noises, particularly at the apex for mid-diastolic murmurs.
- Time the sounds and murmurs by feeling the carotid pulse; systolic murmurs are synchronous with the pulse.
- Listen for radiation of systolic murmurs, over the base of the neck (aortic stenosis) and in the axilla (mitral incompetence).
- Listen over the left sternal border with the patient sitting forward (aortic incompetence), then at the apex with the patient rolled on to the left side (mitral stenosis).

The haemodynamic effects of respiration are discussed on page 532.
See page 560 for analysis and interpretation of heart sounds and murmurs.
Cardiovascular disease is the most frequent cause of adult death in the Western world; in the UK, one-third of men and one-quarter of women will die as a result of ischaemic heart disease. In many developed countries, the incidence of ischaemic heart disease has been falling for the last two or three decades, but it is rising in Eastern Europe and Asia. Cardiovascular disease may thus soon become the leading cause of death on all continents.

Strategies for the treatment and prevention of heart disease can be highly effective and have been subjected to rigorous evaluation. The evidence base for the treatment of cardiovascular disease is stronger than for almost any other disease group.

Valvular heart disease is common, but the aetiology varies in different parts of the world. On the Indian subcontinent and in Africa, it is predominantly due to rheumatic fever, whereas calcific aortic valve disease is the most common problem in developed countries.

Prompt recognition of the development of heart disease is limited by two key factors. Firstly, it is often latent; coronary artery disease can proceed to an advanced stage before the patient notices any symptoms. Secondly, the diversity of symptoms attributable to heart disease is limited, so different pathologies may frequently present with the same symptoms.

**FUNCTIONAL ANATOMY AND PHYSIOLOGY**

### Anatomy

The heart acts as two serial pumps that share several electrical and mechanical components. The right heart circulates blood to the lungs where it is oxygenated, and the left heart receives this and circulates it to the rest of the body (Fig. 18.1). The atria are thin-walled structures that act as priming pumps for the ventricles, which provide most of the energy to the circulation. Within the mediastinum, the atria are situated posteriorly and the left atrium (LA) sits anterior to the oesophagus and descending aorta. The interatrial septum separates the two atria. In 20% of adults, a patent foramen ovale is found; this communication in the fetal circulation between the right and left atria normally closes at birth (p. 629). The right atrium (RA) receives blood from the superior and inferior venae cavae and the coronary sinus. The LA receives blood from four pulmonary veins, two from each of the left and right lungs. The ventricles are thick-walled structures, adapted to circulating blood through large vascular beds under pressure. The atria and ventricles are separated by the annulus fibrosus, which forms the skeleton for the atrioventricular (AV) valves and which electrically insulates the atria from the ventricles. The right ventricle (RV) is roughly triangular in shape and extends from the annulus fibrosus to near the cardiac apex, which is situated to the left of the midline. Its anterosuperior surface is rounded and convex, and its posterior extent is bounded by the interventricular septum, which bulges into the chamber. Its upper extent is conical, forming the conus arteriosus or outflow tract, from which the pulmonary artery arises. The RV sits anterior to, and to the right of, the left ventricle (LV). The LV is more conical in shape and in cross-section is nearly circular. It extends from the LA to the apex of the heart. The LV myocardium is normally around 10 mm thick (c.f. RV thickness of 2–3 mm) because it pumps blood at a higher pressure.

The normal heart occupies less than 50% of the trans-thoracic diameter in the frontal plane, as seen on a chest X-ray. On the patient’s left, the cardiac silhouette is formed by the aortic arch, the pulmonary trunk, the left atrial appendage and the LV. On the right, the RA is joined by superior and inferior venae cavae, and the lower right border is made up by the RV (Fig. 18.2). In disease states or congenital cardiac abnormalities, the

**Fig. 18.1** Direction of blood flow through the heart. The blue arrows show deoxygenated blood moving through the right heart to the lungs. The red arrows show oxygenated blood moving from the lungs to the systemic circulation. The normal pressures are shown for each chamber in mmHg.
posterior descending artery runs in the posterior interventricular groove and supplies the inferior part of the interventricular septum. This vessel is a branch of the RCA in approximately 90% of people (dominant right system) and is supplied by the CX in the remainder (dominant left system). The coronary anatomy varies greatly from person to person and there are many ‘normal variants’.

The RCA supplies the sinoatrial (SA) node in about 60% of individuals and the AV node in about 90%. Proximal occlusion of the RCA therefore often results in sinus bradycardia and may also cause AV nodal block. Abrupt occlusions in the RCA, due to coronary thrombosis, result in infarction of the inferior part of the LV and often the RV. Abrupt occlusion of the LAD or CX causes infarction in the corresponding territory of the LV, and occlusion of the left main coronary artery is usually fatal.

The venous system follows the coronary arteries but drains into the coronary sinus in the atrioventricular groove, and then to the RA. An extensive lymphatic system drains into vessels that travel with the coronary vessels and then into the thoracic duct.

Conducting system of the heart

The SA node is situated at the junction of the superior vena cava and RA (Fig. 18.4). It comprises specialised atrial cells that depolarise at a rate influenced by the autonomic nervous system and by circulating catecholamines. During normal (sinus) rhythm, this depolarisation wave propagates through both atria via sheets of atrial myocytes. The annulus fibrosus forms a conduction barrier between atria and ventricles, and the only pathway through it is the AV node. This is a midline structure, extending from the right side of the interatrial septum, penetrating the annulus fibrosus anteriorly. The AV node conducts relatively slowly, producing a necessary time delay between atrial and ventricular contraction. The His–Purkinje system is comprised of the bundle of His extending from the AV node into the interventricular septum, the right and left bundle branches passing along the ventricular septum and into the respective ventricles, the anterior and posterior fascicles

silhouette may change as a result of hypertrophy or dilatation.

The coronary circulation

The left main and right coronary arteries arise from the left and right sinuses of the aortic root, distal to the aortic valve (Fig. 18.3). Within 2.5 cm of its origin, the left main coronary artery divides into the left anterior descending artery (LAD), which runs in the anterior interventricular groove, and the left circumflex artery (CX), which runs posteriorly in the atrioventricular groove. The LAD gives branches to supply the anterior part of the septum (septal perforators) and the anterior, lateral and apical walls of the LV. The CX gives marginal branches that supply the lateral, posterior and inferior segments of the LV. The right coronary artery (RCA) runs in the right atrioventricular groove, giving branches that supply the RA, RV and inferoposterior aspects of the LV. The posterior descending artery runs in the posterior interventricular groove and supplies the inferior part of the interventricular septum. This vessel is a branch of the RCA in approximately 90% of people (dominant right system) and is supplied by the CX in the remainder (dominant left system). The coronary anatomy varies greatly from person to person and there are many ‘normal variants’.

The RCA supplies the sinoatrial (SA) node in about 60% of individuals and the AV node in about 90%. Proximal occlusion of the RCA therefore often results in sinus bradycardia and may also cause AV nodal block. Abrupt occlusions in the RCA, due to coronary thrombosis, result in infarction of the inferior part of the LV and often the RV. Abrupt occlusion of the LAD or CX causes infarction in the corresponding territory of the LV, and occlusion of the left main coronary artery is usually fatal.

The venous system follows the coronary arteries but drains into the coronary sinus in the atrioventricular groove, and then to the RA. An extensive lymphatic system drains into vessels that travel with the coronary vessels and then into the thoracic duct.

Conducting system of the heart

The SA node is situated at the junction of the superior vena cava and RA (Fig. 18.4). It comprises specialised atrial cells that depolarise at a rate influenced by the autonomic nervous system and by circulating catecholamines. During normal (sinus) rhythm, this depolarisation wave propagates through both atria via sheets of atrial myocytes. The annulus fibrosus forms a conduction barrier between atria and ventricles, and the only pathway through it is the AV node. This is a midline structure, extending from the right side of the interatrial septum, penetrating the annulus fibrosus anteriorly. The AV node conducts relatively slowly, producing a necessary time delay between atrial and ventricular contraction. The His–Purkinje system is comprised of the bundle of His extending from the AV node into the interventricular septum, the right and left bundle branches passing along the ventricular septum and into the respective ventricles, the anterior and posterior fascicles

silhouette may change as a result of hypertrophy or dilatation.

The coronary circulation

The left main and right coronary arteries arise from the left and right sinuses of the aortic root, distal to the aortic valve (Fig. 18.3). Within 2.5 cm of its origin, the left main coronary artery divides into the left anterior descending artery (LAD), which runs in the anterior interventricular groove, and the left circumflex artery (CX), which runs posteriorly in the atrioventricular groove. The LAD gives branches to supply the anterior part of the septum (septal perforators) and the anterior, lateral and apical walls of the LV. The CX gives marginal branches that supply the lateral, posterior and inferior segments of the LV. The right coronary artery (RCA) runs in the right atrioventricular groove, giving branches that supply the RA, RV and inferoposterior aspects of the LV. The posterior descending artery runs in the posterior interventricular groove and supplies the inferior part of the interventricular septum. This vessel is a branch of the RCA in approximately 90% of people (dominant right system) and is supplied by the CX in the remainder (dominant left system). The coronary anatomy varies greatly from person to person and there are many ‘normal variants’.

The RCA supplies the sinoatrial (SA) node in about 60% of individuals and the AV node in about 90%. Proximal occlusion of the RCA therefore often results in sinus bradycardia and may also cause AV nodal block. Abrupt occlusions in the RCA, due to coronary thrombosis, result in infarction of the inferior part of the LV and often the RV. Abrupt occlusion of the LAD or CX causes infarction in the corresponding territory of the LV, and occlusion of the left main coronary artery is usually fatal.

The venous system follows the coronary arteries but drains into the coronary sinus in the atrioventricular groove, and then to the RA. An extensive lymphatic system drains into vessels that travel with the coronary vessels and then into the thoracic duct.
of the left bundle branch, and the smaller Purkinje fibres that ramify through the ventricular myocardium. The tissues of the His–Purkinje system conduct very rapidly and allow near-simultaneous depolarisation of the entire ventricular myocardium.

Nerve supply of the heart

The heart is innervated by both sympathetic and parasympathetic fibres. Adrenergic nerves from the cervical sympathetic chain supply muscle fibres in the atri and ventricles and the electrical conducting system. Positive inotropic and chronotropic effects are mediated by β1-adrenoceptors, whereas β2-adrenoceptors predominate in vascular smooth muscle and mediate vasodilatation. Parasympathetic pre-ganglionic fibres and sensory fibres reach the heart through the vagus nerves. Cholinergic nerves supply the AV and SA nodes via muscarinic (M2) receptors. Under resting conditions, vagal inhibitory activity predominates and the heart rate is slow. Adrenergic stimulation, associated with exercise, emotional stress, fever and so on, causes the heart rate to increase. In disease states, the nerve supply to the heart may be affected. For example, in heart failure the sympathetic system may be up-regulated, and in diabetes mellitus the nerves themselves may be damaged (autonomic neuropathy, p. 831) so that there is little variation in heart rate.

Physiology

The circulation

The RA receives deoxygenated blood from the superior and inferior venae cavae and discharges blood to the RV, which in turn pumps it into the pulmonary artery. Blood passes through the pulmonary arterial and alveolar capillary bed, where it is oxygenated, then drains via four pulmonary veins into the LA. This, in turn, fills the LV, which delivers blood into the aorta (see Fig. 18.1). During ventricular contraction (systole), the tricuspid valve in the right heart and the mitral valve in the left heart close, and the pulmonary and aortic valves open. In diastole, the pulmonary and aortic valves close, and the two AV valves open. Collectively, these atrial and ventricular events constitute the cardiac cycle of filling and ejection of blood from one heartbeat to the next.

Myocardial contraction

Myocardial cells (myocytes) are about 50–100 µm long; each cell branches and interdigitates with adjacent cells. An intercalated disc permits electrical conduction via gap junctions, and mechanical conduction via the fascia adherens, to adjacent cells (Fig. 18.5A). The basic unit of contraction is the sarcomere (2 µm long), which is aligned to those of adjacent myofibrils, giving a striated appearance due to the Z-lines (Fig. 18.5B and C). Actin filaments are attached at right angles to the Z-lines and interdigitate with thicker parallel myosin filaments. The cross-links between actin and myosin molecules contain myofibrillar adenosine triphosphatase (ATPase), which breaks down adenosine triphosphate (ATP) to provide the energy for contraction (Fig. 18.5E). Two chains of actin molecules form a helical structure, with a second molecule, tropomyosin, in the grooves of the actin helix, and a further molecule complex, troponin, attached to every seventh actin molecule (Fig. 18.5D).

During the plateau phase of the action potential, calcium ions enter the cell and are mobilised from the sarcoplasmic reticulum. They bind to troponin and thereby precipitate contraction by shortening of the sarcomere through the interdigitation of the actin and myosin molecules. The force of cardiac muscle contraction, or inotropic state, is regulated by the influx of calcium ions through ‘slow calcium channels’. The extent to which the sarcomere can shorten determines stroke volume of the ventricle. It is maximally shortened in response to powerful inotropic drugs or marked exercise. However, the enlargement of the heart seen in heart failure is due to slippage of the myofilaments and adjacent cells rather than shortening of the sarcomere.

Cardiac output

Cardiac output is the product of stroke volume and heart rate. Stroke volume is the volume of blood ejected in each cardiac cycle (see Fig. 18.36, p. 561), and is dependent upon end-diastolic volume and pressure (preload), myocardial contractility and systolic aortic pressure (afterload). Stretch of cardiac muscle (from increased end-diastolic volume) causes an increase in the force of contraction, producing a greater stroke volume: Starling’s Law of the heart (see Fig. 18.22, p. 547).

The contractile state of the myocardium is controlled by neuro-endocrine factors, such as adrenaline (epinephrine), and can be influenced by inotropic drugs and their antagonists. The response to a physiological change or to a drug can be predicted on the basis of its combined influence on preload, afterload and contractility (see Fig. 18.26, p. 551).

Blood flow

Blood passes from the heart through the large central elastic arteries into muscular arteries before encountering the resistance vessels, and ultimately the capillary bed, where there is exchange of nutrients, oxygen and waste products of metabolism. The central arteries, such as the aorta, are predominantly composed of elastic tissue with little or no vascular smooth muscle cells. When blood is ejected from the heart, the compliant aorta expands to accommodate the volume of blood before the elastic recoil sustains blood pressure (BP) and flow following cessation of cardiac contraction. This ‘Windkessel effect’ prevents excessive rises in systolic BP whilst sustaining diastolic BP, thereby reducing cardiac afterload and maintaining coronary perfusion. These benefits are lost with progressive arterial stiffening: a feature of ageing and advanced renal disease.

Passing down the arterial tree, vascular smooth muscle cells progressively play a greater role until the resistance arterioles are encountered. Although all vessels contribute, the resistance vessels (diameter 50–200 µm) provide the greatest contribution to systemic vascular resistance, with small changes in radius having a marked influence on blood flow; resistance is proportional to the fourth power of the radius (Poiseuille’s Law). The tone of these resistance vessels is tightly regulated by humoral, neuronal and mechanical factors. Neurogenic constriction operates via α-adrenoceptors on vascular smooth muscle, and dilatation
vasoactive mediators that cause vasodilatation, including nitric oxide, prostacyclin and endothelium-derived hyperpolarising factor, and vasoconstriction, including endothelin-1 and angiotensin II. A balance exists whereby the release of such factors contributes to the maintenance and regulation of vascular tone and BP. Damage to the endothelium may disrupt this balance and lead to vascular dysfunction, tissue ischaemia and hypertension.

The endothelium also has a major influence on key regulatory steps in the recruitment of inflammatory cells and on the formation and dissolution of thrombus. Once activated, the endothelium expresses surface receptors such as E-selectin, intercellular adhesion molecule type 1 (ICAM-1) and platelet endothelial cell adhesion molecule type 1 (PECAM-1), which mediate rolling, adhesion and migration of inflammatory leucocytes into the subintima. The endothelium also stores and releases the multimeric glycoprotein, von Willebrand factor, which promotes thrombus formation by linking platelet adhesion to denuded surfaces, especially in the arterial vasculature. In contrast, once intravascular thrombus forms, tissue plasminogen activator is rapidly released from a dynamic storage pool within the endothelium to induce

Endothelial function

The endothelium plays a vital role in the control of vascular homeostasis. It synthesises and releases many
CARDIOVASCULAR DISEASE

fibrinolysis and thrombus dissolution. These processes are critically involved in the development and progression of atherosclerosis, and endothelial function and injury are seen as central to the pathogenesis of many cardiovascular disease states.

Effects of respiration
There is a fall in intrathoracic pressure during inspiration that tends to promote venous flow into the chest, producing an increase in the flow of blood through the right heart. However, a substantial volume of blood is sequestered in the chest as the lungs expand; the increase in the capacitance of the pulmonary vascular bed usually exceeds any increase in the output of the right heart and therefore there is a reduction in the flow of blood into the left heart during inspiration. In contrast, expiration is accompanied by a fall in venous return to the right heart, a reduction in the output of the right heart, a rise in the venous return to the left heart (as blood is squeezed out of the lungs) and an increase in the output of the left heart (Box 18.1).

Fig. 18.6 The electrocardiograph. The components correspond to depolarisation and repolarisation, as depicted in Figure 18.4. The upper limit of the normal range for each interval is given in brackets.

Pulsus paradoxus
This term is used to describe the exaggerated fall in BP during inspiration that is characteristic of cardiac tamponade (pp. 545 and 640) and severe airways obstruction. In airways obstruction, it is due to accentuation of the change in intrathoracic pressure with respiration. In cardiac tamponade, compression of the right heart prevents the normal increase in flow through the right heart on inspiration, which exaggerates the usual drop in venous return to the left heart and produces a marked fall in BP (>10 mmHg fall during inspiration).

INVESTIGATION OF CARDIOVASCULAR DISEASE

Specific investigations may be required to confirm a diagnosis of cardiac disease. Basic tests, such as electrocardiography, chest X-ray and echocardiography, can be performed in an outpatient clinic or at the bedside. Procedures such as cardiac catheterisation, radionuclide imaging, computed tomography (CT) and magnetic resonance imaging (MRI) require specialised facilities.

Electrocardiogram
The electrocardiogram (ECG) is used to assess cardiac rhythm and conduction. It provides information about chamber size and is the main test used to assess for myocardial ischaemia and infarction.

The basis of an ECG recording is that the electrical depolarisation of myocardial tissue produces a small dipole current which can be detected by electrode pairs on the body surface. These signals are amplified and either printed or displayed on a monitor (Fig. 18.6). During sinus rhythm, the SA node triggers atrial depolarisation, producing a P wave. Depolarisation proceeds slowly through the AV node, which is too small to produce a depolarisation wave detectable from the body surface. The bundle of His, bundle branches and Purkinje system are then activated, initiating ventricular myocardial depolarisation, which produces the QRS complex. The muscle mass of the ventricles is much larger than that of the atria, so the QRS complex is larger than the P wave. The interval between the onset of the P wave and the onset of the QRS complex is termed the ‘PR interval’ and largely reflects the duration of AV nodal conduction. Injury to the left or right bundle branch delays ventricular depolarisation, widening the QRS complex. Selective injury of one of the left fascicles (hemiblock, p. 573) affects the electrical axis. Repolarisation is slower and spreads from the epicardium to the endocardium. Atrial repolarisation does not cause a detectable signal but ventricular repolarisation produces the T wave. The QT interval represents the total duration of ventricular depolarisation and repolarisation.

The standard 12-lead ECG
The 12-lead ECG (Box 18.2) is generated from ten physical electrodes that are attached to the skin. One electrode is attached to each limb and six electrodes are attached to the chest. In addition, the left arm, right arm and left leg electrodes are attached to a central terminal acting as an additional virtual electrode in the centre of the chest (the right leg electrode acts as an earthing electrode). The twelve ‘leads’ of the ECG refer to recordings made from pairs or sets of these electrodes. They comprise three groups: three dipole limb leads, three augmented voltage limb leads and six unipole chest leads.

Leads I, II and III are the dipole limb leads and refer to recordings obtained from pairs of limb electrodes.
### 10.2 How to read a 12-lead ECG: examination sequence

<table>
<thead>
<tr>
<th>Rhythm strip (lead II)</th>
<th>To determine heart rate and rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac axis</td>
<td>Normal if QRS complexes +ve in leads I/II</td>
</tr>
<tr>
<td>P-wave shape</td>
<td>Tall P waves denote right atrial enlargement (P pulmonale) and notched P waves denote left atrial enlargement (P mitrale)</td>
</tr>
<tr>
<td>PR interval</td>
<td>Normal = 0.12–0.20 secs. Prolongation denotes impaired AV nodal conduction. A short PR interval occurs in Wolff–Parkinson–White syndrome (p. 568)</td>
</tr>
<tr>
<td>QRS duration</td>
<td>If &gt; 0.12 secs, ventricular conduction is abnormal (left or right bundle branch block)</td>
</tr>
<tr>
<td>QRS amplitude</td>
<td>Large QRS complexes occur in slim young patients and in patients with left ventricular hypertrophy</td>
</tr>
<tr>
<td>Q waves</td>
<td>May signify previous myocardial infarction</td>
</tr>
<tr>
<td>ST segment</td>
<td>ST elevation may signify myocardial infarction, pericarditis or left ventricular aneurysm; ST depression may signify ischaemia or infarction</td>
</tr>
<tr>
<td>T waves</td>
<td>T-wave inversion has many causes, including myocardial ischaemia or infarction, and electrolyte disturbances</td>
</tr>
<tr>
<td>QT interval</td>
<td>Normal &lt; 0.42 secs. QT prolongation may occur with congenital long QT syndrome, low K(^+), Mg(^+) or Ca(^+), and some drugs (see Box 18.34, p. 571)</td>
</tr>
<tr>
<td>ECG conventions</td>
<td>Depolarisation towards electrode: positive deflection Depolarisation away from electrode: negative deflection Sensitivity: 10 mm = 1 mV Paper speed: 25 mm per second Each large (5 mm) square = 0.2 s Each small (1 mm) square = 0.04 s Heart rate = 1500/RR interval (mm) (i.e. 300 + number of large squares between beats)</td>
</tr>
</tbody>
</table>

---

Lead I records the signal between the right (negative) and left (positive) arms. Lead II records the signal between the right arm (negative) and left leg (positive). Lead III records the signal between the left arm (negative) and left (positive) arm. These three leads thus record electrical activity along three different axes in the frontal plane. Leads aVR, aVL and aVF are the augmented voltage limb leads. These record electrical activity between a limb electrode and a modified central terminal. For example, lead aVL records the signal between the left arm (positive) and a central (negative) terminal, formed by connecting the right arm and left leg electrodes (Fig. 18.7). Similarly augmented signals are obtained from the right arm (aVR) and left leg (aVF). These leads also record electrical activity in the frontal plane, with each lead 120° apart. Lead aVF thus examines activity along the axis +90°, and lead aVL along the axis −30°, and so on.

When depolarisation moves towards a positive electrode, it produces a positive deflection in the ECG; depolarisation in the opposite direction produces a negative deflection. The average vector of ventricular depolarisation is known as the frontal cardiac axis. When the vector is at right angles to a lead, the depolarisation in that lead is equally negative and positive (isoelectric). In Figure 18.7A, the QRS complex is isoelectric in aVL, negative in aVR and most strongly positive in lead II; the main vector or axis of depolarisation is therefore 60°. The normal cardiac axis lies between −30° and +90°. Examples of left and right axis deviation are shown in Figures 18.7B and C.

There are six chest leads, V\(_1\)–V\(_6\), derived from electrodes placed on the anterior and lateral left side of the chest, over the heart. Each lead records the signal between the corresponding chest electrode (positive) and the central terminal (negative). Leads V\(_1\) and V\(_2\) lie approximately over the RV, V\(_3\) and V\(_4\) over the...
interventricular septum, and $V_5$ and $V_6$ over the LV (Fig. 18.8). The LV has the greater muscle mass and contributes the major component of the QRS complex.

The shape of the QRS complex varies across the chest leads. Depolarisation of the interventricular septum occurs first and moves from left to right; this generates a small initial negative deflection in lead $V_6$ (Q wave) and an initial positive deflection in lead $V_1$ (R wave). The second phase of depolarisation is activation of the body of the LV, which creates a large positive deflection or R wave in $V_4$ (with reciprocal changes in $V_3$). The third and final phase involves the RV and produces a small negative deflection or S wave in $V_6$.

**The ECG in ischaemia and infarction**

When an area of the myocardium is ischaemic or undergoing infarction, repolarisation and depolarisation become abnormal relative to the surrounding myocardium. In transmural infarction, there is initial ST segment elevation (the current of injury) in the leads facing or overlying the infarct; Q waves (negative deflections) will then appear as the entire thickness of the myocardial wall becomes electrically neutral relative to the adjacent myocardium. The changes occurring in infarction are described in more detail on page 589, and shown in Figures 18.71–18.74 (pp. 592–593). In myocardial ischaemia, the ECG typically shows ST segment depression and/or T-wave inversion; it is usually the subendocardium that most readily becomes ischaemic. Other conditions, such as left ventricular hypertrophy and electrolyte disturbances, can cause similar ST and T-wave changes.

**Exercise (stress) ECG**

Exercise electrocardiography is used to detect myocardial ischaemia during physical stress and is helpful in the diagnosis of coronary artery disease. A 12-lead ECG is recorded during exercise on a treadmill or bicycle ergometer. The limb electrodes are placed on the shoulders and hips rather than the wrists and ankles. The Bruce Protocol is the most commonly used for testing. BP is recorded and symptoms assessed throughout the test. Common indications for exercise testing are shown in Box 18.3. A test is ‘positive’ if anginal pain occurs, BP falls or fails to increase, or if there are ST segment shifts of more than 1 mm (see Fig. 18.64, p. 584). Exercise testing is useful in confirming the diagnosis in patients with suspected angina, and in such patients has good sensitivity and specificity (see Box 18.5). False-negative results can occur in patients with coronary artery disease, and some patients with a positive test will not have coronary disease (false-positive). It is an unreliable population screening tool because, in low-risk individuals (e.g. asymptomatic young or middle-aged women), an abnormal response is more likely to represent a false-positive than a true positive test.

Stress testing is contraindicated in the presence of acute coronary syndrome, decompensated heart failure and severe hypertension.

**Ambulatory ECG**

Continuous (ambulatory) ECG recordings can be obtained using a portable digital recorder. These devices usually provide limb lead ECG recordings only, and can record for between 1 and 7 days. Ambulatory ECG recording is principally used in the investigation of patients with suspected arrhythmia, such as those with intermittent palpitation, dizziness or syncope. For these patients, a 12-lead ECG provides only a snapshot of the cardiac rhythm and is unlikely to detect an intermittent arrhythmia, so a longer period of recording is useful (see Fig. 18.39, p. 563). These devices can also be used to assess rate control in patients with atrial fibrillation, and are sometimes used to detect transient myocardial ischaemia using ST segment analysis. For patients with more infrequent symptoms, small, patient-activated ECG recorders can be issued for several weeks until a symptom episode occurs. The patient places the device on the chest to record the rhythm during the episode. With some devices, the recording can be transmitted to hospital via telephone. Implantable ‘loop recorders’ resemble a leadless pacemaker and are implanted subcutaneously. They have a lifespan of 1–3 years and are used to investigate patients with infrequent but potentially serious symptoms, such as syncope.
Cardiac biomarkers

Plasma or serum biomarkers can be measured to assess myocardial dysfunction and ischaemia.

Brain natriuretic peptide

This is a 32-amino acid peptide and is secreted by the LV along with an inactive 76-amino acid N-terminal fragment (NT-proBNP). The latter is diagnostically more useful, as it has a longer half-life. It is elevated principally in conditions associated with left ventricular systolic dysfunction, and may aid the diagnosis and assess prognosis and response to therapy in patients with heart failure (p. 546).

Cardiac troponins

Troponin I and troponin T are structural cardiac muscle proteins (see Fig. 18.5, p. 531) that are released during myocyte damage and necrosis, and represent the cornerstone of the diagnosis of acute myocardial infarction (MI, p. 593). However, modern assays are extremely sensitive and some have a normal reference range and can detect very low levels of myocardial damage, so that elevated plasma troponin concentrations are seen in other acute conditions, such as pulmonary embolus, septic shock and acute pulmonary oedema. The diagnosis of MI therefore relies on the patient’s clinical presentation (see Box 18.61, p. 590).

Chest X-ray

This is useful for determining the size and shape of the heart, and the state of the pulmonary blood vessels and lung fields. Most information is given by a postero-anterior (PA) projection taken in full inspiration. Antero-posterior (AP) projections are convenient when patient movement is restricted but result in magnification of the cardiac shadow.

An estimate of overall heart size can be made by comparing the maximum width of the cardiac outline with the maximum internal transverse diameter of the thoracic cavity. ‘Cardiomegaly’ is the term used to describe an enlarged cardiac silhouette where the ‘cardiothoracic ratio’ is greater than 0.5. It can be caused by chamber dilatation, especially left ventricular dilatation, or by a pericardial effusion. Artefactual cardiomegaly may be due to a mediastinal mass or pectus excavatum (p. 731), and cannot be reliably assessed from an AP film. Cardiomegaly is not a sensitive indicator of left ventricular systolic dysfunction since the cardiothoracic ratio is normal in many affected patients (false-negative) and also lacks specificity with many patients with apparent cardiomegaly having normal echocardiograms (false-positive).

Dilatation of individual cardiac chambers can be recognised by the characteristic alterations to the cardiac silhouette:

- Left atrial dilatation results in prominence of the left atrial appendage, creating the appearance of a straight left heart border, a double cardiac shadow to the right of the sternum, and widening of the angle of the carina (bifurcation of the trachea) as the left main bronchus is pushed upwards (Fig. 18.9).
- Right atrial enlargement projects from the right heart border towards the right lower lung field.
- Left ventricular dilatation causes prominence of the left heart border and enlargement of the cardiac silhouette. Left ventricular hypertrophy produces rounding of the left heart border (Fig. 18.10).
- Right ventricular dilatation increases heart size, displaces the apex upwards and straightens the left heart border.

Lateral or oblique projections may be useful for detecting pericardial calcification in patients with constrictive pericarditis (p. 641) or a calcified thoracic aortic aneurysm, as these abnormalities may be obscured by the spine on the PA view.

The lung fields on the chest X-ray may show congestion and oedema in patients with heart failure (see Fig. 18.25, p. 550), and an increase in pulmonary blood flow.
CARDIOVASCULAR DISEASE

(pulmonary plethora’) in those with left-to-right shunt. Pleural effusions may also occur in heart failure.

Echocardiography (echo)

**Two-dimensional echocardiography**

Echocardiography, or cardiac ultrasound, is obtained by placing an ultrasound transducer on the chest wall to image the heart structures as a real-time, two-dimensional ‘slice’. This permits the rapid assessment of cardiac structure and function. Left ventricular wall thickness and ejection fraction can be estimated. Common indications for echocardiography are shown in Box 18.4.

**Doppler echocardiography**

This depends on the Doppler principle that sound waves reflected from moving objects, such as intracardiac red blood cells, undergo a frequency shift. The speed and direction of the red cells, and thus of blood, can be detected in the heart chambers and great vessels. The greater the frequency shift, the faster the blood is moving. The derived information can be presented either as a plot of blood velocity against time for a particular point in the heart (Fig. 18.11) or as a colour overlay on a two-dimensional real-time echo picture (colour-flow Doppler, Fig. 18.12). Doppler echocardiography can be used to detect valvular regurgitation, where the direction of

**Box 18.4 Common indications for echocardiography**

- Assessment of left ventricular function
- Diagnosis and quantification of severity of valve disease
- Identification of vegetations in endocarditis
- Identification of structural heart disease in atrial fibrillation, cardiomyopathies or congenital heart disease
- Detection of pericardial effusion
- Identification of structural heart disease or intracardiac thrombus in systemic embolism

**Fig. 18.11 Doppler echocardiography in aortic stenosis.**

A The aortic valve is imaged and a Doppler beam passed directly through the left ventricular outflow tract and the aorta into the turbulent flow beyond the stenosed valve. B The velocity of the blood cells is recorded to determine the maximum velocity and hence the pressure gradient across the valve. In this example, the peak velocity is approximately 450 cm/sec (4.5 m/sec), indicating severe aortic stenosis (peak gradient of 81 mmHg).

**Fig. 18.12 Echocardiographic illustration of the principal cardiac structures in the ‘four-chamber’ view.**

A The major chambers and valves. B Colour-flow Doppler has been used to demonstrate mitral regurgitation: a flame-shaped (yellow/blue) turbulent jet into the left atrium.
blood flow is reversed and turbulence is seen, and is also used to detect high pressure gradients associated with stenosed valves. For example, the normal resting systolic flow velocity across the aortic valve is approximately 1 m/sec; in the presence of aortic stenosis, this is increased as blood accelerates through the narrow orifice. In severe aortic stenosis, the peak aortic velocity may be increased to 5 m/sec (see Fig. 18.11). An estimate of the pressure gradient across a valve or lesion is given by the modified Bernoulli equation:

\[
\text{Pressure gradient (mmHg)} = 4 \times (\text{peak velocity in m/sec})^2
\]

Advanced techniques include three-dimensional echocardiography, intravascular ultrasound (defines vessel wall abnormalities and guides coronary intervention), intracardiac ultrasound (provides high-resolution images) and tissue Doppler imaging (quantifies myocardial contractility and diastolic function).

### Transoesophageal echocardiography

Transthoracic echocardiography sometimes produces poor images, especially if the patient is overweight or has obstructive airways disease. Some structures are difficult to visualise in transthoracic views, such as the left atrial appendage, pulmonary veins, thoracic aorta and interatrial septum. Transoesophageal echocardiography (TOE) uses an endoscope-like ultrasound probe which is passed into the oesophagus under light sedation and positioned behind the left atrium (LA). This produces high-resolution images, which makes the technique particularly valuable for investigating patients with prosthetic (especially mitral) valve dysfunction, congenital abnormalities (e.g. atrial septal defect), aortic dissection, infective endocarditis (vegetations that are too small to be detected by transthoracic echocardiography) and systemic embolism (intracardiac thrombus or masses).

### Stress echocardiography

Stress echocardiography is used to investigate patients with suspected coronary artery disease who are unsuitable for exercise stress testing, such as those with mobility problems or pre-existing bundle branch block. A two-dimensional echo is performed before and after infusion of a moderate to high dose of an inotrope, such as dobutamine. Myocardial segments with poor perfusion become ischaemic and contract poorly under stress, showing as a wall motion abnormality on the scan. Stress echocardiography is sometimes used to examine myocardial viability in patients with impaired left ventricular function. Low-dose dobutamine can induce contraction in ‘hibernating’ myocardium; such patients may benefit from bypass surgery or percutaneous coronary intervention.

### Computed tomographic imaging

Computed tomography (CT) is useful for imaging the cardiac chambers, great vessels, pericardium, and mediastinal structures and masses. Multidetector scanners can acquire up to 320 slices per rotation, allowing very high-resolution imaging. CT is often performed using a timed injection of X-ray contrast to produce clear images of blood vessels and associated pathologies. Contrast scans are very useful for imaging the aorta in suspected aortic dissection (see Fig. 18.82, p. 607), and the pulmonary arteries and branches in suspected pulmonary embolism (p. 721).

Some centres use cardiac CT scans for quantification of coronary artery calcification, which may serve as an index of cardiovascular risk. However, modern multidetector scanning allows non-invasive coronary angiography (Fig. 18.13) with a spatial resolution approaching that of conventional coronary arteriography and at a lower radiation dose. CT coronary angiography is particularly useful in the initial elective assessment of patients with chest pain and a low or intermediate likelihood of disease, since its negative predictive value is very high: that is, excluding the presence of coronary artery disease. Modern volume scanners are also able to assess myocardial perfusion, often at the same sitting.

### Magnetic resonance imaging

Magnetic resonance imaging (MRI) requires no ionising radiation and can be used to generate cross-sectional images of the heart, lungs and mediastinal structures. It provides better differentiation of soft tissue structures than CT but is poor at demonstrating calcification. MRI scans need to be ‘gated’ to the ECG, allowing the scanner to produce moving images of the heart and mediastinal structures throughout the cardiac cycle. MRI is very
useful for imaging the aorta, including suspected dissection (see Fig. 18.81, p. 606), and can define the anatomy of the heart and great vessels in patients with congenital heart disease. It is also useful for detecting infiltrative conditions affecting the heart.

Physiological data can be obtained from the signal returned from moving blood, which allows quantification of blood flow across regurgitant or stenotic valves. It is also possible to analyse regional wall motion in patients with suspected coronary disease or cardiomyopathy. The RV is difficult to assess using echocardiography because of its retrosternal position but is readily visualised with MRI.

MRI can also be employed to assess myocardial perfusion and viability. When a contrast agent, such as gadolinium, is injected, areas of myocardial hypoperfusion can be identified with better spatial resolution than nuclear medicine techniques. Later redistribution of this contrast, so-called delayed enhancement, can be used to identify myocardial scarring and fibrosis (Fig. 18.14). This can help in selecting patients for revascularisation procedures, or in identifying those with myocardial infiltration such as that seen with sarcoid heart disease and right ventricular dysplasia.

**Cardiac catheterisation**

This involves passage of a preshaped catheter via a vein or artery into the heart under X-ray guidance, which allows the measurement of pressure and oxygen saturation in the cardiac chambers and great vessels, and the performance of angiograms by injecting contrast media into a chamber or blood vessel.

Left heart catheterisation involves accessing the arterial circulation, usually via the radial artery, to allow catheterisation of the aorta, LV and coronary arteries. Coronary angiography is the most widely performed procedure, in which the left and right coronary arteries are selectively cannulated and imaged, providing information about the extent and severity of coronary stenoses, thrombus and calcification (Fig. 18.15). This permits

---

**Fig. 18.14** Cardiac magnetic resonance imaging. A Recent inferior myocardial infarction with black area of microvascular obstruction (arrow). B Old anterior myocardial infarction with large area of subendocardial delayed gadolinium enhancement (white area, arrows).

**Fig. 18.15** The left anterior descending and circumflex coronary arteries with a stenosis in the left anterior descending vessel. A Coronary artery angiogram. B Schematic of the vessels and branches.
planning of percutaneous coronary intervention and coronary artery bypass graft surgery. Left ventriculography can be performed during the procedure to determine the size and function of the LV and to demonstrate mitral regurgitation. Aortography defines the size of the aortic root and thoracic aorta, and can help quantify aortic regurgitation. Left heart catheterisation is a day-case procedure and is relatively safe, with serious complications occurring in fewer than 1 in 1000 cases.

Right heart catheterisation is used to assess right heart and pulmonary artery pressures, and to detect intracardiac shunts by measuring oxygen saturations in different chambers. For example, a step up in oxygen saturation from 65% in the RA to 80% in the pulmonary artery is indicative of a large left-to-right shunt that might be due to a ventricular septal defect. Cardiac output can also be measured using thermodilution techniques. Left atrial pressure can be measured directly by puncturing the interatrial septum from the RA with a special catheter. For most purposes, however, a satisfactory approximation to left atrial pressure can be obtained by ‘wedging’ an end-hole or balloon catheter in a branch of the pulmonary artery. Swan–Ganz balloon catheters are often used to monitor pulmonary ‘wedge’ pressure as a guide to left heart filling pressure in critically ill patients (p. 186).

Electrophysiology study

Patients with known or suspected arrhythmia are investigated by percutaneous placement of electrode catheters into the heart via the femoral and neck veins. Electrophysiology study (EPS) is most commonly performed to evaluate patients for catheter ablation, normally done during the same procedure. It is occasionally used for risk stratification of patients suspected of being at risk of ventricular arrhythmias.

Radionuclide imaging

The availability of gamma-emitting radionuclides with a short half-life has made it possible to study cardiac function non-invasively. Two techniques are available, although their use is declining due to the availability of equivalent or superior imaging techniques that have lower or no exposure to ionising radiation.

Blood pool imaging

The isotope is injected intravenously and mixes with the circulating blood. A gamma camera detects the amount of radiation-emitting blood in the heart at different phases of the cardiac cycle, thereby permitting the calculation of ventricular ejection fractions. It also allows the assessment of the size and ‘shape’ of the cardiac chambers.

Myocardial perfusion imaging

This technique involves obtaining scintiscans of the myocardium at rest and during stress after the administration of an intravenous radioactive isotope, such as 99mtechnetium tetrofosmin (see Fig. 18.65, p. 585). More sophisticated quantitative information is obtained with positron emission tomography (PET), which can also be used to assess myocardial metabolism, but this is only available in a few centres.

PRESENTING PROBLEMS IN CARDIOVASCULAR DISEASE

Cardiovascular disease gives rise to a relatively limited range of symptoms. Differential diagnosis depends on careful analysis of the factors that provoke symptoms, the subtle differences in how they are described by the patient, the clinical findings and appropriate investigations. A close relationship between symptoms and exercise is the hallmark of heart disease. The New York Heart Association (NYHA) functional classification is used to grade disability (Box 18.5).

Chest pain

Chest pain is a common presentation of cardiac disease but can also be a manifestation of anxiety or disease of the respiratory, musculoskeletal or gastrointestinal systems (see Box 18.6 below). Some patients deny ‘pain’ in favour of ‘discomfort’ but the significance remains the same.

Characteristics of cardiac pain

Several key characteristics help to distinguish cardiac pain from that of other causes (Fig. 18.16). Diagnosis may be difficult and it is helpful to classify pain as typical, atypical or non-cardiac chest pain, based on the balance of evidence (Fig. 18.17).

- **Site.** Cardiac pain is typically located in the centre of the chest because of the derivation of the nerve supply to the heart and mediastinum.

**Box 18.6 New York Heart Association (NYHA) functional classification**

- Class I: No limitation during ordinary activity
- Class II: Slight limitation during ordinary activity
- Class III: Marked limitation of normal activities without symptoms at rest
- Class IV: Unable to undertake physical activity without symptoms; symptoms may be present at rest

Fig. 18.16 Typical ischaemic cardiac pain. Characteristic hand gestures used to describe cardiac pain. Typical radiation of pain is shown in the schematic.
Cardiovascular disease

1. Radiation. Ischaemic cardiac pain may radiate to the neck, jaw, and upper or even lower arms. Occasionally, cardiac pain may be experienced only at the sites of radiation or in the back. Pain situated over the left anterior chest and radiating laterally is unlikely to be due to cardiac ischaemia and may have many causes, including pleural or lung disorders, musculoskeletal problems and anxiety.

2. Character. Cardiac pain is typically dull, constricting, choking or ‘heavy’, and is usually described as squeezing, crushing, burning or aching but not sharp, stabbing, prickling or knife-like. The sensation can be described as breathlessness. Patients often emphasise that it is a discomfort rather than a pain. They typically use characteristic hand gestures (e.g. open hand or clenched fist) when describing ischaemic pain (see Fig. 18.16).

3. Provocation. Anginal pain occurs during (not after) exertion and is promptly relieved (in less than 5 minutes) by rest. The pain may also be precipitated or exacerbated by emotion but tends to occur more readily during exertion, after a large meal or in a cold wind. In crescendo or unstable angina, similar pain may be precipitated by minimal exertion or at rest. The increase in venous return or preload induced by lying down may also be sufficient to provoke pain in vulnerable patients (decubitus angina). The pain of MI may be preceded by a period of stable or unstable angina but often occurs de novo. In contrast, pleural or pericardial pain is usually described as a ‘sharp’ or ‘catching’ sensation that is exacerbated by breathing, coughing or movement. Pain associated with a specific movement (bending, stretching, turning) is likely to be musculoskeletal in origin.

4. Onset. The pain of MI typically takes several minutes or even longer to develop; similarly, angina builds up gradually in proportion to the intensity of exertion. Pain that occurs after rather than during exertion is usually musculoskeletal or psychological in origin. The pain of aortic dissection, massive pulmonary embolism or pneumothorax is usually very sudden or instantaneous in onset.

5. Associated features. The pain of MI, massive pulmonary embolism or aortic dissection is often accompanied by autonomic disturbance, including sweating, nausea and vomiting. Breathlessness, due to pulmonary congestion arising from transient ischaemic left ventricular dysfunction, is often a prominent and occasionally the dominant feature of MI or angina (angina equivalent). Breathlessness may also accompany any of the respiratory causes of chest pain and can be associated with cough, wheeze or other respiratory symptoms. Gastrointestinal disorders, such as gastro-oesophageal reflux, peptic ulceration or biliary colic, may present with chest pain but effort-related ‘indigestion’ is usually due to heart disease.

**Differential diagnosis of chest pain**

Common causes of chest pain are listed in Box 18.6.

**Psychological aspects of chest pain**

Emotional distress is a common cause of atypical or non-cardiac chest pain. This diagnosis should be considered if there are features of anxiety and the pain lacks a predictable relationship with exercise. However, the prospect of heart disease is a frightening experience, particularly when it has been responsible for the death of a close friend or relative; psychological and organic features therefore often coexist. Anxiety may amplify the effects of organic disease and can create a very confusing picture. Patients who believe they are suffering from heart disease are sometimes afraid to take exercise and this may make it difficult to establish their true effort tolerance; assessment may also be complicated by the impact of physical deconditioning.
Presenting problems in cardiovascular disease

**Myocarditis and pericarditis**

Pain is characteristically felt retrosternally, to the left of the sternum, or in the left or right shoulder, and typically varies in intensity with movement and the phase of respiration. The pain is described as ‘sharp’ and may ‘catch’ the patient during inspiration, coughing or lying flat; there may be a history of a prodromal viral illness.

**Mitral valve prolapse**

Sharp left-sided chest pain that is suggestive of a musculoskeletal problem may be a feature of mitral valve prolapse (p. 618).

**Aortic dissection**

This pain is severe, sharp and tearing, is often felt in or penetrating through to the back, and is typically very abrupt in onset (p. 605). The pain follows the path of the dissection.

**Oesophageal pain**

This can mimic the pain of angina very closely, is sometimes precipitated by exercise and may be relieved by nitrates. However, it is usually possible to elicit a history relating chest pain to supine posture or eating, drinking or oesophageal reflux. It often radiates to the interscapular region and dysphagia may be present.

**Bronchospasm**

Patients with reversible airways obstruction, such as asthma, may describe exertional chest tightness that is relieved by rest. This may be difficult to distinguish from ischaemic chest tightness. Bronchospasm may be associated with wheeze, atopy and cough (p. 654).

**Musculoskeletal chest pain**

This is a common problem that is very variable in site and intensity but does not usually fall into any of the patterns described above. The pain may vary with posture or movement of the upper body and is sometimes accompanied by local tenderness over a rib or costal cartilage. There are numerous causes, including arthritis, costochondritis, intercostal muscle injury and Coxsackie viral infection (epidemic myalgia or Bornholm disease). Many minor soft tissue injuries are related to everyday activities, such as driving, manual work and sport. The differential diagnosis of peripheral or pleural chest pain is discussed on page 658.

**Initial evaluation of suspected cardiac pain**

A careful history is crucial in determining whether pain is cardiac or not. Although the physical findings and subsequent investigations may help to confirm the diagnosis, they are of more value in determining the nature and extent of any underlying heart disease, the risk of a serious adverse event, and the best course of management.

**Stable angina**

Effort-related chest pain is the hallmark of angina pectoris or ‘choking in the chest’ (Fig. 18.18). The reproducibility, predictability and relationship to physical exertion (and occasionally emotion) of the chest pain are

---

18.6 Common causes of chest pain

<table>
<thead>
<tr>
<th>Anxiety/emotion</th>
<th>Cardiac</th>
<th>Aortic</th>
<th>Oesophageal</th>
<th>Lungs/pleura</th>
<th>Musculoskeletal</th>
<th>Neurological</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Myocardial ischaemia (angina)</td>
<td>Mitral valve prolapse</td>
<td>Aortic dissection</td>
<td>Pulmonary embolism</td>
<td>Osteoarthritis</td>
<td>Prolapsed intervertebral disc</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td></td>
<td></td>
<td>Malignancy</td>
<td>Rib fracture/injury</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tuberculosis</td>
<td>Costochondritis (Tietze’s syndrome)</td>
<td>Herpes zoster</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Connective tissue disorders (rare)</td>
<td></td>
<td>Thoracic outlet syndrome</td>
</tr>
</tbody>
</table>

**Stable angina**

- Fixed stenosis
- Stable fibrous plaque
- Demand-led ischaemia
- Related to effort
- Predictable
- Symptoms over long term

**Acute coronary syndrome**

- Dynamic stenosis
- Ruptured or inflamed plaque
- Supply-led ischaemia
- Symptoms at rest
- Unpredictable
- Symptoms over short term
- Frequent or nocturnal symptoms

**Risk assessment**

- Symptoms on minimal exertion
- ECG changes at rest
- ECG changes with symptoms
- Elevation of troponin

**Pathophysiology**

- Exercise testing
  - Duration of exercise
  - Degree of ECG changes
  - Abnormal BP response
  - CT coronary angiogram

**Clinical features**

- Supply/led ischaemia
- Symptoms at rest
- Unpredictable
- Symptoms over short term
- Frequent or nocturnal symptoms

**Fig. 18.18** Pathophysiology, clinical features and risk assessment of patients with stable or unstable coronary heart disease.
CARDIOVASCULAR DISEASE

the most important features. The duration of symptoms should be noted because patients with recent-onset angina are at greater risk than those with long-standing and unchanged symptoms.

Physical examination is often normal but may reveal evidence of risk factors (e.g. xanthoma indicating hyperlipidaemia), left ventricular dysfunction (e.g. dyskinetic apex beat, gallop rhythm), other manifestations of arterial disease (e.g. bruits, signs of peripheral vascular disease) and unrelated conditions that may exacerbate angina (e.g. anaemia, thyroid disease). Stable angina is usually a symptom of coronary artery disease but may be a manifestation of other forms of heart disease, particularly aortic valve disease and hypertrophic cardiomyopathy. In patients with angina in whom a murmur is found, echocardiography should be performed.

A full blood count, fasting blood glucose, lipids, thyroid function tests and a 12-lead ECG are the most important baseline investigations. Exercise testing may confirm the diagnosis and also identify high-risk patients

---

**Fig. 18.19** Summary of treatment for acute coronary syndrome (ACS). *Not required following PCI. Amended from SIGN 93. For details of the GRACE score, see Figure 18.70, p. 591 (ACE = angiotensin-converting enzyme; GP = glycoprotein; LMW = low molecular weight; PCI = percutaneous coronary intervention). From SIGN 93 – see p. 641.
who require further investigation and treatment (p. 534). CT coronary angiography is very useful to exclude the presence of coronary artery disease where doubt exists.

**Acute coronary syndromes**

Prolonged, severe cardiac chest pain may be due to unstable angina (which comprises recent-onset limiting angina, rapidly worsening or crescendo angina, and angina at rest) or acute MI; these are known collectively as the acute coronary syndromes. Although there may be a history of antecedent chronic stable angina, an episode of chest pain at rest is often the first presentation of coronary artery disease. Diagnosis depends on analysis of the character of the pain and its associated features. Physical examination may reveal signs of important comorbidity, such as peripheral or cerebrovascular disease, autonomic disturbance (pallor or sweating) and complications (arrhythmia or heart failure).

Patients presenting with symptoms consistent with an acute coronary syndrome require urgent evaluation because there is a high risk of avoidable complications, such as sudden death and MI. Signs of haemodynamic compromise (hypotension, pulmonary oedema), ECG changes (ST segment elevation or depression) and biochemical markers of cardiac damage, such as elevated troponin I or T, are powerful indicators of short-term risk. A 12-lead ECG is mandatory and is the most useful method of initial triage (Fig. 18.19). The release of markers such as creatine kinase, troponin and myoglobin is relatively slow (p. 593) but can help guide immediate management and treatment.

If the diagnosis is unclear, patients with a suspected acute coronary syndrome should be observed in hospital. Repeated ECG recordings are valuable, particularly if obtained during an episode of pain. Plasma troponin concentrations should be measured at presentation and, if normal, repeated 6–12 hours after the onset of symptoms or hospital admission. New ECG changes or an elevated plasma troponin concentration confirm the diagnosis of an acute coronary syndrome. The subsequent management is described on page 593.

If the pain has not recurred, troponin concentrations are not elevated and there are no new ECG changes, the patient may be discharged from hospital. At this stage, an exercise test or CT coronary angiogram may help diagnose underlying coronary artery disease.

### Breathlessness (dyspnoea)

Dyspnoea of cardiac origin may vary in severity from an uncomfortable awareness of breathing to a frightening sensation of ‘fighting for breath’. The sensation of dyspnoea originates in the cerebral cortex and is described in detail on page 655.

There are several causes of cardiac dyspnoea: acute left heart failure, chronic heart failure, arrhythmia and angina equivalent (Box 18.7). The assessment and treatment of heart failure is described on pages 548–553, and arrhythmias on pages 562–571.

### Acute left heart failure

Acute left heart failure may be triggered by a major event, such as MI, in a previously healthy heart, or by a relatively minor event, such as the onset of atrial fibrillation, in a diseased heart. An increase in the left ventricular diastolic pressure causes the pressure in the LA, pulmonary veins and pulmonary capillaries to rise. When the hydrostatic pressure of the pulmonary capillaries exceeds the oncotic pressure of plasma (about 25–30 mmHg), fluid moves from the capillaries into the alveoli. This stimulates respiration through a series of autonomic reflexes, producing rapid shallow respiration. Congestion of the bronchial mucosa may cause wheeze (cardiac asthma).

Acute pulmonary oedema is a terrifying experience because of the sensation of ‘fighting for breath’. Sitting upright or standing may provide some relief by helping to reduce congestion at the apices of the lungs. The patient may be unable to speak and is typically distressed, agitated, sweaty and pale. Respiration is rapid, with recruitment of accessory muscles, coughing and wheezing. Sputum may be profuse, frothy and blood-streaked or pink. Extensive crepitations and rhonchi are usually audible in the chest and there may also be signs of right heart failure.

### Chronic heart failure

Chronic heart failure is the most common cardiac cause of chronic dyspnoea. Symptoms may first present on moderate exertion, such as walking up a steep hill, and may be described as a difficulty in ‘catching my breath’. As heart failure progresses, the dyspnoea is provoked by less exertion and, ultimately, the patient may be breathless walking from room to room, washing, dressing or trying to hold a conversation. Other symptoms may include:

- **Orthopnoea.** Lying down increases the venous return to the heart and provokes breathlessness. Patients may prop themselves up with pillows to prevent this.
**Arrhythmia**

Any arrhythmia may cause breathlessness but usually only does so if the heart is structurally abnormal, such as with the onset of atrial fibrillation in a patient with mitral stenosis.

**Angina equivalent**

Breathlessness is a common feature of angina. Patients will sometimes describe chest tightness as ‘breathlessness’. However, myocardial ischaemia may also induce true breathlessness by provoking transient left ventricular dysfunction or heart failure. When breathlessness is the dominant or sole feature of myocardial ischaemia, it is known as ‘angina equivalent’. A history of chest tightness, the close correlation with exercise, and objective evidence of myocardial ischaemia from stress testing may all help to establish the diagnosis.

**Acute circulatory failure (cardiogenic shock)**

‘Shock’ is used to describe the clinical syndrome that develops when there is critical impairment of tissue perfusion due to some form of acute circulatory failure. There are numerous causes of shock, described in detail on page 190. The important features and causes (Fig. 18.20) of acute heart failure or cardiogenic shock are described here.

---

**Fig. 18.20** Some common causes of cardiogenic shock.
Myocardial infarction

Shock in acute MI is due to left ventricular dysfunction in more than 70% of cases. However, it may also be due to infarction of the RV and a variety of mechanical complications, including tamponade (due to infarction and rupture of the free wall), an acquired ventricular septal defect (due to infarction and rupture of the septum) and acute mitral regurgitation (due to infarction or rupture of the papillary muscles).

Severe myocardial systolic dysfunction causes a fall in cardiac output, BP and coronary perfusion pressure. Diastolic dysfunction causes a rise in left ventricular end-diastolic pressure, pulmonary congestion and oedema, leading to hypoxaemia that worsens myocardial ischaemia. This is further exacerbated by peripheral vasoconstriction. These factors combine to create the ‘downward spiral’ of cardiogenic shock (Fig. 18.21).

Hypotension, oliguria, confusion and cold, clammy peripheries are the manifestations of a low cardiac output, whereas breathlessness, hypoxaemia, cyanosis and inspiratory crackles at the lung bases are typical features of pulmonary oedema. A chest X-ray (see Fig. 18.25, p. 550) may reveal signs of pulmonary congestion when clinical examination is normal. If necessary, a Swan–Ganz catheter can be used to measure the pulmonary artery wedge pressure and to guide fluid replacement. The findings can be used to categorise patients with acute MI into four haemodynamic subsets (Box 18.8). Those with cardiogenic shock should be considered for immediate coronary revascularisation.

The viable myocardium surrounding a fresh infarct may contract poorly for a few days and then recover. This phenomenon is known as myocardial stunning. Myocardial infarction may contract poorly for a few days and then recover.

Cardiac output | Pulmonary oedema
---|---
Normal | Good prognosis and requires no treatment for heart failure | Due to moderate left ventricular dysfunction. Treat with vasodilators and diuretics
Low | Due to right ventricular dysfunction or concomitant hypovolaemia. Give fluid challenge and consider pulmonary artery catheter to guide therapy | Extensive MI and poor prognosis. Consider intra-aortic balloon pump, vasodilators, diuretics and inotropes

Acute massive pulmonary embolism

This may complicate leg or pelvic vein thrombosis and usually presents with sudden collapse. The clinical features and treatment are discussed on page 721. Bedside echocardiography may demonstrate a small, under-filled, vigorous LV with a dilated RV; it is sometimes possible to see thrombus in the right ventricular outflow tract or main pulmonary artery. CT pulmonary angiography usually provides a definitive diagnosis.

Cardiac tamponade

This is due to a collection of fluid or blood in the pericardial sac, compressing the heart; the effusion may be small and is very occasionally less than 100 mL. Sudden deterioration (Box 18.9) may be due to bleeding into the pericardial space. Tamponade may complicate any form of pericarditis but can be caused by malignant disease. Other causes include trauma and rupture of the free wall of the myocardium following MI.

An ECG may show features of the underlying disease, such as pericarditis or acute MI. When there is a large pericardial effusion, the ECG complexes are small and there may be electrical alternans: a changing axis with alternate beats caused by the heart swinging from side to side in the pericardial fluid. A chest X-ray shows an enlarged globular heart but can look normal. Echocardiography is the best way of confirming the diagnosis.

**Clinical features of pericardial tamponade**

- Dyspnoea
- Collapse
- Tachycardia
- Hypotension
- Gross elevation of the JVP
- Soft heart sounds with an early third heart sound
- Pulsus paradoxus (a large fall in BP during inspiration, when the pulse may be impalpable)
- Kussmaul’s sign (a paradoxical rise in the JVP during inspiration)

(JVP = jugular venous pressure)
and helps to identify the optimum site for aspiration of the fluid. Prompt recognition of tamponade is important because the patient usually responds dramatically to percutaneous pericardiocentesis (p. 640) or surgical drainage.

**Valvular heart disease**

Acute left ventricular failure and shock may be due to the sudden onset of aortic regurgitation, mitral regurgitation or prosthetic valve dysfunction (Box 18.10).

The clinical diagnosis of acute valvular dysfunction is sometimes difficult. Murmurs are often unimpressive because there is usually a tachycardia and a low cardiac output. Transthoracic echocardiography will establish the diagnosis in most cases; however, transoesophageal echocardiography is sometimes required, especially in patients with prosthetic mitral valves.

Patients with acute valve failure usually require cardiac surgery and should be referred for urgent assessment in a cardiac centre.

Aortic dissection may lead to shock by causing aortic regurgitation, coronary dissection, tamponade or blood loss (p. 605).

**Management of shock**

This is discussed in detail on page 190.

### Heart failure

Heart failure describes the clinical syndrome that develops when the heart cannot maintain adequate output, or can do so only at the expense of elevated ventricular filling pressure. In mild to moderate forms of heart failure, cardiac output is normal at rest and only becomes impaired when the metabolic demand increases during exercise or some other form of stress. In practice, heart failure may be diagnosed when a patient with significant heart disease develops the signs or symptoms of a low cardiac output, pulmonary congestion or systemic venous congestion.

Almost all forms of heart disease can lead to heart failure. An accurate aetiological diagnosis (Box 18.11) is

---

<table>
<thead>
<tr>
<th>18.10 Causes of acute valve failure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aortic regurgitation</strong></td>
</tr>
<tr>
<td>• Aortic dissection</td>
</tr>
<tr>
<td>• Infective endocarditis</td>
</tr>
<tr>
<td><strong>Mitral regurgitation</strong></td>
</tr>
<tr>
<td>• Papillary muscle rupture due to acute MI</td>
</tr>
<tr>
<td>• Infective endocarditis</td>
</tr>
<tr>
<td>• Rupture of chordae due to myxomatous degeneration</td>
</tr>
<tr>
<td><strong>Prosthetic valve failure</strong></td>
</tr>
<tr>
<td>• Mechanical valves: fracture, jamming, thrombosis, dehiscence</td>
</tr>
<tr>
<td>• Biological valves: degeneration with cusp tear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>18.11 Mechanisms of heart failure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cause</strong></td>
</tr>
<tr>
<td><strong>Examples</strong></td>
</tr>
<tr>
<td><strong>Features</strong></td>
</tr>
<tr>
<td>Reduced ventricular contractility</td>
</tr>
<tr>
<td>MI (segmental dysfunction)</td>
</tr>
<tr>
<td>Myocarditis/cardiomyopathy (global dysfunction)</td>
</tr>
<tr>
<td>In coronary artery disease, ‘akinetic’ or ‘dyskinetic’ segments contract poorly and may impede the function of normal segments by distorting their contraction and relaxation patterns. Progressive ventricular dilatation</td>
</tr>
<tr>
<td>Ventricular outflow obstruction</td>
</tr>
<tr>
<td>Hypertension, aortic stenosis (left heart failure)</td>
</tr>
<tr>
<td>Pulmonary hypertension, pulmonary valve stenosis (right heart failure)</td>
</tr>
<tr>
<td>Initially, concentric ventricular hypertrophy allows the ventricle to maintain a normal output by generating a high systolic pressure. Later, secondary changes in the myocardium and increasing obstruction lead to failure with ventricular dilatation and rapid clinical deterioration</td>
</tr>
<tr>
<td>Ventricular inflow obstruction</td>
</tr>
<tr>
<td>Mitral stenosis, tricuspid stenosis</td>
</tr>
<tr>
<td>Small, vigorous ventricle, dilated hypertrophied atrium. Atrial fibrillation is common and often causes marked deterioration because ventricular filling depends heavily on atrial contraction</td>
</tr>
<tr>
<td>Ventricular volume overload</td>
</tr>
<tr>
<td>Ventricular septal defect</td>
</tr>
<tr>
<td>Right ventricular volume overload (e.g. atrial septal defect)</td>
</tr>
<tr>
<td>Increased metabolic demand (high output)</td>
</tr>
<tr>
<td>Dilatation and hypertrophy allow the ventricle to generate a high stroke volume and help to maintain a normal cardiac output. However, secondary changes in the myocardium lead to impaired contractility and worsening heart failure</td>
</tr>
<tr>
<td>Arrhythmia</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>Tachycardia cardiomyopathy</td>
</tr>
<tr>
<td>Complete heart block</td>
</tr>
<tr>
<td>Tachycardia does not allow for adequate filling of the heart, resulting in reduced cardiac output and back pressure. Incessant tachycardia causes myocardial fatigue. Bradycardia limits cardiac output, even if stroke volume is normal</td>
</tr>
<tr>
<td>Diastolic dysfunction</td>
</tr>
<tr>
<td>Constrictive pericarditis</td>
</tr>
<tr>
<td>Restrictive cardiomyopathy</td>
</tr>
<tr>
<td>Left ventricular hypertrophy and fibrosis</td>
</tr>
<tr>
<td>Cardiac tamponade</td>
</tr>
<tr>
<td>Marked fluid retention and peripheral oedema, ascites, pleural effusions and elevated jugular veins. Bi-atral enlargement (restrictive filling pattern and high atrial pressures). Atrial fibrillation may cause deterioration. Good systolic function but poor diastolic filling. Hypotension, elevated jugular veins, pulsus paradoxus, poor urine output</td>
</tr>
</tbody>
</table>
important because treatment of the underlying cause may reverse heart failure or prevent its progression.

Heart failure is most common in the elderly. The prevalence of heart failure rises from 1% in those aged 50–59 years to over 10% in those aged 80–89 years. In the UK, most patients admitted to hospital with heart failure are more than 70 years old; they remain hospitalised for a week or more and may be left with chronic disability. The most common aetiology is coronary artery disease and myocardial infarction.

Although the outlook depends, to some extent, on the underlying cause of the problem, untreated heart failure carries a poor prognosis; approximately 50% of patients with severe heart failure due to left ventricular dysfunction will die within 2 years, because of either pump failure or malignant ventricular arrhythmias.

**Pathophysiology**

Cardiac output is determined by preload (the volume and pressure of blood in the ventricles at the end of diastole), afterload (the volume and pressure of blood in the ventricles during systole) and myocardial contractility; this is the basis of Starling’s Law (Fig. 18.22).

In patients without valvular disease, the primary abnormality is impairment of ventricular myocardial function, leading to a fall in cardiac output. This can occur because of impaired systolic contraction, impaired diastolic relaxation, or both. This activates counter-regulatory neurohumoral mechanisms that, in normal physiological circumstances, would support cardiac function but, in the setting of impaired ventricular function, can lead to a deleterious increase in both afterload and preload (Fig. 18.23). A vicious circle may be established because any additional fall in cardiac output will cause further neurohumoral activation and increasing peripheral vascular resistance.

Stimulation of the renin–angiotensin–aldosterone system leads to vasoconstriction, sodium and water retention, and sympathetic nervous system activation. This is mediated by angiotensin II, a potent constrictor of arterioles, in both the kidney and the systemic circulation (see Fig. 18.23). Activation of the sympathetic nervous system may initially sustain cardiac output through increased myocardial contractility (inotropy) and heart rate (chronotropy). Prolonged sympathetic stimulation also causes negative effects, including cardiac myocyte apoptosis, hypertrophy and focal myocardial necrosis. Sympathetic stimulation also causes peripheral vasoconstriction and arrhythmias. Sodium and water retention is promoted by the release of aldosterone, endothelin-1 (a potent vasoconstrictor peptide with marked effects on the renal vasculature) and, in severe heart failure, antidiuretic hormone (ADH). Natriuretic peptides are released from the atria in response to atrial stretch, and act as physiological antagonists to the fluid-conserving effect of aldosterone.

After MI, cardiac contractility is impaired and neurohumoral activation causes hypertrophy of non-infarcted segments, with thinning, dilatation and expansion of the...
inflamed segment (remodelling; see Fig. 18.77, p. 597). This leads to further deterioration in ventricular function and worsening heart failure.

Pulmonary and peripheral oedema occurs because of high left and right atrial pressures, respectively; this is compounded by sodium and water retention, caused by impairment of renal perfusion and by secondary hyperaldosteronism.

Types of heart failure

**Left, right and biventricular heart failure**

The left side of the heart comprises the functional unit of the LA and LV, together with the mitral and aortic valves; the right heart comprises the RA, RV, and tricuspid and pulmonary valves.

- **Left-sided heart failure.** There is a reduction in left ventricular output and an increase in left atrial and pulmonary venous pressure. An acute increase in left atrial pressure causes pulmonary congestion or pulmonary oedema; a more gradual increase in left atrial pressure, as occurs with mitral stenosis, leads to reflex pulmonary vasoconstriction, which protects the patient from pulmonary oedema. This increases pulmonary vascular resistance and causes pulmonary hypertension, which can, in turn, impair right ventricular function.

- **Right-sided heart failure.** There is a reduction in right ventricular output and an increase in right atrial and systemic venous pressure. Causes of isolated right heart failure include chronic lung disease (cor pulmonale), pulmonary embolism and pulmonary valvular stenosis.

- **Biventricular heart failure.** Failure of the left and right heart may develop because the disease process, such as dilated cardiomyopathy or ischaemic heart disease, affects both ventricles or because disease of the left heart leads to chronic elevation of the left atrial pressure, pulmonary hypertension and right heart failure.

**Diastolic and systolic dysfunction**

Heart failure may develop as a result of impaired myocardial contraction (systolic dysfunction) but can also be due to poor ventricular filling and high filling pressures stemming from abnormal ventricular relaxation (diastolic dysfunction). The latter is caused by a stiff, non-compliant ventricle and is commonly found in patients with left ventricular hypertrophy. Systolic and diastolic dysfunction often coexist, particularly in patients with coronary artery disease.

**High-output failure**

A large arteriovenous shunt, beri-beri (p. 128), severe anaemia or thyrotoxicosis can occasionally cause heart failure due to an excessively high cardiac output.

**Acute and chronic heart failure**

Heart failure may develop suddenly, as in MI, or gradually, as in progressive valvular heart disease. When there is gradual impairment of cardiac function, several compensatory changes may take place.

The term ‘compensated heart failure’ is sometimes used to describe the condition of those with impaired cardiac function, in whom adaptive changes have prevented the development of overt heart failure. A minor event, such as an intercurrent infection or development of atrial fibrillation, may precipitate overt or acute heart failure (Box 18.12). Acute left heart failure occurs, either de novo or as an acute decompensated episode, on a background of chronic heart failure: so-called acute-on-chronic heart failure.

**Clinical assessment**

**Acute left heart failure**

Acute de novo left ventricular failure presents with a sudden onset of dyspnoea at rest that rapidly progresses to acute respiratory distress, orthopnoea and prostration. The precipitant, such as acute MI, is often apparent from the history.

The patient appears agitated, pale and clammy. The peripheries are cool to the touch and the pulse is rapid. Inappropriate bradycardia or excessive tachycardia should be identified promptly, as this may be the precipitant for the acute episode of heart failure. The BP is usually high because of sympathetic nervous system activation, but may be normal or low if the patient is in cardiogenic shock.

The jugular venous pressure (JVP) is usually elevated, particularly with associated fluid overload or right heart failure. In acute de novo heart failure, there has been no time for ventricular dilatation and the apex is not displaced. A ‘gallop’ rhythm, with a third heart sound, is heard quite early in the development of acute left-sided heart failure. A new systolic murmur may signify acute mitral regurgitation or ventricular septal rupture. Auscultatory findings in pulmonary oedema are crepitations at the lung bases, or throughout the lungs if pulmonary oedema is severe. Expiratory wheeze often accompanies this.

Acute-on-chronic heart failure will have additional features of long-standing heart failure (see below). Potential precipitants, such as an upper respiratory tract infection or inappropriate cessation of diuretic medication, should be identified.

**Chronic heart failure**

Patients with chronic heart failure commonly follow a relapsing and remitting course, with periods of stability and episodes of decompensation, leading to worsening symptoms that may necessitate hospitalisation. The clinical picture depends on the nature of the underlying heart disease, the type of heart failure that it has evoked, and the neurohumoral changes that have developed (see Box 18.11 and Fig. 18.24).
Low cardiac output causes fatigue, listlessness and a poor effort tolerance; the peripheries are cold and the BP is low. To maintain perfusion of vital organs, blood flow is diverted away from skeletal muscle and this may contribute to fatigue and weakness. Poor renal perfusion leads to oliguria and uraemia.

Pulmonary oedema due to left heart failure presents as above and with inspiratory crepitations over the lung bases. In contrast, right heart failure produces a high JVP (jugular venous pressure) and is a feature of severe heart failure and is a poor prognostic sign. It may be caused by diuretic therapy, inappropriate water retention due to high ADH secretion, or failure of the cell membrane ion pump.

**Thromboembolism.** Deep vein thrombosis and pulmonary embolism may occur due to the effects of a low cardiac output and enforced immobility. Systemic emboli occur in patients with atrial fibrillation or flutter, or with intracardiac thrombus complicating conditions such as mitral stenosis, MI or left ventricular aneurysm.

**Atrial and ventricular arrhythmias** are very common and may be related to electrolyte changes (e.g. hypokalaemia, hypomagnesaemia), the underlying cardiac disease, and the pro-arrhythmic effects of sympathetic activation. Atrial fibrillation occurs in approximately 20% of patients with heart failure and causes further impairment of cardiac function. Sudden death occurs in up to 50% of patients with heart failure and is often due to a ventricular arrhythmia. Frequent ventricular ectopic beats and runs of non-sustained ventricular tachycardia are common findings in patients with heart failure and are associated with an adverse prognosis.

**Investigations**

Serum urea, creatinine and electrolytes, haemoglobin, thyroid function, ECG and chest X-ray may help to establish the nature and severity of the underlying heart disease and detect any complications. Brain natriuretic peptide (BNP) is elevated in heart failure and is a marker of risk; it is useful in the investigation of patients with breathlessness or peripheral oedema.

Echocardiography is very useful and should be considered in all patients with heart failure in order to:

- Determine the aetiology
- Detect hitherto unsuspected valvular heart disease, such as occult mitral stenosis, and other conditions that may be amenable to specific remedies
- Identify patients who will benefit from long-term drug therapy, e.g. ACE inhibitors (see below).

**Complications**

In advanced heart failure, the following may occur:

- **Renal failure** is caused by poor renal perfusion due to low cardiac output and may be exacerbated by diuretic therapy, angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers.

- **Cardiac failure:** right or combined left and right heart failure, pericardial constriction, cardiomyopathy
- **Chronic venous insufficiency:** varicose veins
- **Hypoalbuminaemia:** nephrotic syndrome, liver disease, protein-losing enteropathy; often widespread, can affect arms and face
- **Drugs:**
  - Sodium retention: fludrocortisone, NSAIDs
  - Increasing capillary permeability: nifedipine, amlodipine
  - Idiopathic: women > men
  - Chronic lymphatic obstruction

- **Hyponatraemia** may be the result of treatment with potassium-losing diuretics or hyperaldosteronism caused by activation of the renin–angiotensin system and impaired aldosterone metabolism due to hepatic congestion. Most of the body’s potassium is intracellular and there may be substantial depletion of potassium stores, even when the plasma concentration is in the reference range.

- **Hyperkalaemia** may be due to the effects of drugs which promote renal resorption of potassium, in particular the combination of ACE inhibitors (or angiotensin receptor blockers) and mineralocorticoid receptor antagonists. These effects are amplified if there is renal dysfunction due to low cardiac output or atherosclerotic renal vascular disease.

- **Hypokalaemia** may be caused by hepatic venous congestion and poor arterial perfusion, which frequently cause mild jaundice and abnormal liver function tests; reduced synthesis of clotting factors can make anticoagulant control difficult.

- **Thromboembolism.** Deep vein thrombosis and pulmonary embolism may occur due to the effects of a low cardiac output and enforced immobility. Systemic emboli occur in patients with atrial fibrillation or flutter, or with intracardiac thrombus complicating conditions such as mitral stenosis, MI or left ventricular aneurysm.

- **Atrial and ventricular arrhythmias** are very common and may be related to electrolyte changes (e.g. hypokalaemia, hypomagnesaemia), the underlying cardiac disease, and the pro-arrhythmic effects of sympathetic activation. Atrial fibrillation occurs in approximately 20% of patients with heart failure and causes further impairment of cardiac function. Sudden death occurs in up to 50% of patients with heart failure and is often due to a ventricular arrhythmia. Frequent ventricular ectopic beats and runs of non-sustained ventricular tachycardia are common findings in patients with heart failure and are associated with an adverse prognosis.

**18.13 Differential diagnosis of peripheral oedema**

- Cardiac failure: right or combined left and right heart failure, pericardial constriction, cardiomyopathy
- Chronic venous insufficiency: varicose veins
- Hypoalbuminaemia: nephrotic syndrome, liver disease, protein-losing enteropathy; often widespread, can affect arms and face
- Drugs:
  - Sodium retention: fludrocortisone, NSAIDs
  - Increasing capillary permeability: nifedipine, amlodipine
  - Idiopathic: women > men
  - Chronic lymphatic obstruction

**Fig. 18.24 Clinical features of left and right heart failure.**

(JVP = jugular venous pressure)
Cardiovascular disease

Management of acute pulmonary oedema

This is an acute medical emergency:
• Sit the patient up to reduce pulmonary congestion.
• Give oxygen (high-flow, high-concentration).
• Non-invasive positive pressure ventilation (continuous positive airways pressure (CPAP) of 5–10 mmHg) by a tight-fitting facemask results in a more rapid clinical improvement.
• Administer nitrates, such as IV glyceryl trinitrate (10–200 µg/min or buccal glyceryl trinitrate 2–5 mg, titrated upwards every 10 minutes), until clinical improvement occurs or systolic BP falls to less than 110 mmHg.
• Administer a loop diuretic, such as furosemide (50–100 mg IV).

The patient should initially be kept rested, with continuous monitoring of cardiac rhythm, BP and pulse oximetry. Intravenous opiates must be used sparingly in distressed patients, as they may cause respiratory depression and exacerbation of hypoxaemia and hypercapnia.

If these measures prove ineffective, inotropic agents may be required to augment cardiac output, particularly in hypotensive patients. Insertion of an intra-aortic balloon pump may be beneficial in patients with acute cardiogenic pulmonary oedema and shock.

Management of chronic heart failure

General measures

Education of patients and their relatives about the causes and treatment of heart failure can help adherence to a management plan (Box 18.14). Some patients may need to weigh themselves daily, as a measure of fluid load, and adjust their diuretic therapy accordingly. Treatment of the underlying cause of heart failure (e.g. coronary artery disease) is important to prevent its progression.

Drug therapy

Cardiac function can be improved by increasing contractility, optimising preload or decreasing afterload (see Fig. 18.23). Drugs that reduce preload are appropriate in patients with high end-diastolic filling pressures and evidence of pulmonary or systemic venous congestion. Those that reduce afterload or increase myocardial

Chest X-ray

High pulmonary venous pressure in left-sided heart failure first shows on the chest X-ray (Fig. 18.25) as an abnormal distension of the upper lobe pulmonary veins (with the patient in the erect position). The vascularity of the lung fields becomes more prominent, and the right and left pulmonary arteries dilate. Subsequently, interstitial oedema causes thickened interlobular septa and dilated lymphatics. These are evident as horizontal lines in the costophrenic angles (septal or ‘Kerley B’ lines). More advanced changes due to alveolar oedema cause a hazy opacification spreading from the hilar regions, and pleural effusions.

18.14 General measures for the management of heart failure

<table>
<thead>
<tr>
<th>Education</th>
<th>Diet</th>
<th>Alcohol</th>
<th>Smoking</th>
<th>Exercise</th>
<th>Vaccination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanation of nature of disease, treatment and self-help strategies</td>
<td>Good general nutrition and weight reduction for the obese</td>
<td>Moderation or elimination of alcohol consumption. Alcohol-induced cardiomyopathy requires abstinence</td>
<td>Cessation</td>
<td>Regular moderate aerobic exercise within limits of symptoms</td>
<td>Consider influenza and pneumococcal vaccination</td>
</tr>
</tbody>
</table>
contractility are more useful in patients with signs and symptoms of a low cardiac output.

**Diuretic therapy**

In heart failure, diuretics produce an increase in urinary sodium and water excretion, leading to reduction in blood and plasma volume (p. 434). Diuretic therapy reduces preload and improves pulmonary and systemic venous congestion. It may also reduce afterload and ventricular volume, leading to a fall in ventricular wall tension and increased cardiac efficiency.

Although a fall in preload (ventricular filling pressure) tends to reduce cardiac output, the ‘Starling curve’ in heart failure is flat, so there may be a substantial and beneficial fall in filling pressure with little change in cardiac output (see Figs 18.22 and 18.26). Nevertheless, excessive diuretic therapy may cause an undesirable fall in cardiac output, especially in patients with a marked diastolic component to their heart failure. This leads to hypotension, lethargy and renal failure.

In some patients with severe chronic heart failure, particularly if there is associated renal impairment, oedema may persist, despite oral loop diuretic therapy. In such patients, an intravenous infusion of furosemide (5–10 mg/hr) may initiate a diuresis. Combining a loop diuretic with a thiazide diuretic (e.g. bendroflumethiazide 5 mg daily) may prove effective, but this can cause an excessive diuresis.

---

18.15 Congestive cardiac failure in old age

- **Incidence**: rises with age and affects 5–10% of those in their eighties.
- **Common causes**: coronary artery disease, hypertension and calcific degenerative valvular disease.
- **Diastolic dysfunction**: often prominent, particularly in those with a history of hypertension.
- **ACE inhibitors**: improve symptoms and mortality but are more frequently associated with postural hypotension and renal impairment than in younger patients.
- **Loop diuretics**: usually required but may be poorly tolerated in those with urinary incontinence and men with prostate enlargement.

---

![Cardiac output or ventricular performance](image)

**Fig. 18.26** The effect of treatment on ventricular performance curves in heart failure. Diuretics and venodilators (A), angiotensin-converting enzyme (ACE) inhibitors and mixed vasodilators (B), and positive inotropic agents (C).

---

**Angiotensin-converting enzyme inhibition therapy**

Angiotensin-converting enzyme (ACE) inhibition therapy interrupts the vicious circle of neurohumoral activation that is characteristic of moderate and severe heart failure by preventing the conversion of angiotensin I to angiotensin II, thereby preventing peripheral vasoconstriction, activation of the sympathetic nervous system (Fig. 18.27), and salt and water retention due to aldosterone release. These drugs also prevent the undesirable activation of the renin–angiotensin system caused by diuretic therapy.

In moderate and severe heart failure, ACE inhibitors can produce a substantial improvement in effort tolerance and in mortality. They can also improve outcome and prevent the onset of overt heart failure in patients with poor residual left ventricular function following MI (Box 18.16).

ACE inhibitors can cause symptomatic hypotension and impairment of renal function, especially in patients with bilateral renal artery stenosis or those with pre-existing renal disease. An increase in serum potassium concentration may occur that can offset hypokalaemia associated with loop diuretic therapy. Short-acting ACE inhibitors can cause marked falls in BP, particularly in the elderly or when started in the presence of hypotension, hypovolaemia or hyponatraemia. In stable patients without hypotension (systolic BP over 100 mmHg), ACE inhibitors can usually be safely started in the community. However, in other patients, it is usually advisable to withhold diuretics for 24 hours before starting treatment with a small dose of a long-acting agent, preferably given at night (Box 18.17). Renal function and serum potassium must be monitored and should be checked 1–2 weeks after starting therapy.

---

**EBM 18.16 ACE inhibitors and treatment of chronic heart failure**

ACE inhibitors in chronic heart failure due to ventricular dysfunction reduce mortality and re-admission rates; average MNTs for 3 years to prevent 1 death = 26 and for the combined endpoint of death or re-admission = 19.’


For further information: [www.sign.ac.uk/guidelines/fulltext/95/contents.html](http://www.sign.ac.uk/guidelines/fulltext/95/contents.html)

---

**Angiotensin receptor blocker therapy**

Angiotensin receptor blockers (ARBs; see Box 18.17) act by blocking the action of angiotensin II on the heart, peripheral vasculature and kidney. In heart failure, they produce beneficial haemodynamic changes that are similar to the effects of ACE inhibitors (see Fig. 18.27) but are generally better tolerated. They have comparable effects on mortality and are a useful alternative for patients who cannot tolerate ACE inhibitors (Box 18.18).
Unfortunately, they share all the more serious adverse effects of ACE inhibitors, including renal dysfunction and hyperkalaemia. ARBs are normally used as an alternative to ACE inhibitors, but the two can be combined in patients with resistant or recurrent heart failure.

**Vasodilator therapy**

These drugs are valuable in chronic heart failure, when ACE inhibitor or ARB drugs are contraindicated (e.g. in severe renal failure). Venodilators, such as nitrates, reduce preload, and arterial dilators, such as hydralazine, reduce afterload (see Fig. 18.26). Their use is limited by pharmacological tolerance and hypotension.

**Fig. 18.27** Neurohumoral activation and sites of action of drugs used in the treatment of heart failure.

---

**Table 18.17 ACE inhibitor and angiotensin receptor blocker (ARB) dosages in heart failure**

<table>
<thead>
<tr>
<th>ACE inhibitors</th>
<th>Starting dose</th>
<th>Target dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enalapril</td>
<td>2.5 mg twice daily</td>
<td>10 mg twice daily</td>
</tr>
<tr>
<td>Lisinopril</td>
<td>2.5 mg daily</td>
<td>20 mg daily</td>
</tr>
<tr>
<td>Ramipril</td>
<td>1.25 mg daily</td>
<td>10 mg daily</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Angiotensin receptor blockers</th>
<th>Starting dose</th>
<th>Target dose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losartan</td>
<td>25 mg daily</td>
<td>100 mg daily</td>
</tr>
<tr>
<td>Candesartan</td>
<td>4 mg daily</td>
<td>32 mg daily</td>
</tr>
<tr>
<td>Valsartan</td>
<td>40 mg daily</td>
<td>160 mg daily</td>
</tr>
</tbody>
</table>

**Table 18.18 Angiotensin receptor blockers (ARBs) and chronic heart failure**

‘Compared with ACE inhibitors, ARBs are better tolerated and have similar efficacy in reducing cardiovascular events. ARBs reduce cardiovascular morbidity and mortality in patients with symptomatic heart failure who are intolerant of ACE inhibitors. NNT\(_5\) for 5 years to prevent 1 death or hospitalisation for heart failure = 8. The addition of an ARB to an ACE inhibitor produces further additional benefit. NNT\(_5\) for 5 years to prevent 1 death or hospitalisation for heart failure = 16.’


**Beta-adrenoceptor blocker therapy**

Beta-blockade helps to counteract the deleterious effects of enhanced sympathetic stimulation and reduces the risk of arrhythmias and sudden death. When initiated in standard doses, they may precipitate acute-on-chronic heart failure, but when given in small incremental doses (e.g. bisoprolol started at a dose of 1.25 mg daily, and increased gradually over a 12-week period to a target maintenance dose of 10 mg daily), they can increase ejection fraction, improve symptoms, reduce the frequency of hospitalisation and reduce mortality in patients with chronic heart failure (Box 18.19). Beta-blockers are more effective at reducing mortality than ACE inhibitors: relative risk reduction of 33% versus 20%, respectively.

**Ivabradine**

Ivabradine acts on the I\(_f\) inward current in the SA node, resulting in reduction of heart rate. It reduces hospital admission and mortality rates in patients with heart failure due to moderate or severe left ventricular systolic impairment. In trials, its effects were most marked in patients with a relatively high heart rate (over 77/min), so ivabradine is best suited to patients who cannot take β-blockers or in whom the heart rate remains high despite β-blockade. It is ineffective in patients in atrial fibrillation.
Digoxin
Digoxin (p. 576) can be used to provide rate control in patients with heart failure and atrial fibrillation. In patients with severe heart failure (NYHA class III–IV, see Box 18.5, p. 539), digoxin reduces the likelihood of hospitalisation for heart failure, although it has no effect on long-term survival.

Amiodarone
This is a potent anti-arrhythmic drug (p. 576) that has little negative inotropic effect and may be valuable in patients with poor left ventricular function. It is only effective in the treatment of symptomatic arrhythmias, and should not be used as a preventative agent in asymptomatic patients.

**Implantable cardiac defibrillators and resynchronisation therapy**
Patients with symptomatic ventricular arrhythmias and heart failure have a very poor prognosis. Irrespective of their response to anti-arrhythmic drug therapy, all should be considered for implantation of a cardiac defibrillator because it improves survival (p. 579). In patients with marked intraventricular conduction delay, prolonged depolarisation may lead to uncoordinated left ventricular contraction. When this is associated with severe symptomatic heart failure, cardiac resynchronisation therapy should be considered. Here, both the LV and RV are paced simultaneously (Fig. 18.28) to generate a more coordinated left ventricular contraction and improve cardiac output. This is associated with improved symptoms and survival.

**Coronary revascularisation**
Coronary artery bypass surgery or percutaneous coronary intervention may improve function in areas of the myocardium that are 'hibernating' because of inadequate blood supply, and can be used to treat carefully selected patients with heart failure and coronary artery disease. If necessary, ‘hibernating’ myocardium can be identified by stress echocardiography and specialised nuclear or MR imaging.

**Heart transplantation**
Cardiac transplantation is an established and successful treatment for patients with intractable heart failure. Coronary artery disease and dilated cardiomyopathy are the most common indications. The introduction of ciclosporin for immunosuppression (p. 96) has improved survival, which is around 80% at 1 year. The use of transplantation is limited by the efficacy of modern drug and device therapies, as well as the availability of donor hearts, so it is generally reserved for young patients with severe symptoms despite optimal therapy.

Conventional heart transplantation is contraindicated in patients with pulmonary vascular disease due to long-standing left heart failure, complex congenital heart disease (e.g. Eisenmenger’s syndrome) or primary pulmonary hypertension because the RV of the donor heart may fail in the face of high pulmonary vascular resistance. However, heart–lung transplantation can be successful in patients with Eisenmenger’s syndrome. Lung transplantation has been used for primary pulmonary hypertension.

Although cardiac transplantation usually produces a dramatic improvement in the recipient’s quality of life, serious complications may occur:
- **Rejection.** In spite of routine therapy with ciclosporin A, azathioprine and corticosteroids, episodes of rejection are common and may present with heart failure, arrhythmias or subtle ECG changes; cardiac biopsy is often used to confirm the diagnosis before starting treatment with high-dose corticosteroids.
- **Accelerated atherosclerosis.** Recurrent heart failure is often due to progressive atherosclerosis in the coronary arteries of the donor heart. This is not confined to patients who underwent transplantation for coronary artery disease and is probably a manifestation of chronic rejection. Angina is rare because the heart has been denervated.
- **Infection.** Opportunistic infection with organisms such as cytomegalovirus or *Aspergillus* remains a major cause of death in transplant recipients.

**Ventricular assist devices**
Because of the limited supply of donor organs, ventricular assist devices (VADs) have been employed as:
- a bridge to cardiac transplantation
- potential long-term therapy
- short-term restoration therapy following a potentially reversible insult, e.g. viral myocarditis.

VADs assist cardiac output by using a roller, centrifugal or pulsatile pump that, in some cases, is implantable and portable. They withdraw blood through cannulae inserted in the atria or ventricular apex and pump it into the pulmonary artery or aorta. They are designed not only to unload the ventricles but also to provide support to the pulmonary and systemic circulations. Their more widespread application is limited by high complication rates (haemorrhage, systemic embolism, infection, neurological and renal sequelae), although some improvements in survival and quality of life have been demonstrated in patients with severe heart failure.
The term ‘syncope’ refers to sudden loss of consciousness due to reduced cerebral perfusion. ‘Presyncope’ refers to lightheadedness in which the individual thinks he or she may black out. Syncope affects around 20% of the population at some time and accounts for more than 5% of hospital admissions. Dizziness and presyncope are very common in old age (p. 173). Symptoms are disabling, undermine confidence and independence, and can affect an individual’s ability to work or to drive. There are three principal mechanisms that underlie recurrent presyncope or syncope:

- cardiac syncope due to mechanical cardiac dysfunction or arrhythmia
- neurocardiogenic syncope, in which an abnormal autonomic reflex causes bradycardia and/or hypotension
- postural hypotension, in which physiological peripheral vasoconstriction on standing is impaired, lead to hypotension.

Loss of consciousness can also be caused by non-cardiac pathology, such as epilepsy, cerebrovascular ischaemia or hypoglycaemia (Fig. 18.29).

**Syncope and presyncope**

Differential diagnosis

History-taking, from the patient or a witness, is the key to establishing a diagnosis. Attention should be paid to potential triggers (e.g. medication, exertion, posture), the victim’s appearance (e.g. colour, seizure activity), the duration of the episode and the speed of recovery (Box 18.20). Cardiac syncope is usually sudden but can be associated with premonitory light-headedness, palpitation or chest discomfort. The blackout is usually brief and recovery rapid. Neurocardiogenic syncope will often be associated with a situational trigger, and the patient may experience flushing, nausea and malaise for several minutes afterwards. Patients with seizures do not exhibit pallor, may have abnormal movements, usually take more than 5 minutes to recover and are often confused. A history of rotational vertigo is suggestive of a labyrinthine or vestibular disorder (p. 1167). The pattern and description of the patient’s symptoms should indicate the probable mechanism and help to determine subsequent investigations (Fig. 18.30). Postural hypotension is normally obvious from the history, with presyncope or, less commonly, syncope, occurring within a few seconds of standing.

**Fig. 18.29** The differential diagnosis of syncope and presyncope.
Cardiac syncope

**Arrhythmia**

Lightheadedness may occur with many arrhythmias, but blackouts (Stokes–Adams attacks, p. 572) are usually due to profound bradycardia or malignant ventricular tachyarrhythmias. The 12-lead ECG may show evidence of conducting system disease (e.g. sinus bradycardia, atrioventricular block, bundle branch block or axis deviation), which would predispose a patient to bradycardia, but the key to establishing a diagnosis is to obtain an ECG recording while symptoms are present. Since minor rhythm disturbances are common, especially in old age, symptoms must occur at the same time as a recorded arrhythmia before a diagnosis can be made. Ambulatory ECG recordings are helpful only if symptoms occur several times per week. Patient-activated ECG recorders are useful for examining the rhythm in patients with recurrent dizziness but are not useful in assessing sudden blackouts. When these investigations fail to establish a cause in patients with presyncope or syncope, an implantable ECG recorder can be sited subcutaneously over the upper left chest. This device continuously records the cardiac rhythm and will activate automatically if extreme bradycardia or tachycardia occurs. The ECG memory can also be tagged by the patient, using a hand-held activator. Stored ECGs can be accessed by the implanting centre, using a telemetry device in a clinic, or using a home monitoring system via an online link.

**Structural heart disease**

Severe aortic stenosis and hypertrophic obstructive cardiomyopathy can lead to lightheadedness or syncope on exertion. This is caused by profound hypotension due to a fall in cardiac output, or failure to increase output during exertion, coupled with exercise-induced peripheral vasodilatation. Severe coronary artery disease can produce the same symptoms because of ischaemic left ventricular dysfunction. Exertional arrhythmias also occur in these patients.

**Neurocardiogenic syncope**

This encompasses a family of syndromes in which bradycardia and/or hypotension occur because of a series of abnormal autonomic reflexes. The two main conditions are hypersensitive carotid sinus syndrome and malignant vasovagal syncope.

**Situational syncope**

This is the collective name given to some variants of neurocardiogenic syncope that occur in the presence of identifiable triggers (e.g. cough syncope, micturition syncope).

**Vasovagal syncope**

This is normally triggered by a reduction in venous return due to prolonged standing, excessive heat or a large meal. It is mediated by the Bezold–Jarisch reflex, in which a combination of sympathetic activation, and...
reduced venous return due to an impaired vasoconstrictor response to standing, leads to vigorous contraction of relatively under-filled ventricles. This stimulates ventricular mechanoreceptors, producing parasympathetic (vagal) activation and sympathetic withdrawal, causing bradycardia, vasodilatation or both. Head-up tilt-table testing is a provocation test used to establish the diagnosis, and involves positioning the patient supine on a padded table that is then tilted to an angle of 60–70° for up to 45 minutes, while the ECG and BP responses are monitored. A positive test is characterised by bradycardia (cardio-inhibitory response) and/or hypotension (vasodepressor response) associated with typical symptoms. Initial management involves lifestyle modification (salt supplementation and avoiding prolonged standing, dehydration or missing meals). In resistant cases, drug therapy can be tried, although efficacy is inconsistent in clinical trials. Fludrocortisone (causes sodium and water retention and expands plasma volume), disopyramide (a vagolytic agent) or midodrine (a vasoconstrictor α-adrenoceptor agonist) may be helpful. Beta-blockers (inhibit the initial sympathetic activation) are effective and are rarely used. In patients with resistant vasovagal syncope in which bradycardia is the predominant response, a dual-chamber pacemaker can be useful. Patients with a urinary sodium excretion of less than 170 mmol/day may respond to salt loading.

Hypersensitive carotid sinus syndrome

Hypersensitive carotid sinus syndrome (HCSS) causes presyncope or syncope because of reflex bradycardia and vasodilatation. Carotid baroreceptors are involved in BP regulation and are activated by increased BP, resulting in a vagal discharge that causes a compensatory drop in BP. In HCSS, the baroreceptor is sensitive to external pressure (e.g. during neck movement or if a tight collar is worn), so that pressure over the carotid artery causes an inappropriate and intense vagal discharge. The diagnosis can be established by monitoring the ECG and BP during carotid sinus pressure. This manoeuvre should not be attempted in patients with a carotid bruit or with a history of cerebrovascular disease because of the risk of embolic stroke. A positive cardio-inhibitory response is defined as a sinus pause of 3 seconds or more; a positive vasodepressor response is defined as a fall in systolic BP of more than 50 mmHg. Carotid sinus pressure will produce positive findings in about 10% of elderly individuals but less than 25% of these experience spontaneous syncope. Symptoms should not therefore be attributed to HCSS unless they are reproduced by carotid sinus pressure. Dual-chamber pacemaker implantation usually prevents syncope in patients with the more common cardio-inhibitory response.

Postural hypotension

This is caused by a failure of normal postural compensatory mechanisms. Relative hypovolaemia (often due to excessive diuretic therapy), sympathetic degeneration (diabetes mellitus, Parkinson’s disease, ageing) and drug therapy (vasodilators, antidepressants) can all cause or aggravate the problem. Treatment is often ineffective; however, withdrawing unnecessary medication and advising the patient to wear graduated elastic stockings and to get up slowly may be helpful. Fludrocortisone, which can expand blood volume through sodium and water retention, may be of value.

**Palpitation**

Palpitation is a very common and sometimes frightening symptom. Patients use the term to describe many sensations, including an unusually erratic, fast, slow or forceful heart beat, or even chest pain or breathlessness. Initial evaluation should concentrate on determining its likely mechanism, and whether or not there is significant underlying heart disease.

A detailed description of the sensation is essential and patients should be asked to describe their symptoms clearly, or to demonstrate the sensation of rhythm by tapping with their hand. A provisional diagnosis can usually be made on the basis of a thorough history (Box 18.21 and Fig. 18.31). The diagnosis should be confirmed by an ECG recording during an episode using an ambulatory ECG monitor or a patient-activated ECG recorder.

---

**18.21 How to evaluate palpitation**

- Is the palpitation continuous or intermittent?
- Is the heart beat regular or irregular?
- What is the approximate heart rate?
- Do symptoms occur in discrete attacks?
- Is the onset abrupt? How do attacks terminate?
- Are there any associated symptoms?
  - e.g. Chest pain, lightheadedness, polyuria (a feature of supraventricular tachycardia, p. 567)
- Are there any precipitating factors, e.g. exercise, alcohol?
- Is there a history of structural heart disease, e.g. coronary artery disease, valvular heart disease?

---

**Fig. 18.31** A simple approach to the diagnosis of palpitation.
Recurrent but short-lived bouts of an irregular heart beat are usually due to atrial or ventricular extrasystoles (ectopic beats). Some patients will describe the experience as a ‘flip’ or a ‘jolt’ in the chest, while others report dropped or missed beats. Extrasystoles are often more frequent during periods of stress or debility; they can be triggered by alcohol or nicotine.

Episodes of a pounding, forceful and relatively fast (90–120 / min) heart beat are a common manifestation of anxiety. These may also reflect a hyperdynamic circulation, such as anaemia, pregnancy and thyrotoxicosis, and can occur in some forms of valve disease (e.g. aortic regurgitation). Discrete bouts of a very rapid (over 120/min) heart beat are more likely to be due to a paroxysmal tachyarrhythmia. Supraventricular and ventricular tachycardias may present in this way. In contrast, episodes of atrial fibrillation typically present with irregular and usually rapid palpitation.

Palpitation is usually benign and, even if the patient’s symptoms are due to an arrhythmia, the outlook is good if there is no underlying structural heart disease. Most cases are due to an awareness of the normal heart beat, a sinus tachycardia or benign extrasystoles, in which case an explanation and reassurance may be all that is required. Palpitation associated with presyncope or syncpe may reflect more serious structural or electrical disease and should be investigated without delay.

The diagnosis and management of individual arrhythmias are considered on pages 562–579.

Cardiac arrest and sudden cardiac death

Cardiac arrest describes the sudden and complete loss of cardiac output due to asystole, ventricular tachycardia or fibrillation, or loss of mechanical cardiac contraction (pulseless electrical activity). The clinical diagnosis is based on the victim being unconscious and pulseless; breathing may take some time to stop completely after cardiac arrest. Death is virtually inevitable, unless effective treatment is given promptly.

Sudden cardiac death is usually caused by a catastrophic arrhythmia and accounts for 25–30% of deaths from cardiovascular disease, claiming an estimated 70000–90000 lives each year in the UK. Many of these deaths are potentially preventable. Arrhythmias complicate many types of heart disease and can sometimes occur in the absence of recognisable structural abnormalities (causes are listed in Box 18.22). Sudden death less often occurs because of an acute mechanical catastrophe such as cardiac rupture or aortic dissection (pp. 597 and 605).

Coronary artery disease is the most common condition leading to cardiac arrest. Ventricular fibrillation or ventricular tachycardia is common in the first few hours of MI and many victims die before medical help is sought. Up to one-third of people developing MI die before reaching hospital, emphasising the importance of educating the public to recognise symptoms and to seek medical help quickly. Acute myocardial ischaemia (in the absence of infarction) can also cause these arrhythmias, although less commonly. Patients with a history of previous MI may be at risk of sudden arrhythmic death, especially if there is extensive left ventricular scarring and impairment, or if there is ongoing myocardial ischaemia. In these patients, the risk is reduced by the treatment of heart failure with β-blockers and ACE inhibitors, and by coronary revascularisation. For some patients, the risk of sudden death is reduced by the implantation of a cardiac defibrillator (p. 579).

Aetiology of cardiac arrest

Cardiac arrest may be caused by ventricular fibrillation, pulseless ventricular tachycardia, asystole or pulseless electrical activity.

Ventricular fibrillation and pulseless ventricular tachycardia

These are the most common and most easily treatable cardiac arrest rhythms. Ventricular fibrillation produces rapid, ineffective, uncoordinated movement of the ventricles, which therefore produces no pulse. The ECG (Fig. 18.32) shows rapid, bizarre and irregular ventricular complexes. Ventricular tachycardia (p. 569) can cause cardiac arrest if the ventricular rate is so rapid that effective mechanical contraction and relaxation cannot occur, especially in the presence of severe left ventricular impairment. It may degenerate into ventricular fibrillation. Defibrillation will restore cardiac output in more than 80% of patients, if delivered immediately. However, the chances of survival fall by at least 10% with each minute’s delay, and by more if basic life support is not given (see below); thus provision of these is the key to survival.

**Cardiac arrest and sudden cardiac death**

Cardiac arrest describes the sudden and complete loss of cardiac output due to asystole, ventricular tachycardia or fibrillation, or loss of mechanical cardiac contraction (pulseless electrical activity). The clinical diagnosis is based on the victim being unconscious and pulseless; breathing may take some time to stop completely after cardiac arrest. Death is virtually inevitable, unless effective treatment is given promptly.

**Fig. 18.32** Ventricular fibrillation. A bizarre chaotic rhythm, initiated in this case by two ventricular ectopic beats in rapid succession.
CARDIOVASCULAR DISEASE

Asystole
This occurs when there is no electrical activity within the ventricles and is usually due to failure of the conducting tissue or massive ventricular damage complicating MI. A precordial thump, external cardiac massage, or administration of intravenous atropine or adrenaline (epinephrine) may restore cardiac activity. When due to conducting tissue failure, permanent pacemaker implantation will be required if the individual survives.

Pulseless electrical activity
This occurs when there is no effective cardiac output despite the presence of organised electrical activity. It may be caused by reversible conditions, such as hypovolaemia, cardiac tamponade or tension pneumothorax (see Fig. 18.35 below), but is often due to a catastrophic event, such as cardiac rupture or massive pulmonary embolism, and therefore carries an extremely poor prognosis.

Management of cardiac arrest
The Chain of Survival
This term refers to the sequence of events that is necessary to maximise the chances of a cardiac arrest victim surviving (Fig. 18.33). Survival is most likely if all links in the chain are strong: that is, if the arrest is witnessed, help is called immediately, basic life support is administered by a trained individual, the emergency medical services respond promptly, and defibrillation is achieved within a few minutes. Good training in both basic and advanced life support is essential and should be maintained by regular refresher courses. In recent years, public access defibrillation has been introduced in places of high population density, particularly where traffic congestion may impede the response of emergency services, such as railway stations, airports and sports stadia. Designated individuals can respond to a cardiac arrest using basic life support and an automated external defibrillator.

Basic life support
Basic life support (BLS) encompasses manoeuvres that aim to maintain a low level of circulation until more definitive treatment with advanced life support can be given. The ABCDE approach to management of the collapsed patient should be followed: prompt assessment and restoration of the Airway, maintenance of ventilation using rescue Breathing (‘mouth-to-mouth’ breathing), maintenance of the Circulation using chest compressions; Disability, in resuscitated patients, refers to assessment of neurological status, and Exposure entails removal of clothes to enable defibrillation, auscultation of the chest, and assessment for a rash caused by anaphylaxis, injuries or so on (Fig. 18.34). Chest compression-only (‘hands-only’) CPR is easier for members of the public to learn and administer, and is now advocated in public education campaigns.

Advanced life support (ALS)
ALS (Fig. 18.35) aims to restore normal cardiac rhythm by defibrillation when the cause of cardiac arrest is due to a tachyarrhythmia, or to restore cardiac output by correcting other reversible causes of cardiac arrest. ALS can also involve administration of intravenous drugs to
support the circulation, and endotracheal intubation to ventilate the lungs.

If cardiac arrest is witnessed, a precordial thump may sometimes convert ventricular fibrillation or tachycardia to normal rhythm, but this is futile if cardiac arrest has lasted longer than a few seconds. The priority is to assess the patient’s cardiac rhythm by attaching a defibrillator or monitor. Ventricular fibrillation or pulseless ventricular tachycardia is treated with immediate defibrillation. Defibrillation is more likely to be effective if a biphasic shock defibrillator is used, where the polarity of the shock is reversed midway through its delivery. Defibrillation is usually administered using a 150-Joule biphasic shock, and CPR resumed immediately for 2 minutes without attempting to confirm restoration of a pulse, because restoration of mechanical cardiac output rarely occurs immediately after successful defibrillation. If, after 2 minutes, a pulse is not restored, a further biphasic shock of 150–200 joules is given. Thereafter, additional biphasic shocks of 150–200 joules are given every 2 minutes after each cycle of cardiopulmonary resuscitation (CPR). During resuscitation, adrenaline (epinephrine, 1 mg IV) should be given every 3–5 minutes and consideration given to the use of intravenous amiodarone, especially if ventricular fibrillation or ventricular tachycardia re-initiates after successful defibrillation.

Ventricular fibrillation of low amplitude, or ‘fine VF’, may mimic asystole. If asystole cannot be confidently diagnosed, the patient should be treated for VF and defibrillated. If an electrical rhythm is observed that would be expected to produce a cardiac output, ‘pulseless electrical activity’ is present. Pulseless electrical activity is treated by continuing CPR and adrenaline (epinephrine) administration whilst seeking such causes. Asystole is treated similarly, with the additional support of atropine and sometimes external or transvenous pacing in an attempt to generate an electrical rhythm.

There are many potentially reversible causes of cardiac arrest and the main causes can be easily remembered as a list of four Hs and four Ts (see Fig. 18.35).

**Survivors of cardiac arrest**

Patients who survive a cardiac arrest caused by acute MI need no specific treatment beyond that given to those recovering from an uncomplicated infarct, since their prognosis is similar (p. 599). Those with reversible causes, such as exercise-induced ischaemia or aortic stenosis, should have the underlying cause treated if possible. Survivors of ventricular tachycardia or ventricular fibrillation

---

Fig. 18.35 Algorithm for adult advanced life support. For further information see www.resus.org.uk (CPR = cardiopulmonary resuscitation; PEA = pulseless electrical activity; VF = ventricular fibrillation; VT = pulseless ventricular tachycardia). From Resuscitation Council (UK) guidelines – see p. 641.
fibrillation arrest in whom no reversible cause can be identified may be at risk of another episode, and should be considered for an implantable cardiac defibrillator (p. 579) and anti-arrhythmic drug therapy.

### Abnormal heart sounds and murmurs

The first indication of heart disease may be the discovery of an abnormal sound on auscultation (Box 18.23). This may be incidental – for example, during a routine childhood examination – or may be prompted by symptoms of heart disease. Clinical evaluation is helpful, and is supported by more detailed evaluation of the abnormal sound or murmur using echocardiography.

**Is the sound cardiac?**

Additional heart sounds and murmurs demonstrate a consistent relationship to a specific part of the cardiac cycle but extracardiac sounds (e.g. pleural rub or venous hum) do not. Pericardial friction produces a characteristic scratching noise (a pericardial ‘rub’), which may have two components corresponding to atrial and ventricular systole, and may vary with posture and respiration.

**Is the sound pathological?**

Pathological sounds and murmurs are the product of turbulent blood flow or rapid ventricular filling due to abnormal loading conditions. Some added sounds are physiological but may also occur in pathological conditions; for example, a third sound is common in young people and in pregnancy but is also a feature of heart failure (see Box 18.23). Similarly, a systolic murmur due to turbulence across the right ventricular outflow tract may occur in hyperdynamic states (e.g. anaemia, pregnancy), but may also be due to pulmonary stenosis or an intracardiac shunt leading to volume overload of the RV (e.g. atrial septal defect).

Benign murmurs do not occur in diastole (Box 18.24), and systolic murmurs that radiate or are associated with a thrill are almost always pathological.

### Auscultatory evaluation of a heart murmur

Timing, intensity, location, radiation and quality are all useful clues to the origin and nature of a heart murmur (Box 18.25). Radiation of a murmur is determined by the direction of turbulent blood flow and is only detectable when there is a high-velocity jet, such as in mitral regurgitation (radiation from apex to axilla) or aortic stenosis (radiation from base to neck). Similarly, the pitch and quality of the sound can help to distinguish the murmur,

---

**18.23 Normal and abnormal heart sounds**

<table>
<thead>
<tr>
<th>Sound</th>
<th>Timing</th>
<th>Characteristics</th>
<th>Mechanisms</th>
<th>Variable features</th>
</tr>
</thead>
<tbody>
<tr>
<td>First heart sound (S1)</td>
<td>Onset of systole</td>
<td>Usually single or narrowly split</td>
<td>Closure of mitral and tricuspid valves</td>
<td>Loud: hyperdynamic circulation (anaemia, pregnancy, thyrotoxicosis); mitral stenosis Soft: heart failure; mitral regurgitation</td>
</tr>
<tr>
<td>Second heart sound (S2)</td>
<td>End of systole</td>
<td>Split on inspiration Single on expiration (p. 532)</td>
<td>Closure of aortic and pulmonary valve A2: first P2: second</td>
<td>Fixed wide splitting with atrial septal defect Wide but variable splitting with delayed right heart emptying (e.g. right bundle branch block) Reversed splitting due to delayed left heart emptying (e.g. left bundle branch block)</td>
</tr>
<tr>
<td>Third heart sound (S3)</td>
<td>Early in diastole, just after S2</td>
<td>Low pitch, often heard as ‘gallop’</td>
<td>From ventricular wall due to abrupt cessation of rapid filling</td>
<td>Physiological: young people, pregnancy Pathological: heart failure, mitral regurgitation</td>
</tr>
<tr>
<td>Fourth heart sound (S4)</td>
<td>End of diastole, just before S1</td>
<td>Low pitch</td>
<td>Ventricular origin (stiff ventricle and augmented atrial contraction) related to atrial filling</td>
<td>Absent in atrial fibrillation A feature of severe left ventricular hypertrophy (e.g. hypertrophic cardiomyopathy)</td>
</tr>
<tr>
<td>Systolic clicks</td>
<td>Early or mid-systole</td>
<td>Brief, high-intensity sound</td>
<td>Valvular aortic stenosis Valvular pulmonary stenosis Floppy mitral valve Prosthetic heart sounds from opening and closing of normally functioning mechanical valves</td>
<td>Click may be lost when stenotic valve becomes thickened or calcified Prosthetic clicks lost when valve obstructed by thrombus or vegetations</td>
</tr>
<tr>
<td>Opening snap (OS)</td>
<td>Early in diastole</td>
<td>High pitch, brief duration</td>
<td>Opening of stenosed leaflets of mitral valve Prosthetic heart sounds</td>
<td>Moves closer to S2 as mitral stenosis becomes more severe. May be absent in calcific mitral stenosis</td>
</tr>
</tbody>
</table>

---

**18.24 Features of a benign or innocent heart murmur**

- Soft
- Mid-systolic
- Heard at left sternal edge
- No radiation
- No other cardiac abnormalities
such as the ‘blowing’ murmur of mitral regurgitation or the ‘rasping’ murmur of aortic stenosis.

The position of a murmur in relation to the cardiac cycle is crucial and should be assessed by timing it with the heart sounds, carotid pulse and apex beat (Figs 18.36 and 18.37).

**Systolic murmurs**

Ejection systolic murmurs are associated with ventricular outflow tract obstruction and occur in mid-systole with a crescendo–decrescendo pattern, reflecting the changing velocity of blood flow (Box 18.26). Pansystolic murmurs maintain a constant intensity and extend from the first heart sound throughout systole to the second heart sound, sometimes obscuring it. They occur when blood leaks from a ventricle into a low-pressure chamber at an even or constant velocity. Mitral regurgitation, tricuspid regurgitation and ventricular septal defect are the only causes of a pansystolic murmur. Late systolic murmurs are unusual but may occur in mitral valve prolapse, if the mitral regurgitation is confined to late systole, and hypertrophic obstructive cardiomyopathy, if dynamic obstruction occurs late in systole.

**Diastolic murmurs**

These are due to accelerated or turbulent flow across the mitral or tricuspid valves. They are low-pitched noises that are often difficult to hear and should be evaluated with the bell of the stethoscope. A mid-diastolic murmur may be due to mitral stenosis (located at the apex and axilla), tricuspid stenosis (located at the left sternal edge), increased flow across the mitral valve (e.g. the to-and-fro murmur of severe mitral regurgitation) or increased flow across the tricuspid valve (e.g. left-to-right shunt through a large atrial septal defect). Early diastolic murmurs have a soft, blowing quality with a...
If the sinus rate becomes unduly slow, another, more distal part of the conducting system may assume the role of pacemaker. This is known as an escape rhythm and may arise in the atrioventricular (AV) node or His bundle (junctional rhythm) or the ventricles (idioventricular rhythm).

A cardiac arrhythmia is a disturbance of the electrical rhythm of the heart. Arrhythmias are often a manifestation of structural heart disease but may also occur because of abnormal conduction or depolarisation in an otherwise healthy heart. A heart rate of more than 100/min is called a tachycardia, and a heart rate of less than 60/min is called a bradycardia.

There are three main mechanisms of tachycardia:

1. **Increased automaticity.** The tachycardia is produced by repeated spontaneous depolarisation of an ectopic focus, often in response to catecholamines.
2. **Re-entry.** The tachycardia is initiated by an ectopic beat and sustained by a re-entry circuit (Fig. 18.38). Most tachyarrhythmias are due to re-entry.
3. **Triggered activity.** This can cause ventricular arrhythmias in patients with coronary artery disease. It is a form of secondary depolarisation arising from an incompletely repolarised cell membrane.

Bradyarrhythmia may be due to:

1. **Reduced automaticity, e.g. sinus bradycardia.
2. **Blocked or abnormally slow conduction, e.g. AV block.

Continuous murmurs

These result from a combination of systolic and diastolic flow (e.g. persistent ductus arteriosus), and must be distinguished from extracardiac noises such as bruits from arterial shunts, venous hums (high rates of venous flow in children) and pericardial friction rubs.

The characteristics of specific valve defects and congenital anomalies are described on pages 613 and 629.

---

**DISORDERS OF HEART RATE, RHYTHM AND CONDUCTION**

The heart beat is normally initiated by an electrical discharge from the sinoatrial (sinus) node. The atria and ventricles then depolarise sequentially as electrical depolarisation passes through specialised conducting tissues (see Fig. 18.4, p. 529). The sinus node acts as a pacemaker and its intrinsic rate is regulated by the autonomic nervous system; vagal activity decreases the heart rate, and sympathetic activity increases it via cardiac sympathetic nerves and circulating catecholamines.

If the sinus rate becomes unduly slow, another, more distal part of the conducting system may assume the role of pacemaker. This is known as an escape rhythm and may arise in the atrioventricular (AV) node or His bundle (junctional rhythm) or the ventricles (idioventricular rhythm).

A cardiac arrhythmia is a disturbance of the electrical rhythm of the heart. Arrhythmias are often a manifestation of structural heart disease but may also occur because of abnormal conduction or depolarisation in an otherwise healthy heart. A heart rate of more than 100/min is called a tachycardia, and a heart rate of less than 60/min is called a bradycardia.

1. **Increased automaticity.** The tachycardia is produced by repeated spontaneous depolarisation of an ectopic focus, often in response to catecholamines.
2. **Re-entry.** The tachycardia is initiated by an ectopic beat and sustained by a re-entry circuit (Fig. 18.38). Most tachyarrhythmias are due to re-entry.
3. **Triggered activity.** This can cause ventricular arrhythmias in patients with coronary artery disease. It is a form of secondary depolarisation arising from an incompletely repolarised cell membrane.

Bradyarrhythmia may be due to:

1. **Reduced automaticity, e.g. sinus bradycardia.
2. **Blocked or abnormally slow conduction, e.g. AV block.

---

---

**Fig. 18.38 The mechanism of re-entry.** Re-entry can occur when there are two alternative pathways with different conducting properties (e.g. the AV node and an accessory pathway, or an area of normal and an area of ischaemic tissue). Here, pathway A conducts slowly and recovers quickly, while pathway B conducts rapidly and recovers slowly. (1) In sinus rhythm, each impulse passes down both pathways before entering a common distal pathway. (2) As the pathways recover at different rates, a premature impulse may find pathway A open and B closed. (3) Pathway B may recover while the premature impulse is travelling selectively down pathway A. The impulse can then travel retrogradely up pathway B, setting up a closed loop or re-entry circuit. (4) This may initiate a tachycardia that continues until the circuit is interrupted by a change in conduction rates or electrical depolarisation.
An arrhythmia may be ‘supraventricular’ (sinus, atrial or junctional) or ventricular in origin. Supraventricular rhythms usually produce narrow QRS complexes because the ventricles are depolarised in their normal sequence via the AV node and bundle of His. In contrast, ventricular rhythms produce broad, bizarre QRS complexes because the ventricles are activated in an abnormal sequence. Occasionally, however, a supraventricular rhythm can produce broad or wide QRS complexes due to coexisting bundle branch block or the presence of an additional atrioventricular connection (accessory pathway, see below).

Bradycardias cause symptoms that reflect low cardiac output: fatigue, lightheadedness and syncope. Tachycardias cause rapid palpitation, dizziness, chest discomfort or breathlessness. Extreme tachycardias can cause syncope because the heart is unable to contract or relax properly at extreme rates. Extreme bradycardias or tachycardias can precipitate sudden death or cardiac arrest.

**Sinoatrial nodal rhythms**

**Sinus arrhythmia**

Phasic alteration of the heart rate during respiration (the sinus rate increases during inspiration and slows during expiration) is a consequence of normal parasympathetic nervous system activity and can be pronounced in children. Absence of this normal variation in heart rate with breathing or with changes in posture may be a feature of autonomic neuropathy (p. 831).

**Sinus bradycardia**

A sinus rate of less than 60/min may occur in healthy people at rest and is a common finding in athletes. Some pathological causes are listed in Box 18.27. Asymptomatic sinus bradycardia requires no treatment. Symptomatic acute sinus bradycardia usually responds to intravenous atropine 0.6–1.2 mg. Patients with recurrent or persistent symptomatic sinus bradycardia should be considered for pacemaker implantation.

**Sinus tachycardia**

This is defined as a sinus rate of more than 100/min, and is usually due to an increase in sympathetic activity associated with exercise, emotion, pregnancy or pathology (see Box 18.27). Young adults can produce a rapid sinus rate, up to 200/min, during intense exercise.

**Sinoatrial disease (sick sinus syndrome)**

Sinoatrial disease can occur at any age but is most common in older people. The underlying pathology involves fibrosis, degenerative changes or ischaemia of the SA (sinus) node. The condition is characterised by a variety of arrhythmias (Box 18.28) and may present with palpitation, dizzy spells or syncope, due to intermittent tachycardia, bradycardia, or pauses with no atrial or ventricular activity (SA block or sinus arrest) (Fig. 18.39).
A permanent pacemaker may benefit patients with troublesome symptoms due to spontaneous bradyarrhythmias, or those with symptomatic bradycardias induced by drugs required to prevent tachyarrhythmias. Atrial pacing may prevent episodes of atrial fibrillation. Pacing improves symptoms but not prognosis, and is not indicated in patients who are asymptomatic.

## Atrial tachyarrhythmias

### Atrial ectopic beats (extrasystoles, premature beats)
These usually cause no symptoms but can give the sensation of a missed beat or an abnormally strong beat. The ECG (Fig. 18.40) shows a premature but otherwise normal QRS complex; if visible, the preceding P wave has a different morphology because the atria activate from an abnormal site. In most cases, these are of no consequence, although very frequent atrial ectopic beats may herald the onset of atrial fibrillation. Treatment is rarely necessary but β-blockers can be used if symptoms are intrusive.

![Fig. 18.40 Atrial ectopic beats.](image)

**Fig. 18.40** Atrial ectopic beats. The first, second and fifth complexes are normal sinus beats. The third, fourth and sixth complexes are atrial ectopic beats with identical QRS complexes and abnormal (sometimes barely visible) P waves.

### Atrial tachycardia
Atrial tachycardia may be a manifestation of increased atrial automaticity, sinoatrial disease or digoxin toxicity. It produces a narrow-complex tachycardia with abnormal P-wave morphology, sometimes associated with AV block if the atrial rate is rapid. It may respond to β-blockers, which reduce automaticity, or class I or III anti-arrhythmic drugs (see Box 18.38, p. 575). The ventricular response in rapid atrial tachycardias may be controlled by AV node-blocking drugs. Catheter ablation (p. 577) can be used to target the ectopic site and should be offered as an alternative to anti-arrhythmic drugs in patients with recurrent atrial tachycardia.

### Atrial flutter
Atrial flutter is characterised by a large (macro) re-entry circuit, usually within the right atrium encircling the tricuspid annulus. The atrial rate is approximately 300/min, and is usually associated with 2:1, 3:1 or 4:1 AV block (with corresponding heart rates of 150, 100 or 75/min). Rarely, in young patients, every beat is conducted, producing a rate of 300/min and, potentially, haemodynamic compromise. The ECG shows saw-toothed flutter waves (Fig. 18.41). When there is regular 2:1 AV block, it may be difficult to identify flutter waves that are buried in QRS complexes and T waves. Atrial flutter should always be suspected when there is a narrow-complex tachycardia of 150/min. Carotid sinus pressure or intravenous adenosine may help to establish the diagnosis by temporarily increasing the degree of AV block and revealing flutter waves (Fig. 18.42).

### Management
Digoxin, β-blockers or verapamil can control the ventricular rate (pp. 574–576). However, in many cases, it may be preferable to try to restore sinus rhythm by direct current (DC) cardioversion or by using intravenous amiodarone. Beta-blockers or amiodarone can also be used to prevent recurrent episodes of atrial flutter. Although flecainide can also be used for acute treatment or prophylaxis, it should be avoided because there is a risk of slowing the flutter circuit and facilitating 1:1 AV nodal conduction. This can cause a paradoxical tachycardia and haemodynamic compromise. If used, it should always be prescribed along with an AV node-blocking drug, such as a β-blocker. Catheter ablation offers a 90% chance of complete cure and is the treatment of choice for patients with persistent symptoms.

### Atrial fibrillation
Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, with an overall prevalence of 0.5% in the adult population of the UK. The prevalence rises with age, affecting 1% of those aged 60–64 years, increasing to 9% of those aged over 80 years. AF is a complex arrhythmia characterised by both abnormal automatic firing and the presence of multiple interacting re-entry
AF can be classified as paroxysmal (intermittent episodes which self-terminate within 7 days), persistent (prolonged episodes that can be terminated by electrical or chemical cardioversion) or permanent. In patients with AF seen for the first time, it can be difficult to identify which of these is present. Unfortunately for many patients, paroxysmal AF will become permanent as the underlying disease process that predisposes to AF progresses. Electrophysiological changes occur in the atria within a few hours of the onset of AF that tend to maintain fibrillation: electrical remodelling. When AF persists for a period of months, structural remodelling occurs, with atrial fibrosis and dilatation that further predispose to AF. Thus early treatment of AF will prevent re-initiation of the arrhythmia.

AF may be the first manifestation of many forms of heart disease (Box 18.29), particularly those that are associated with enlargement or dilatation of the atria. Alcohol excess, hyperthyroidism and chronic lung disease are also common causes of AF, although multiple aetiiological factors often coexist, such as the combination of alcohol, hypertension and coronary artery disease. About 50% of all patients with paroxysmal AF and 20% of patients with persistent or permanent AF have structurally normal hearts; this is known as ‘lone atrial fibrillation’.

AF can cause palpitation, breathlessness and fatigue. In patients with poor ventricular function or valve disease, it may precipitate or aggravate cardiac failure because of loss of atrial function and heart rate control. A fall in BP may cause lightheadedness, and chest pain may occur with underlying coronary artery disease. In older patients, AF may not be associated with a rapid ventricular rate and is thus often asymptomatic, in which case it is usually discovered as a result of a routine examination or ECG.

AF is associated with significant morbidity and a twofold increase in mortality (mainly because of its association with other underlying heart disease). By far the most disabling consequence is its association with stroke and systemic embolism. Careful assessment, risk stratification and therapy can markedly improve prognosis.

**Management**

Assessment of patients with newly diagnosed AF includes a full history, physical examination, 12-lead ECG, echocardiogram and thyroid function tests. Additional investigations may be needed to determine the nature and extent of any underlying heart disease. Biochemical evidence of hyperthyroidism is found in a small minority of patients with otherwise unexplained AF.

When AF complicates an acute illness (e.g. chest infection, pulmonary embolism), effective treatment of the primary disorder will often restore sinus rhythm. Otherwise, the main objectives are restoration of sinus rhythm (when possible), prevention of recurrent AF, optimisation of the heart rate during periods of AF, reduction of the risk of thromboembolism, and treatment of underlying cardiac disease.

**Paroxysmal atrial fibrillation**

Occasional attacks that are well tolerated do not necessarily require treatment. Beta-blockers are normally
used as first-line therapy if symptoms are troublesome, and are particularly useful for treating patients with AF associated with coronary artery disease, hypertension and cardiac failure. Beta-blockers reduce the ectopic firing that normally initiates AF. Class Ic drugs (see Box 18.38, p. 575), such as propafenone or flecainide, are also effective at preventing episodes but should not be given to patients with coronary artery disease or left ventricular dysfunction. Flecainide is usually prescribed along with a rate limiting β-blocker because it occasionally precipitates atrial flutter. Class III drugs can also be used; amiodarone is the most effective agent for preventing AF but its side-effects restrict its use to patients in whom other measures fail. Dronedarone is an effective alternative, but is contraindicated in patients with heart failure or significant left ventricular impairment. Digoxin and verapamil are not effective drugs for preventing paroxysms of AF, although they do limit the heart rate when AF occurs by blocking the AV node. In patients with AF in whom anti-arrhythmic drug therapy is ineffective or causes side-effects, catheter ablation can be considered. Ablation is used to disconnect the pulmonary veins from the LA electrically, preventing ectopic triggering of AF. In addition, lines of conduction block can be created within the atria to prevent re-entry. Ablation prevents AF in approximately 75% of patients with prior drug-resistant episodes, although a repeat procedure is sometimes required before this is achieved. Ablation for AF is an attractive treatment for patients in whom drugs are ineffective or poorly tolerated but it is associated with a risk of cardiac tamponade, stroke and other complications.

### 18.30 Atrial fibrillation in old age

- **Prevalence:** rises with age, reaching 9% in those over 80 yrs of age.
- **Symptoms:** sometimes asymptomatic but often accompanied by diastolic heart failure.
- **Hyperthyroidism:** atrial fibrillation may emerge as the dominant feature of otherwise silent or occult hyperthyroidism.
- **Cardioversion:** followed by high rates (~70% at 1 yr) of recurrent atrial fibrillation.
- **Stroke:** atrial fibrillation is an important cause of cerebral embolism, found in 15% of all stroke patients and 2–8% of those with transient ischaemic attacks (TIAs).
- **Anticoagulation:** although the risk of thromboembolism rises, the hazards of anticoagulation also become greater with age because of increased comorbidity, particularly cognitive impairment and falls.
- **Target INR:** if anticoagulation is recommended in those over 75 yrs, care should be taken to maintain an INR below 3.0 because of the increased risk of intracranial haemorrhage.
- **Direct thrombin (e.g. dabigatran) and factor Xa (e.g. rivaroxaban) inhibitors:** alternatives to warfarin. No blood monitoring is required, there are fewer drug interactions, and fixed dosing may aid compliance. Dabigatran dose is reduced from 150 mg twice daily to 110 mg twice daily in the over-eighties or if creatinine clearance is less than 30 mL/min. Rivaroxaban dose is reduced from 20 mg once daily to 15 mg once daily if creatinine clearance is 30–49 mL/min, and contraindicated if below 30 mL/min.

---

### Persistent and permanent atrial fibrillation

There are two options for treating persistent AF:
- **rhythm control:** attempting to restore and maintain sinus rhythm
- **rate control:** accepting that AF will be permanent and using treatments to control the ventricular rate and to prevent embolic complications.

#### Rhythm control

An attempt to restore sinus rhythm is particularly appropriate if the arrhythmia causes troublesome symptoms and if there is a modifiable or treatable underlying cause. Electrical cardioversion (p. 577) is initially successful in three-quarters of patients but relapse is frequent (25–50% at 1 month and 70–90% at 1 year). Attempts to restore and maintain sinus rhythm are most successful if AF has been present for less than 3 months, the patient is young and there is no important structural heart disease.

Immediate cardioversion, after administration of intravenous heparin, is appropriate if AF has been present for under 48 hours. In stable patients with no history of structural heart disease, intravenous flecainide (2 mg/kg over 30 mins, maximum dose 150 mg) can be used for pharmacological cardioversion and will restore sinus rhythm in 75% of patients within 8 hours. In patients with structural or ischaemic heart disease, intravenous amiodarone can be given via a central venous catheter. Electrical cardioversion, using a DC shock, is an alternative and is often effective when drugs fail. In other situations, DC cardioversion should be deferred until the patient has been established on warfarin, with an international normalised ratio (INR) of more than 2.0 for a minimum of 4 weeks, and any underlying problems, such as hypertension or alcohol excess, have been eliminated. Anticoagulation should be maintained for at least 3 months following successful cardioversion. If AF recurs, further cardioversion may be appropriate but consideration should be given to pretreatment with amiodarone to reduce the risk of recurrence. Catheter ablation is sometimes used to help restore and maintain sinus rhythm in resistant cases, but is a less effective treatment than for paroxysmal AF.

#### Rate control

If sinus rhythm cannot be restored, treatment should be directed at maintaining an appropriate heart rate. Digoxin, β-blockers and rate-limiting calcium antagonists, such as verapamil or diltiazem (pp. 574–576), reduce the ventricular rate by increasing the degree of AV block. This alone may produce a striking improvement in cardiac function, particularly in patients with mitral stenosis. Beta-blockers and rate-limiting calcium antagonists are more effective than digoxin at controlling the heart rate during exercise and have additional benefits in patients with hypertension or structural heart disease. Combination therapy (e.g. digoxin and atenolol) is often advisable but rate-limiting calcium channel antagonists should not be used with β-blockers because of the risk of bradycardia.

In exceptional cases, poorly controlled and symptomatic AF can be treated by deliberately inducing complete AV nodal block with catheter ablation; a permanent pacemaker must be implanted beforehand. This is known as the ‘pace and ablate’ strategy.

### Prevention of thromboembolism

Loss of atrial contraction and left atrial dilatation cause stasis of blood in the LA and may lead to thrombus...
formation in the left atrial appendage. This predisposes patients to stroke and other forms of systemic embolism. The annual risk of stroke in patients with AF (Box 18.31) is influenced by many factors, and in each patient a decision has to be made about the risk of stroke versus the risk of anticoagulation.

Several large randomised trials have shown that treatment with adjusted-dose warfarin (target INR 2.0–3.0) reduces the risk of stroke by about two-thirds, at the cost of an annual risk of bleeding of 1%–1.5%, whereas treatment with aspirin reduces the risk of stroke by only one-fifth, is associated with significant bleeding risk and, although still included in European guidelines, has a very limited role. Warfarin is thus indicated for patients with AF who have specific risk factors for stroke. In intermittent AF, the risk of stroke is only loosely related to the frequency and duration of AF episodes, so stroke prevention guidelines do not distinguish between those with paroxysmal, persistent and permanent AF.

An assessment of the risk of embolism helps to define the possible benefits of antithrombotic therapy (see Box 18.31), which must be balanced against its potential hazards. Risk stratification is based on clinical factors using the CHA₂DS₂-VASc scoring system. Echocardiographic assessment (e.g. left atrial size) is of limited value in predicting stroke risk and is mainly used to identify associated structural disease. Oral anticoagulation is indicated in patients at moderate or high risk of stroke, whereas there is no current evidence of benefit for the direct thrombin inhibitors. Young patients (under 65 years) with no evidence of structural heart disease have a very low risk of stroke and may not require oral anticoagulation.

### ‘Supraventricular’ tachycardias

The term ‘supraventricular tachycardia’ (SVT) is commonly used to describe regular tachycardias that have a similar appearance on ECG. These are usually associated with a narrow QRS complex and are characterised by a re-entry circuit or automatic focus involving the atria. The term SVT is misleading, as, in many cases, the ventricles also form part of the re-entry circuit, such as in patients with AV re-entrant tachycardia.

### Atrioventricular nodal re-entrant tachycardia

Atrioventricular nodal re-entrant tachycardia (AVNRT) is due to re-entry in a circuit involving the AV node and its two right atrial input pathways: a superior ‘fast’ pathway and an inferior ‘slow’ pathway (see Fig. 18.46A below). This produces a regular tachycardia with a rate of 120–240/min. It tends to occur in the absence of structural heart disease and episodes may last from a few seconds to many hours. The patient is usually aware of a rapid, very forceful, regular heart beat and may experience chest discomfort, lightheadedness or breathlessness. Polyuria, mainly due to the release of atrial natriuretic peptide, is sometimes a feature. The ECG (Fig. 18.45) usually shows a tachycardia with normal drug interactions, are all relative contraindications. In warfarin-treated patients, anticoagulation can be reversed by administering vitamin K or clotting factors, but there are no current antidotes for the direct thrombin inhibitors. Young patients (under 65 years) with no evidence of structural heart disease have a very low risk of stroke and may not require oral anticoagulation.
QRS complexes but occasionally there may be rate-dependent bundle branch block.

**Management**

Treatment is not always necessary. However, an episode may be terminated by carotid sinus pressure or by the Valsalva manoeuvre. Adenosine (3–12 mg rapidly IV in incremental doses until tachycardia stops) or verapamil (5 mg IV over 1 min) will restore sinus rhythm in most cases. Intravenous β-blocker or flecainide can also be used. In rare cases, when there is severe haemodynamic compromise, the tachycardia should be terminated by DC cardioversion (p. 577).

In patients with recurrent SVT, catheter ablation (p. 577) is the most effective therapy and will permanently prevent SVT in more than 90% of cases. Alternatively, prophylaxis with oral β-blocker, verapamil or flecainide may be used but commits predominantly young patients to long-term drug therapy and can create difficulty in female patients, as these drugs are normally avoided during pregnancy.

**Wolff–Parkinson–White syndrome and atrioventricular re-entrant tachycardia**

Here, an abnormal band of conducting tissue connects the atria and ventricles. This ‘accessory pathway’ comprises rapidly conducting fibres which resemble Purkinje tissue, in that they conduct very rapidly and are rich in sodium channels. In around half of cases, this pathway only conducts in the retrograde direction (from ventricles to atria) and thus does not alter the appearance of the ECG in sinus rhythm. This is known as a concealed accessory pathway. In the rest, the pathway also conducts antegrade (from atria to ventricles) so AV conduction in sinus rhythm is mediated via both the AV node and the accessory pathway, distorting the QRS complex. Premature ventricular activation via the pathway shortens the PR interval and produces a ‘slurred’ initial deflection of the QRS complex, called a delta wave (Fig. 18.46B). This is known as a manifest accessory pathway. As the AV node and accessory pathway have different conduction speeds and refractory periods, a re-entry circuit can develop, causing tachycardia (Fig. 18.46C); when associated with symptoms, the condition is known as Wolff–Parkinson–White syndrome. The ECG during this tachycardia is almost indistinguishable from that of AVNRT (Fig. 18.46A). Carotid sinus pressure or intravenous adenosine can terminate the tachycardia. If atrial fibrillation occurs, it may produce a dangerously rapid ventricular rate because the accessory pathway lacks the rate-limiting properties of the AV node (Fig. 18.46D). This is known as pre-excited atrial fibrillation and may cause collapse, syncope and even death. It should be treated as an emergency, usually with DC cardioversion.

Catheter ablation is first-line treatment in symptomatic patients and is nearly always curative. Alternatively, prophylactic anti-arrhythmic drugs, such as

---

**Fig. 18.46 AV nodal re-entrant tachycardia (AVNRT) and Wolff–Parkinson–White (WPW) syndrome.**

A AV node re-entrant tachycardia. The mechanism of AVNRT occurs via two right atrial AV nodal input pathways: the slow (S) and fast (F) pathways. Antegrade conduction occurs via the slow pathway; the wavefront enters the AV node and passes into the ventricles, at the same time re-entering the atria via the fast pathway.

In WPW syndrome, there is a strip of accessory conducting tissue that allows electricity to bypass the AV node and spread from the atria to the ventricles rapidly and without delay. When the ventricles are depolarised through the AV node, the ECG is normal, but when the ventricles are depolarised through the accessory conducting tissue, the ECG shows a very short PR interval and a broad QRS complex. B Sinus rhythm. In sinus rhythm, the ventricles are depolarised through (1) the AV node and (2) the accessory pathway, producing an ECG with a short PR interval and broadened QRS complexes; the characteristic slurring of the upstroke of the QRS complex is known as a delta wave. The degree of pre-excitation (the proportion of activation passing down the accessory pathway) and therefore the ECG appearances may vary a lot, and at times the ECG can look normal.

C Orthodromic tachycardia. This is the most common form of tachycardia in WPW. The re-entry circuit passes antegrade through the AV node and retrogradely through the accessory pathway. The ventricles are therefore depolarised in the normal way, producing a narrow-complex tachycardia that is indistinguishable from other forms of SVT. D Atrial fibrillation. In this rhythm, the ventricles are largely depolarised through the accessory pathway, producing an irregular broad-complex tachycardia which is often more rapid than the example shown.
Disorders of heart rate, rhythm and conduction

Ventricular ectopic beats (extrasystoles, premature beats)

QRS complexes in sinus rhythm are normally narrow because the ventricles are activated rapidly and simultaneously via the His–Purkinje system. The complexes of ventricular ectopic beats (VEBs) are premature, broad and bizarre because the ventricles are activated sequentially rather than simultaneously. The complexes may be unifocal (identical beats arising from a single ectopic focus) or multifocal (varying morphology with multiple foci, Fig. 18.47). ‘Couplet’ and ‘triplet’ are the terms used to describe two or three successive ectopic beats. A run of alternating sinus and ventricular ectopic beats is known as ventricular ‘bigeminy’. Ectopic beats produce a low stroke volume because left ventricular contraction occurs before filling is complete. The pulse is therefore irregular, with weak or missed beats (see Fig. 18.47). Patients are usually asymptomatic but may complain of an irregular heart beat, missed beats or abnormally strong beats (due to the increased output of the post-ectopic sinus beat). The significance of VEBs depends on the presence or absence of underlying heart disease.

Ventricular ectopic beats in otherwise healthy subjects

VEBs are frequently found in healthy people and their prevalence increases with age. Ectopic beats in patients with otherwise normal hearts are more prominent at rest and disappear with exercise. Treatment is not necessary, unless the patient is highly symptomatic, in which case β-blockers or, in some situations, catheter ablation can be used. VEBs are sometimes a manifestation of otherwise subclinical heart disease, such as coronary artery disease or cardiomyopathy. There is no evidence that anti-arrhythmic therapy improves prognosis but the discovery of very frequent VEBs should prompt investigations, such as an echocardiogram (looking for structural heart disease) and an exercise stress test (to detect underlying ischaemic heart disease).

Ventricular ectopic beats associated with heart disease

Frequent VEBs often occur during acute MI but need no treatment. Persistent, frequent (over 10/hr) VEBs in patients who have survived the acute phase of MI indicate a poorer long-term outcome. Other than β-blockers, anti-arrhythmic drugs do not improve and may even worsen prognosis.

VEBs are common in patients with heart failure of any cause, including cardiomyopathy. While they are associated with an adverse prognosis, this is not improved by anti-arrhythmic drugs. Effective treatment of the heart failure may suppress the ectopic beats.

VEBs are also a feature of digoxin toxicity, and may occur as ‘escape beats’ in patients with bradycardia. Treatment is that of the underlying condition.

Ventricular tachycardia

Ventricular tachycardia (VT) occurs most commonly in the settings of acute MI, chronic coronary artery disease, and cardiomyopathy. It occurs when there is extensive ventricular disease, such as impaired left ventricular function or a left ventricular aneurysm. In these settings, VT may cause haemodynamic compromise or degenerate into ventricular fibrillation (p. 557). It is caused by abnormal automaticity or triggered activity in ischaemic tissue, or by re-entry within scarred ventricular tissue. Patients may complain of palpitation or symptoms of low cardiac output, e.g. dizziness, dyspnoea or syncope. The ECG shows tachycardia and broad, abnormal QRS complexes with a rate of more than 120/min (Fig. 18.48). VT may be difficult to distinguish from SVT with bundle branch block or pre-excitation (WPW syndrome). Features in favour of a diagnosis of VT are listed in Box 18.33. A 12-lead ECG (Fig. 18.49) or electrophysiology
Cardiovascular disease

Arrhythmia). In patients at high risk of arrhythmic death (e.g. those with poor left ventricular function, or where VT is associated with haemodynamic compromise), the use of an implantable cardiac defibrillator is recommended (p. 579). Rarely, surgery (e.g. aneurysm resection) or catheter ablation can be used to interrupt the arrhythmia focus or circuit in patients with VT associated with a myocardial infarct scar.

Torsades de pointes (ventricular tachycardia)

This form of polymorphic VT is a complication of prolonged ventricular repolarisation (prolonged QT interval). The ECG shows rapid irregular complexes that oscillate from an upright to an inverted position and seem to twist around the baseline as the mean QRS axis changes (Fig. 18.50). The arrhythmia is usually non-sustained and repetitive, but may degenerate into ventricular fibrillation. During periods of sinus rhythm, the ECG will usually show a prolonged QT interval (> 0.43 s in men, > 0.45 s in women when corrected to a heart rate of 60/min).

Management

Prompt action to restore sinus rhythm is required and should usually be followed by prophylactic therapy. Synchronised DC cardioversion is the treatment of choice if systolic BP is less than 90 mmHg. If the arrhythmia is well tolerated, intravenous amiodarone may be given as a bolus, followed by a continuous infusion (p. 576). Intravenous lidocaine can be used but may depress left ventricular function, causing hypotension or acute heart failure. Hypokalaemia, hypomagnesaemia, acidosis and hypoxaemia should be corrected.

Beta-blockers are effective at preventing VT by reducing ventricular automaticity. Amiodarone can be added if additional control is needed. Class Ic anti-arrhythmic drugs should not be used for prevention of VT in patients with coronary artery disease or heart failure because they depress myocardial function and can be pro-arrhythmic (increase the likelihood of a dangerous arrhythmia). In patients at high risk of arrhythmic death (e.g. those with poor left ventricular function, or where VT is associated with haemodynamic compromise), the use of an implantable cardiac defibrillator is recommended (p. 579). Rarely, surgery (e.g. aneurysm resection) or catheter ablation can be used to interrupt the arrhythmia focus or circuit in patients with VT associated with a myocardial infarct scar.

Fig. 18.48 Ventricular tachycardia: fusion beat (arrow). In ventricular tachycardia, there is independent atrial and ventricular activity. Occasionally, a P wave is conducted to the ventricles through the AV node, producing a normal sinus beat in the middle of the tachycardia (a capture beat); more commonly, however, the conducted impulse fuses with an impulse from the tachycardia (a fusion beat). This can only occur when there is AV dissociation and is therefore diagnostic of ventricular tachycardia.

18.33 Features more in keeping with ventricular tachycardia

- History of MI
- AV dissociation (pathognomonic)
- Capture/fusion beats (pathognomonic; see Fig. 18.48)
- Extreme left axis deviation
- Very broad QRS complexes (> 140 ms)
- No response to carotid sinus massage or IV adenosine

Fig. 18.49 Ventricular tachycardia: 12-lead ECG. There are typically very broad QRS complexes and marked left axis deviation. There is also AV dissociation; some P waves are visible and others are buried in the QRS complexes (arrows).
Disorders of heart rate, rhythm and conduction

The Brugada syndrome is a related genetic disorder that may present with polymorphic VT or sudden death. It is characterised by a defect in sodium channel function and an abnormal ECG (right bundle branch block and ST elevation in V1 and V2 but not usually prolongation of the QT interval). The only known effective treatment is an implantable defibrillator.

Atrioventricular and bundle branch block

Atrioventricular block

Atrioventricular conduction is influenced by autonomic activity. AV block can therefore be intermittent and only evident when the conducting tissue is stressed by a rapid atrial rate. Accordingly, atrial tachyarrhythmias are often associated with AV block (see Fig. 18.44, p. 565).

First-degree atrioventricular block

In this condition, AV conduction is delayed and so the PR interval is prolonged (> 0.20 s; Fig. 18.51). It rarely causes symptoms.

Second-degree atrioventricular block

In this, dropped beats occur because some impulses from the atria fail to conduct to the ventricles. In Mobitz type I second-degree AV block (Fig. 18.52), there is progressive lengthening of successive PR intervals, culminating in a dropped beat. The cycle then repeats itself. This is known as the Wenckebach phenomenon and is usually due to impaired conduction in the AV node itself. The phenomenon may be physiological and is sometimes observed at rest or during sleep in athletic young adults with high vagal tone.

In Mobitz type II second-degree AV block (Fig. 18.53), the PR interval of the conducted impulses remains constant but some P waves are not conducted. This is usually caused by disease of the His–Purkinje system and carries a risk of asystole.

In 2:1 AV block (Fig. 18.54), alternate P waves are conducted, so it is impossible to distinguish between Mobitz type I and type II block.

Some of the common causes are listed in Box 18.34. The arrhythmia is more common in women and is often triggered by a combination of aetiological factors (e.g. QT-prolonging medications and hypokalaemia). The congenital long QT syndromes are a family of genetic disorders that are characterised by mutations in genes that code for cardiac sodium or potassium channels. Long QT syndrome subtypes have different triggers, which are important when counselling patients. Adrenergic stimulation (e.g. exercise) is a common trigger in long QT type 1, and a sudden noise (e.g. an alarm clock) may trigger arrhythmias in long QT type 2. Arrhythmias are more common during sleep in type 3.

Treatment should be directed at the underlying cause. Intravenous magnesium (8 mmol over 15 mins, then 72 mmol over 24 hrs) should be given in all cases. Atrial pacing will usually suppress the arrhythmia through rate-dependent shortening of the QT interval. Intravenous isoprenaline is a reasonable alternative to pacing but should be avoided in patients with the congenital long QT syndromes.

Long-term therapy may not be necessary if the underlying cause can be removed. Beta-blockers are effective at preventing syncope in patients with congenital long QT syndrome. Some patients, particularly those with extreme QT interval prolongation (> 500 ms) or certain high-risk genotypes should be considered for implantation of a defibrillator. Left stellate ganglion block may be of value in patients with resistant arrhythmias.

Some of the common causes are listed in Box 18.34. The arrhythmia is more common in women and is often triggered by a combination of aetiological factors (e.g. QT-prolonging medications and hypokalaemia). The congenital long QT syndromes are a family of genetic disorders that are characterised by mutations in genes that code for cardiac sodium or potassium channels. Long QT syndrome subtypes have different triggers, which are important when counselling patients. Adrenergic stimulation (e.g. exercise) is a common trigger in long QT type 1, and a sudden noise (e.g. an alarm clock) may trigger arrhythmias in long QT type 2. Arrhythmias are more common during sleep in type 3.

Treatment should be directed at the underlying cause. Intravenous magnesium (8 mmol over 15 mins, then 72 mmol over 24 hrs) should be given in all cases. Atrial pacing will usually suppress the arrhythmia through rate-dependent shortening of the QT interval. Intravenous isoprenaline is a reasonable alternative to pacing but should be avoided in patients with the congenital long QT syndromes.

Long-term therapy may not be necessary if the underlying cause can be removed. Beta-blockers are effective at preventing syncope in patients with congenital long QT syndrome. Some patients, particularly those with extreme QT interval prolongation (> 500 ms) or certain high-risk genotypes should be considered for implantation of a defibrillator. Left stellate ganglion block may be of value in patients with resistant arrhythmias.

The Brugada syndrome is a related genetic disorder that may present with polymorphic VT or sudden death. It is characterised by a defect in sodium channel function and an abnormal ECG (right bundle branch block and ST elevation in V1 and V2 but not usually prolongation of the QT interval). The only known effective treatment is an implantable defibrillator.
**Cardiovascular Disease**

**Fig. 18.52** Second-degree AV block (Mobitz type I – Wenckebach’s phenomenon). The PR interval progressively increases until a P wave is not conducted. The cycle then repeats itself. In this example, conduction is at a ratio of 4:3, leading to groupings of three ventricular complexes in a row.

**Fig. 18.53** Second-degree AV block (Mobitz type II). The PR interval of conducted beats is normal but some P waves are not conducted. The constant PR interval distinguishes this from Wenckebach’s phenomenon.

**Fig. 18.54** Second-degree AV block with fixed 2:1 block. Alternate P waves are not conducted. This may be due to Mobitz type I or II block.

**Fig. 18.55** Complete (third-degree) AV block. There is complete dissociation of atrial and ventricular complexes. The atrial rate is 80/min and the ventricular rate is 38/min.

**Third-degree (complete) atrioventricular block**

When AV conduction fails completely, the atria and ventricles beat independently (AV dissociation, Fig. 18.55). Ventricular activity is maintained by an escape rhythm arising in the AV node or bundle of His (narrow QRS complexes) or the distal Purkinje tissues (broad QRS complexes). Distal escape rhythms tend to be slower and less reliable.

Complete AV block (Box 18.35) produces a slow (25–50/min), regular pulse that, except in the case of congenital complete AV block, does not vary with exercise. There is usually a compensatory increase in stroke volume, producing a large-volume pulse. Cannon waves may be visible in the neck and the intensity of the first heart sound varies due to the loss of AV synchrony.

**Stokes–Adams attacks**

Episodes of ventricular asystole may complicate complete heart block or Mobitz type II second-degree AV block, or occur in patients with sinoatrial disease (see Fig. 18.39, p. 563). This may cause recurrent syncope or ‘Stokes-Adams’ attacks.

A typical episode is characterised by sudden loss of consciousness that occurs without warning and results in collapse. A brief anoxic seizure (due to cerebral ischaemia) may occur if there is prolonged asystole. There is pallor and a death-like appearance during the attack, but when the heart starts beating again, there is a characteristic flush. Unlike in epilepsy, recovery is rapid. Sinoatrial disease and neurocardiogenic syncope (p. 555) may cause similar symptoms.

**Management**

**Atrioventricular block complicating acute myocardial infarction**

Acute inferior MI is often complicated by transient AV block because the right coronary artery (RCA) supplies the AV node. There is usually a reliable escape rhythm and, if the patient remains well, no treatment is required. Symptomatic second- or third-degree AV block may respond to atropine (0.6 mg IV, repeated as necessary) or, if this fails, a temporary pacemaker. In most cases, the AV block will resolve within 7–10 days.

Second- or third-degree AV heart block complicating acute anterior MI indicates extensive ventricular damage involving both bundle branches and carries a
Disorders of heart rate, rhythm and conduction

Bundle branch block and hemiblock

Conduction block in the right or left bundle branch can occur as a result of many pathologies, including ischaemic or hypertensive heart disease or cardiomyopathies (Box 18.36). Depolarisation proceeds through a slow myocardial route in the affected ventricle rather than through the rapidly conducting Purkinje tissues that constitute the bundle branches. This causes delayed conduction into the LV or RV, broadens the QRS complex (≥ 0.12 s) and produces the characteristic alterations in QRS morphology (Figs 18.56 and 18.57). Right bundle branch block (RBBB) can occur in healthy people but left bundle branch block (LBBB) often signifies important underlying heart disease.

The left bundle branch divides into an anterior and a posterior fascicle. Damage to the conducting tissue at this point (hemiblock) does not broaden the QRS complex but alters the mean direction of ventricular depolarisation (mean QRS axis), causing left axis deviation in left anterior hemiblock and right axis deviation in left posterior hemiblock (see Fig. 18.7, p. 533). The combination of right bundle branch block and left anterior or posterior hemiblock is known as bifascicular block.

Anti-arrhythmic drug therapy

Classification

Anti-arrhythmic drugs may be classified according to their mode or site of action (Box 18.37 and Fig. 18.58).
Identification of ion channel subtypes has led to refinement of drug classifications, according to the specific mechanisms targeted. The Vaughan-Williams classification is a simple system but is convenient for describing the main mode of action of anti-arrhythmic drugs (Box 18.38) that should be used following guiding principles (Box 18.39). Anti-arrhythmic drugs can also be more accurately categorised by referring to the cardiac ion channels and receptors on which they act.

**Class I drugs**

Class I drugs act principally by suppressing excitability and slowing conduction in atrial or ventricular muscle. They block sodium channels, of which there are several types in cardiac tissue. These drugs should generally be avoided in patients with heart failure because they depress myocardial function, and class Ia and Ic drugs are often pro-arrhythmic.

**Class Ia drugs**

These prolong cardiac action potential duration and increase the tissue refractory period. They are used to prevent both atrial and ventricular arrhythmias. **Disopyramide.** An effective drug but causes anticholinergic side-effects, such as urinary retention, and can precipitate glaucoma. It can depress myocardial function and should be avoided in cardiac failure. **Quinidine.** Now rarely used, as it increases mortality and causes gastrointestinal upset.

**Class Ib drugs**

These shorten the action potential and tissue refractory period. They act on channels found predominantly in ventricular myocardium and so are used to treat or prevent ventricular tachycardia and ventricular fibrillation. **Lidocaine.** Must be given intravenously and has a very short plasma half-life. **Mexiletine.** Can be given intravenously or orally, but has many side-effects (see Box 18.38).

**Class Ic drugs**

These affect the slope of the action potential without altering its duration or refractory period. They are used mainly for prophylaxis of atrial fibrillation but are effective in prophylaxis and treatment of supraventricular or ventricular arrhythmias. They are useful for WPW syndrome because they block conduction in accessory pathways. They should not be used as oral prophylaxis in patients with previous MI because of pro-arrhythmia. **Flecainide.** Effective for prevention of atrial fibrillation, and an intravenous infusion may be used for pharmacological cardioversion of atrial fibrillation of less than 24 hours’ duration. It should be prescribed along with an AV node-blocking drug, such as a β-blocker, to prevent pro-arrhythmia. **Propafenone.** Also has some β-blocker (class II) properties. Important interactions with digoxin, warfarin and cimetidine have been described.

**Class II drugs**

This group comprises the β-adrenoceptor antagonists (β-blockers). These agents reduce the rate of SA node depolarisation and cause relative block in the AV node, making them useful for rate control in atrial flutter and atrial fibrillation. They can be used to prevent supraventricular and ventricular tachycardia. They reduce myocardial excitability and the risk of arrhythmic death in patients with coronary artery disease and heart failure. **‘Non-selective’ β-blockers.** Act on both β₁ and β₂ receptors. Beta₂ blockade causes side-effects, such as bronchospasm and peripheral vasoconstriction. **Propranolol** is non-selective and is subject to extensive first-pass metabolism in the liver. The effective oral dose is therefore unpredictable and must be titrated after treatment is started with a small dose. Other non-selective drugs include **nadolol** and **carvedilol.**
### 18.38 The main uses, dosages and side-effects of the most widely used anti-arrhythmic drugs

<table>
<thead>
<tr>
<th>Drug</th>
<th>Main uses</th>
<th>Route</th>
<th>Dose (adult)</th>
<th>Important side-effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disopyramide</td>
<td>Prevention and treatment of atrial and ventricular</td>
<td>IV</td>
<td>2 mg/kg at 30 mg/min, then 0.4 mg/kg/hr (max 800 mg/day)</td>
<td>Myocardial depression, hypotension, dry mouth, urinary retention</td>
</tr>
<tr>
<td></td>
<td>tachyarrhythmias</td>
<td>Oral</td>
<td>300–800 mg daily in divided dosage</td>
<td></td>
</tr>
<tr>
<td>Lidocaine</td>
<td>Treatment and short-term prevention of VT and VF</td>
<td>IV</td>
<td>Bolus 50–100 mg, 4 mg/min for 30 mins, then 2 mg/min for 2 hrs, then 1 mg/min for 24 hrs</td>
<td>Myocardial depression, confusion, convulsions</td>
</tr>
<tr>
<td>Mexiletine</td>
<td>Prevention and treatment of ventricular tachyarrhythmias</td>
<td>IV</td>
<td>Loading dose: 100–250 mg at 25 mg/min, then 250 mg in 1 hr, then 250 mg in 2 hrs</td>
<td>Myocardial depression, gastrointestinal irritation, confusion, dizziness, tremor, nystagmus, ataxia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td>Maintenance therapy: 0.5 mg/min, 200–250 mg 3 times daily</td>
<td></td>
</tr>
<tr>
<td>Flecaïnide</td>
<td>Prevention and treatment of atrial and ventricular</td>
<td>IV</td>
<td>2 mg/kg over 10 mins, then 1.5 mg/kg/hr for 1 hr, then 0.1 mg/kg/hr</td>
<td>Myocardial depression, dizziness</td>
</tr>
<tr>
<td></td>
<td>tachyarrhythmias</td>
<td>Oral</td>
<td>50–100 mg twice daily</td>
<td></td>
</tr>
<tr>
<td>Propafenone</td>
<td>Prevention and treatment of atrial and ventricular</td>
<td>Oral</td>
<td>150 mg 3 times daily for 1 wk, then 300 mg twice daily</td>
<td>Myocardial depression, dizziness</td>
</tr>
<tr>
<td></td>
<td>tachyarrhythmias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atenolol</td>
<td>Treatment and prevention of SVT and AF</td>
<td>IV</td>
<td>2.5 mg at 1 mg/min, repeated at 5-min intervals (max 10 mg)</td>
<td>Myocardial depression, bradycardia, bronchospasm, fatigue, depression, nightmares, cold peripheries</td>
</tr>
<tr>
<td>Bisoprolol</td>
<td>Prevention of VEs and exercise-induced VT</td>
<td>Oral</td>
<td>25–100 mg daily</td>
<td></td>
</tr>
<tr>
<td>Metoprolol</td>
<td></td>
<td>IV</td>
<td>2.5–10 mg daily</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td>5 mg over 2 mins to a maximum of 15 mg</td>
<td></td>
</tr>
<tr>
<td><strong>Class III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amiodarone</td>
<td>Serious or resistant atrial and ventricular</td>
<td>IV</td>
<td>5 mg/kg over 20–120 mins, then up to 15 mg/kg/24 hrs</td>
<td>Photosensitivity, skin discoloration, corneal deposits, thyroid dysfunction, alveolitis, nausea and vomiting, hepatotoxicity, peripheral neuropathy, torsades de pointes; potentiates digoxin and warfarin</td>
</tr>
<tr>
<td></td>
<td>tachyarrhythmias</td>
<td>Oral</td>
<td>Initially 600–1200 mg/day, then 100–400 mg daily</td>
<td></td>
</tr>
<tr>
<td>Dronedarone</td>
<td>Paroxysmal atrial fibrillation</td>
<td>Oral</td>
<td>400 mg twice daily</td>
<td>Renal and hepatic dysfunction requiring regular blood monitoring</td>
</tr>
<tr>
<td>Sotalol</td>
<td>Atrial fibrillation, rarely ventricular tachyarrhythmias</td>
<td>IV</td>
<td>10–20 mg slowly</td>
<td>Can cause torsades de pointes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td>40–160 mg twice daily</td>
<td></td>
</tr>
<tr>
<td><strong>Class IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verapamil</td>
<td>Treatment of SVT, control of AF</td>
<td>IV</td>
<td>5–10 mg over 30 secs 40–120 mg 3 times daily or 240 mg SR daily</td>
<td>Myocardial depression, hypotension, bradycardia, constipation</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atropine</td>
<td>Treatment of bradycardia and/or hypotension due to</td>
<td>IV</td>
<td>0.6–3 mg</td>
<td>Dry mouth, thirst, blurred vision, atrial and ventricular extrasystoles</td>
</tr>
<tr>
<td></td>
<td>vagal overactivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenosine</td>
<td>Treatment of SVT, aid to diagnosis in unidentified</td>
<td>IV</td>
<td>3 mg over 2 secs, followed if necessary by 6 mg, then 12 mg at intervals of</td>
<td>Flushing, dyspnoea, chest pain AVOID IN ASTHMA</td>
</tr>
<tr>
<td></td>
<td>tachycardia</td>
<td></td>
<td>1–2 mins</td>
<td></td>
</tr>
<tr>
<td>Digoxin</td>
<td>Treatment and prevention of SVT, rate control of AF</td>
<td>IV</td>
<td>Loading dose: 0.5–1 mg (total), 0.5 mg over 30 mins, then 0.25–0.5 mg 3–6 times daily, to maximum total of 1 mg, assessing response before each additional dose</td>
<td>Gastrointestinal disturbance, xanthispia, arrhythmias (see Box 18.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral</td>
<td>0.5 mg 4 times daily for 2 doses, then 0.125–0.25 mg daily</td>
<td></td>
</tr>
</tbody>
</table>

(AF = atrial fibrillation; SR = sustained-release formulation; SVT = supraventricular tachycardia; VE = ventricular ectopic; VF = ventricular fibrillation; VT = ventricular tachycardia)
CARDIOVASCULAR DISEASE

18.39 Anti-arrhythmic drugs: principles of use

Anti-arrhythmic drugs are potentially toxic and should be used carefully according to the following principles:
- Many arrhythmias are benign and do not require specific treatment
- Precipitating or causal factors should be corrected if possible, e.g. excess alcohol or caffeine consumption, myocardial ischaemia, hyperthyroidism, acidosis, hypokalaemia and hypomagnesaemia
- If drug therapy is required, it is best to use as few drugs as possible
- In difficult cases, programmed electrical stimulation (electrophysiological study) may help to identify the optimum therapy
- When managing life-threatening arrhythmias, it is essential to ensure that prophylactic treatment is effective. Ambulatory monitoring and exercise testing may be of value
- Patients on long-term anti-arrhythmic drugs should be reviewed regularly and attempts made to withdraw therapy if the factors which precipitated the arrhythmias are no longer operative
- For patients with recurrent SVT, radiofrequency ablation is often preferable to long-term drug therapy

‘Cardioselective’ β-blockers. Act mainly on myocardial β, receptors and are relatively well tolerated. Bisoprolol and metoprolol are examples of cardioselective β-blockers.

Sotalol. A racemic mixture of two isomers with non-selective β-blocker (mainly l-sotalol) and class III (mainly d-sotalol) activity. It may cause torsades de pointes.

Class III drugs

Class III drugs act by prolonging the plateau phase of the action potential, thus lengthening the refractory period. These drugs are very effective at preventing atrial and ventricular tachyarrhythmias. They cause QT interval prolongation and can predispose to torsades de pointes and ventricular tachycardia (p. 570), especially in patients with other predisposing risk factors (see Box 18.38, p. 571).

Amiodarone. The principal drug in this class, although both disopyramide and sotalol have class III activity. Amiodarone is a complex drug that also has class I, II and IV activity. It is probably the most effective drug currently available for controlling paroxysmal atrial fibrillation. It is also used to prevent episodes of recurrent ventricular tachycardia, particularly in patients with poor left ventricular function or those with implantable defibrillators (to prevent unnecessary DC shocks). Amiodarone has a very long tissue half-life (25–110 days). An intravenous or oral loading regime is often used to achieve therapeutic tissue concentrations rapidly. The drug’s effects may last for weeks or months after treatment has been stopped. Side-effects are common (up to one-third of patients), numerous and potentially serious. Drug interactions are also common (see Box 18.38).

Dronedarone. A related drug that has a short tissue half-life and fewer side-effects. It has recently been shown to be effective at preventing episodes of atrial flutter and fibrillation. It is contraindicated in patients with permanent atrial fibrillation, or if there is heart failure or left ventricular impairment, because it increases mortality. Regular liver function test monitoring is required.

Class IV drugs

These block the ‘slow calcium channel’, which is important for impulse generation and conduction in atrial and nodal tissue, although it is also present in ventricular muscle. Their main indications are prevention of supraventricular tachycardia (by blocking the AV node) and rate control in patients with atrial fibrillation.

Verapamil. The most widely used drug in this class. Intravenous verapamil may cause profound bradycardia or hypotension, and should not be used in conjunction with β-blockers.

Diltiazem. Has similar properties.

Other anti-arrhythmic drugs

Adenosine sulphate (0.6 mg IV, repeated if necessary to a maximum of 3 mg). Increases the sinus rate and SA and AV conduction, and is the treatment of choice for severe bradycardia or hypotension due to vagal overactivity. It is used for initial management of symptomatic bradyarrhythmias complicating inferior MI, and in cardiac arrest due to asystole. Repeat dosing may be necessary because the drug disappears rapidly from the circulation after parenteral administration. Side-effects are listed in Box 18.38.

Dronedarone. A purified glycoside from the European foxglove, Digitalis lanata, which slows conduction and prolongs the refractory period in the AV node. This effect helps to control the ventricular rate in atrial fibrillation and may interrupt supraventricular tachycardias involving the AV node. Dronedarone also shortens refractory periods and enhances excitability and conduction in other parts of the heart (including accessory conduction pathways). It may therefore increase atrial and ventricular ectopic activity and can lead to more complex atrial
and ventricular tachyarrhythmias. Digoxin is largely excreted by the kidneys, and the maintenance dose (see Box 18.38) should be reduced in children, older people and those with renal impairment. It is widely distributed and has a long tissue half-life (36 hours), so that effects may persist for several days. Measurement of plasma digoxin concentration helps identify digoxin toxicity or under-treatment (Box 18.41).

**Therapeutic procedures**

**External defibrillation and cardioversion**

The heart can be completely depolarised by passing a sufficiently large electrical current through it from an external source. This will interrupt any arrhythmia and produce a brief period of asystole that is usually followed by the resumption of sinus rhythm. Defibrillators deliver a DC, high-energy, short-duration shock via two large electrodes or paddles coated with conducting jelly or a gel pad, positioned over the upper right sternal edge and the apex. Modern units deliver a biphasic shock, during which the shock polarity is reversed mid-shock. This reduces the total shock energy required to depolarise the heart.

**Electrical cardioversion**

This is the termination of an organised rhythm, such as atrial fibrillation orventricular tachycardia, with a synchronised shock, usually under general anaesthesia. The shock is delivered immediately after detection of the R wave because, if it is applied during ventricular repolarisation (on the T wave), it may provoke ventricular fibrillation. High-energy shocks may cause chest wall pain post-procedure, so, if there is no urgency, it is appropriate to begin with a lower-amplitude shock (e.g. 50 joules), going on to larger shocks if necessary. Patients with atrial fibrillation or flutter of more than 48 hours’ duration are at risk of left atrial appendage thrombus, and thus systemic embolism after cardioversion. In such cases, cardioversion should be delayed until effective anticoagulation has been given for at least 4 weeks.

**Defibrillation**

This is the delivery of an unsynchronised shock during a cardiac arrest caused by ventricular fibrillation. The precise timing of the discharge is not important in this situation. In ventricular fibrillation and other emergencies, the energy of the first and second shocks should be 150 joules and thereafter up to 200 joules; there is no need for an anaesthetic, as the patient is unconscious.

**Catheter ablation**

Catheter ablation therapy is the treatment of choice for patients with supraventricular tachycardia or atrial flutter, and is a useful treatment for some patients with atrial fibrillation or ventricular arrhythmias (Fig. 18.59). A series of catheter electrodes are inserted into the heart via the venous system and are used to record the activation sequence of the heart in sinus rhythm, during tachycardia and after pacing manoeuvres. Once the arrhythmia focus or circuit is identified (e.g. an accessory pathway in WPW syndrome), a catheter is used to ablate the culprit tissue using heat (via radiofrequency current) or sometimes by freezing (cryoablation). The procedure takes approximately 1–4 hours and does not require a general anaesthetic. The patient may experience some discomfort during the ablation itself. Serious complications are rare (< 1%) but include inadvertent complete heart block requiring pacemaker implantation, and cardiac tamponade. For many arrhythmias, radiofrequency ablation is very attractive because it offers the prospect of a lifetime cure, thereby eliminating the need for long-term drug therapy.

The technique has revolutionised the management of many arrhythmias and is now the treatment of choice for AVNRT and AV re-entrant (accessory pathway) tachycardias, when it is curative in over 90% of cases. Focal atrial tachycardias and atrial flutter can also be eliminated by radiofrequency ablation, although some patients subsequently experience episodes of atrial fibrillation. The applications of the technique are expanding and it can now be used to treat some forms of ventricular tachycardia. Catheter ablation techniques are also used to prevent atrial fibrillation. This involves ablation at two sites: the ostia of the pulmonary veins, from which ectopic beats may trigger paroxysms of arrhythmia, and in the LA itself, where re-entry circuits maintain atrial fibrillation, once established. This is effective at reducing episodes of atrial fibrillation in
around 70–80% of younger patients with structurally normal hearts, and tends to be reserved for patients with drug-resistant atrial fibrillation.

In patients with permanent atrial fibrillation and poor rate control, in whom drugs are ineffective or are not tolerated, rate control can be achieved by: (i) implantation of a permanent pacemaker, followed by (ii) ablation of the AV node to induce complete AV block and brady-cardia, thus allowing the pacemaker to assume control of the heart rate.

Temporary pacemakers
Temporary pacing involves delivery of an electrical impulse into the heart to initiate tissue depolarisation and to trigger cardiac contraction. This is achieved by inserting a bipolar pacing electrode via the internal jugular, subclavian or femoral vein and positioning it at the apex of the RV, using fluoroscopic imaging. The electrode is connected to an external pacemaker with an adjustable energy output and pacing rate. The ECG of right ventricular pacing is characterised by regular broad QRS complexes with a left bundle branch block pattern. Each complex is immediately preceded by a ‘pacing spike’ (Fig. 18.60). Nearly all pulse generators are used in the ‘demand’ mode, so that the pacemaker will only operate if the heart rate falls below a preset level. Occasionally, temporary atrial or dual-chamber pacing (see below) is used.

Temporary pacing may be indicated in the management of transient AV block and other arrhythmias complicating acute MI or cardiac surgery, to maintain the rhythm in other situations of reversible bradycardia (i.e. due to metabolic disturbance or drug overdose), or as a bridge to permanent pacing. Complications include pneumothorax, brachial plexus or subclavian artery injury, local infection or septicaemia (usually Staphylococcus aureus), and pericarditis. Failure of the system may be due to lead displacement or a progressive increase in the threshold (exit block) caused by tissue oedema. Complication rates increase with time and so a temporary pacing system should ideally not be used for more than 7 days.

Transcutaneous pacing is administered by delivering an electrical stimulus through two large adhesive gel pad electrodes placed over the apex and upper right sternal edge, or over the anterior and posterior chest. It is easy and quick to set up, but causes discomfort because it induces forceful pectoral and intercostal muscle contraction. Modern external cardiac defibrillators often incorporate a transcutaneous pacing system that can be used during an emergency until transvenous pacing is established.

Permanent pacemakers
Permanent pacemakers are small, flat, metal devices that are implanted under the skin, usually in the pectoral area. They contain a battery, a pulse generator, and programmable electronics that allow adjustment of pacing and memory functions. Pacing electrodes (leads) can be placed via the subclavian or cephalic veins into the RV (usually at the apex), the right atrial appendage or, to maintain AV synchrony, both.

Permanent pacemakers are programmed using an external programmer via a wireless telemetry system. Pacing rate, output, timing and other parameters can be adjusted. This allows the device to be set to the optimum settings to suit the patient’s needs. Pacemakers store useful diagnostic data about the patient’s heart rate trends and the occurrence of tachyarrhythmias, such as ventricular tachycardia.

Single-chamber atrial pacing is used in patients with sinoatrial disease without AV block (the pacemaker acts as an external sinus node). Single-chamber ventricular pacing is used in patients with continuous atrial fibrillation and bradycardia. Dual-chamber pacing is most often used in patients with second- or third-degree AV block; here, the atrial electrode is used to detect spontaneous atrial activity and trigger ventricular pacing (see Fig. 18.60), thereby preserving AV synchrony and allowing the ventricular rate to increase, together with the sinus node rate, during exercise and other forms of stress. Dual-chamber pacing has many advantages over ventricular pacing; these include superior haemodynamics, leading to a better effort tolerance, a lower prevalence of atrial arrhythmias in patients with sinoatrial disease, and avoidance of ‘pacemaker syndrome’ (a fall in BP and dizziness precipitated by loss of AV synchrony).

A code is used to signify the pacing mode (Box 18.42). For example, a system that paces the atrium, senses the atrium and is inhibited if it senses spontaneous activity is designated AAI. Most dual-chamber pacemakers are programmed to a mode termed DDD; in this case, ventricular pacing is triggered by a sensed sinus P wave and inhibited by a sensed spontaneous QRS complex. A fourth letter, ‘R’, is added if the pacemaker has a rate response function (e.g. AAIR = atrial demand pacemaker with rate response function). Rate-responsive pacemakers are used in patients with chronotropic incompetence, who are unable to increase their heart rate during exercise. These devices have a sensor that triggers an increase in heart rate in response to movement or increased respiratory rate. The sensitivity of

---

**Fig. 18.60 Dual-chamber pacing.** The first three beats show atrial and ventricular pacing with narrow pacing spikes in front of each P wave and QRS complex. The last four beats show spontaneous P waves with a different morphology and no pacing spike; the pacemaker senses or tracks these P waves and maintains AV synchrony by pacing the ventricle after an appropriate interval.

---

### 18.42 International generic pacemaker code

<table>
<thead>
<tr>
<th>Chamber paced</th>
<th>Chamber sensed</th>
<th>Response to sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>O = none</td>
<td>O = none</td>
<td>O = none</td>
</tr>
<tr>
<td>A = atrium</td>
<td>A = atrium</td>
<td>T = triggered</td>
</tr>
<tr>
<td>V = ventricle</td>
<td>V = ventricle</td>
<td>I = inhibited</td>
</tr>
<tr>
<td>D = both</td>
<td>D = both</td>
<td>D = both</td>
</tr>
</tbody>
</table>
the sensor is programmable, as is the maximum paced heart rate.

Early complications of permanent pacing include pneumothorax, cardiac tamponade, infection and lead displacement. Late complications include infection (which usually necessitates removing the pacing system), erosion of the generator or lead, chronic pain related to the implant site, and lead fracture due to mechanical fatigue.

**Implantable cardiac defibrillators**

In addition to the functions of a permanent pacemaker, implantable cardiac defibrillators (ICDs) can also detect and terminate life-threatening ventricular tachyarrhythmias. ICDs are larger than pacemakers mainly because of the need for a large battery and capacitor to enable cardioversion or defibrillation. ICD leads are similar to pacing leads but have one or two shock coils along the length of the lead, used for delivering defibrillation. ICDs treat ventricular tachyarrhythmias using overdive pacing, cardioversion or defibrillation. They are implanted in a similar manner to pacemakers and carry a similar risk of complications. In addition, patients can be prone to psychological problems and anxiety, particularly if they have experienced repeated shocks from their device.

The evidence-based indications for ICD implantation are shown in Box 18.43. These can be divided into ‘secondary prevention’ indications, when patients have already had a potentially life-threatening ventricular arrhythmia, and ‘primary prevention’ indications, when patients are considered to be at significant future risk of arrhythmic death. ICDs may be used prophylactically in selected patients with inherited conditions associated with a high risk of sudden cardiac death, such as long QT syndrome (p. 571), hypertrophic cardiomyopathy and arrhythmogenic right ventricular dysplasia (pp. 637 and 638). ICD treatment is expensive and so the indications for which the devices are routinely implanted depend on the health-care resources available.

<table>
<thead>
<tr>
<th>18.43 Key indications for ICD therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary prevention</strong></td>
</tr>
<tr>
<td>• After MI, if LV ejection fraction &lt; 30%</td>
</tr>
<tr>
<td>• Mild to moderate symptomatic heart failure on optimal drug therapy, with LV ejection fraction &lt; 35%</td>
</tr>
<tr>
<td><strong>Secondary prevention</strong></td>
</tr>
<tr>
<td>• Survivors of ventricular fibrillation or ventricular tachycardia cardiac arrest not due to transient or reversible cause</td>
</tr>
<tr>
<td>• Ventricular tachycardia with haemodynamic compromise or significant LV impairment (LV ejection fraction &lt; 35%)</td>
</tr>
</tbody>
</table>

**Cardiac resynchronisation therapy**

Cardiac resynchronisation therapy (CRT) is a treatment for selected patients with heart failure, in whom cardiac function is further impaired by the presence of left bundle branch block. This conduction defect is associated with left ventricular dys-synchrony (poorly coordinated left ventricular contraction) and can aggravate heart failure in susceptible patients. CRT systems have an additional lead that is placed via the coronary sinus into one of the veins on the epicardial surface of the LV (see Fig. 18.28, p. 553). Simultaneous septal and left ventricular epicardial pacing resynchronises left ventricular contraction. These devices improve effort tolerance, reduce heart failure symptoms (Box 18.44), and are more effective in patients in sinus rhythm than in those with atrial fibrillation. Most devices are also defibrillators (CRT-D) because many patients with heart failure are predisposed to ventricular arrhythmias. CRT-pacemakers (CRT-P) are used in patients considered to be at relatively low risk of these arrhythmias.

**EBM 18.44 Cardiac resynchronisation therapy (CRT) for heart failure**

‘CRT improves symptoms and quality of life, and reduces mortality in patients with moderate to severe (NYHA class III–IV) heart failure who are in sinus rhythm, with left bundle branch block and LV ejection fraction ≤ 35%. CRT also prevents heart failure progression in similar patients with mild (NYHA class I–II) heart failure symptoms.’


**ATHEROSCLEROSIS**

Atherosclerosis can affect any artery in the body. When it occurs in the heart, it may cause angina, MI and sudden death; in the brain, stroke and transient ischaemic attack (TIA); and in the limbs, claudication and critical limb ischaemia. Occult coronary artery disease is common in those who present with other forms of atherosclerotic vascular disease, such as intermittent claudication or stroke, and is an important cause of morbidity and mortality in these patients.

**Pathophysiology**

Atherosclerosis is a progressive inflammatory disorder of the arterial wall that is characterised by focal lipid-rich deposits of atheroma that remain clinically silent until they become large enough to impair tissue perfusion, or until ulceration and disruption of the lesion result in thrombotic occlusion or distal embolisation of the vessel. These mechanisms are common to the entire vascular tree, and the clinical manifestations of atherosclerosis depend upon the site of the lesion and the vulnerability of the organ supplied.

Atherosclerosis begins early in life. Abnormalities of arterial function have been detected among high-risk children and adolescents, such as cigarette smokers and those with familial hyperlipidaemia or hypertension. Early lesions have been found in the arteries of victims of accidental death in the second and third decades of life. Nevertheless, clinical manifestations often do not appear until the sixth, seventh or eighth decade.

**Early atherosclerosis**

Fatty streaks tend to occur at sites of altered arterial shear stress, such as bifurcations, and are associated with abnormal endothelial function. They develop when inflammatory cells, predominantly monocytes, bind to
Cardiovascular disease

Receptors expressed by endothelial cells, migrate into the intima, take up oxidised low-density lipoprotein (LDL) particles and become lipid-laden macrophages or foam cells. Extracellular lipid pools appear in the intimal space when foam cells die and release their contents (Fig. 18.61). In response to cytokines and growth factors produced by activated macrophages, smooth muscle cells migrate from the media of the arterial wall into the intima, and change from a contractile to a repair phenotype in an attempt to stabilise the atherosclerotic lesion. If this is successful, the lipid core will be covered by smooth muscle cells and matrix, producing a stable atherosclerotic plaque that will remain asymptomatic until it becomes large enough to obstruct arterial flow.

Advanced atherosclerosis

In an established atherosclerotic plaque, macrophages mediate inflammation and smooth muscle cells promote repair. If inflammation predominates, the plaque becomes active or unstable and may be complicated by ulceration and thrombosis. Cytokines, such as interleukin-1, tumour necrosis factor-alpha, interferon-gamma, platelet-derived growth factors, and matrix metalloproteinases are released by activated macrophages; they cause the intimal smooth muscle cells overlying the plaque to become senescent and collagen cross-struts within the plaque to degrade. This results in thinning of the protective fibrous cap, making the lesion vulnerable to mechanical stress that ultimately causes erosion, fissuring or rupture of the plaque surface (see Fig. 18.61). Any breach in the integrity of the plaque will expose its contents to blood and will trigger platelet aggregation and thrombosis that extend into the atheromatous plaque and the arterial lumen. This type of plaque event may cause partial or complete obstruction at the site of the lesion or distal embolisation resulting in infarction or ischaemia of the affected organ. This common mechanism underlies many of the acute manifestations of atherosclerotic vascular disease, such as acute lower limb ischaemia, MI and stroke.

The number and complexity of arterial plaques increase with age and with risk factors (see below) but the rate of progression of individual plaques is variable. There is a complex and dynamic interaction between mechanical wall stress and atherosclerotic lesions. ‘Vulnerable’ plaques are characterised by a lipid-rich core, a thin fibrocellular cap, speckled calcification and an increase in inflammatory cells that release specific enzymes to degrade matrix proteins. In contrast, stable plaques are typified by a small lipid pool, a thick fibrous cap, heavy calcification and plentiful collagenous cross-struts. Fissuring or rupture tends to occur at sites of maximal mechanical stress, particularly the margins of an eccentric plaque, and may be triggered by a surge in BP, such as during exercise or emotional stress. Surprisingly, the majority of plaque events are subclinical and heal spontaneously, although this may allow thrombus to be incorporated into the lesion, producing plaque growth and further obstruction to flow.

Atherosclerosis may induce complex changes in the media that lead to arterial remodelling. Some arterial segments may slowly constrict (negative remodelling),

---

**Fig. 18.61 The six stages of atherosclerosis.** American Heart Association classification. From Stary, et al. 1995 – see p. 641.
whilst others may gradually enlarge (positive remodeling). These changes are important because they may amplify or minimise the degree to which atheroma encroaches into the arterial lumen.

**Risk factors**

The role and relative importance of many risk factors for the development of coronary, peripheral and cerebrovascular disease have been defined in experimental animal studies, epidemiological studies and clinical interventional trials. Key factors have emerged but do not explain all the risk, and unknown factors may account for up to 40% of the variation in risk from one person to the next.

The impact of genetic risk is illustrated by twin studies; a monozygotic twin of an affected individual has an eightfold increased risk and a dizygotic twin a fourfold increased risk of dying from coronary artery disease, compared to the general population.

The effect of risk factors is multiplicative rather than additive. People with a combination of risk factors are at greatest risk and so assessment should take account of all identifiable risk factors. It is important to distinguish between relative risk (the proportional increase in risk) and absolute risk (the actual chance of an event). Thus, a man of 35 years with a plasma cholesterol of 7 mmol/L (approximately 170 mg/dL), who smokes 40 cigarettes a day, is relatively much more likely to die from coronary disease within the next decade than a non-smoking woman of the same age with a normal cholesterol, but the absolute likelihood of his dying during this time is still small (high relative risk, low absolute risk).

- **Age and sex.** Age is the most powerful independent risk factor for atherosclerosis. Pre-menopausal women have lower rates of disease than men, although this sex difference disappears after the menopause. However, hormone replacement therapy has no role in the primary or secondary prevention of coronary artery disease, and isolated oestrogen therapy may cause an increased cardiovascular event rate.

- **Family history.** Atherosclerotic vascular disease often runs in families, due to a combination of shared genetic, environmental and lifestyle factors. The most common inherited risk characteristics (hypertension, hyperlipidaemia, diabetes mellitus) are polygenic. A ‘positive’ family history is present when clinical problems in first-degree relatives occur at relatively young age, such as below 50 years for men and below 55 years for women.

- **Smoking.** This is probably the most important avoidable cause of atherosclerotic vascular disease. There is a strong, consistent and dose-linked relationship between cigarette smoking and coronary artery disease, especially in younger (< 70 years) individuals.

- **Hypertension** (see below). The incidence of atherosclerosis increases as BP rises, and this excess risk is related to both systolic and diastolic BP, as well as pulse pressure. Antihypertensive therapy reduces cardiovascular mortality, stroke and heart failure.

- **Hypercholesterolaemia** (p. 453). Risk rises with increasing serum cholesterol concentrations. Lowering serum total and LDL cholesterol concentrations reduces the risk of cardiovascular events, including death, MI, stroke and coronary revascularisation.

- **Diabetes mellitus.** This is a potent risk factor for all forms of atherosclerosis and is often associated with diffuse disease that is difficult to treat. Insulin resistance (normal glucose homeostasis with high levels of insulin) is associated with obesity and physical inactivity, and is a risk factor for coronary artery disease (p. 805). Glucose intolerance accounts for a major part of the high incidence of ischaemic heart disease in certain ethnic groups, e.g. South Asians.

- **Haemostatic factors.** Platelet activation and high plasma fibrinogen concentrations are associated with an increased risk of coronary thrombosis. Antiphospholipid antibodies are associated with recurrent arterial thromboses (p. 1055).

- **Physical activity.** Physical inactivity roughly doubles the risk of coronary artery disease and is a major risk factor for stroke. Regular exercise (brisk walking, cycling or swimming for 20 minutes two or three times a week) has a protective effect that may be related to increased serum high-density lipoprotein (HDL) cholesterol concentrations, lower BP, and collateral vessel development.

- **Obesity** (p. 115). Obesity, particularly if central or truncal, is an independent risk factor, although it is often associated with other adverse factors, such as hypertension, diabetes mellitus and physical inactivity.

- **Alcohol.** Alcohol consumption is associated with reduced rates of coronary artery disease. Excess alcohol consumption is associated with hypertension and cerebrovascular disease.

- **Other dietary factors.** Dietary deficiencies in fresh fruit, vegetables and polyunsaturated fatty acids are associated with an increased risk of cardiovascular disease. The introduction of a Mediterranean-style diet reduces cardiovascular events. However, dietary supplements, such as vitamin C and E, beta-carotene, folate and fish oils, do not reduce cardiovascular events and, in some cases, have been associated with harm.

- **Personality.** Certain personality traits are associated with an increased risk of coronary disease. Nevertheless, there is little or no evidence to support the popular belief that stress is a major cause of coronary artery disease.

- **Social deprivation.** Health inequalities have a major influence on cardiovascular disease. The impact of established risk factors is amplified in patients who are socially deprived and current guidelines recommend that treatment thresholds should be lowered for them.

**Primary prevention**

Two complementary strategies can be used to prevent atherosclerosis in apparently healthy but at-risk individuals: population and targeted strategies.

The population strategy aims to modify the risk factors of the whole population through diet and lifestyle advice, on the basis that even a small reduction in
smoking or average cholesterol, or modification of exercise and diet will produce worthwhile benefits (Box 18.45). Some risk factors for atheroma, such as obesity and smoking, are also associated with a higher risk of other diseases and should be actively discouraged through public health measures. Legislation restricting smoking in public places is associated with reductions in rates of MI.

The targeted strategy aims to identify and treat high-risk individuals, who usually have a combination of risk factors and can be identified by using composite scoring systems (Fig. 18.62). It is important to consider the absolute risk of atheromatous cardiovascular disease that an individual is facing before contemplating specific antihypertensive or lipid-lowering therapy because this will help to determine whether the possible benefits of intervention are likely to outweigh the expense, inconvenience and possible side-effects of treatment. For example, a 65-year-old man with an average BP of 150/90 mmHg, who smokes and has diabetes mellitus, a total:HDL cholesterol ratio of 8 and left ventricular hypertrophy on ECG, will have a 10-year risk of coronary artery disease of 68% and a 10-year risk of any cardiovascular event of 90%. Lowering his cholesterol will reduce these risks by 30% and lowering his BP will produce a further 20% reduction; both would obviously be worthwhile. Conversely, a 55-year-old woman who has an identical BP, is a non-smoker, does not have diabetes mellitus and has a normal ECG and a total:HDL cholesterol ratio of 6 has a much better outlook, with a predicted coronary artery disease risk of 14% and a 10-year risk of any cardiovascular event of 90%. Lowering her cholesterol and BP would also reduce risk by 30% and 20% respectively, the value of either or both treatments is questionable.

Secondary prevention

Patients who already have evidence of atheromatous vascular disease are at high risk of future cardiovascular events and should be offered treatments and measures to improve their outlook. The energetic correction of...
modifiable risk factors, particularly smoking, hypertension and hypercholesterolaemia, is particularly important because the absolute risk of further vascular events is high. All patients with coronary artery disease should be given statin therapy, irrespective of their serum cholesterol concentration (Box 18.46). BP should be treated to a target of 140/85 mmHg or lower (p. 610). Aspirin and ACE inhibitors are of benefit in patients with evidence of vascular disease (Boxes 18.47 and 18.48). Beta-blockers benefit patients with a history of MI (see below) or heart failure.

Many clinical events offer an unrivalled opportunity to introduce effective secondary preventive measures; patients who have just survived an MI or undergone bypass surgery are usually keen to help themselves and may be particularly receptive to lifestyle advice, such as dietary modification and smoking cessation.

**EBM** 18.46 Use of statins in prevention of atherosclerotic disease

**Primary prevention**

‘In patients without evidence of coronary disease but with high serum cholesterol concentrations, cholesterol-lowering with statins does not lower mortality but does prevent coronary events (angina and MI).’

**Secondary prevention**

‘In patients with established coronary disease (MI or angina), statin therapy can safely reduce the 5-year incidence of all-cause death, as well as major coronary events, coronary revascularisation and stroke. Benefit depends on the overall risk of the study population but the NNT₉ for 5 years to prevent 1 death ranges from 10 to 90.’


For further information: [www.sign.ac.uk/guidelines/fulltext/93-97/index.html](http://www.sign.ac.uk/guidelines/fulltext/93-97/index.html)

**EBM** 18.47 ACE inhibitors and secondary prevention of atherosclerotic disease

‘ACE inhibitor therapy reduces the risk of death, MI and stroke in patients with atherosclerotic vascular disease without apparent left ventricular systolic dysfunction or heart failure. NNT₉ to avoid 1 event over 4 years ranges from 6 to 50, depending upon the level of cardiovascular risk.’


**EBM** 18.48 Aspirin and secondary prevention in atherosclerotic vascular disease

‘In patients with established coronary artery disease, peripheral vascular disease or thrombotic stroke, aspirin is effective in reducing morbidity and mortality (non-fatal MI, stroke and cardiovascular death). In patients at high risk of future vascular events, the overall risk reduction is 22%.’


For further information: [www.clinicalevidence.org](http://www.clinicalevidence.org)

---

**CORONARY ARTERY DISEASE**

Coronary artery disease (CAD) is the most common form of heart disease and the single most important cause of premature death in Europe, the Baltic states, Russia, North and South America, Australia and New Zealand. By 2020, it is estimated that it will be the major cause of death in all regions of the world.

In the UK, 1 in 3 men and 1 in 4 women die from CAD, an estimated 330000 people have a myocardial infarct each year, and approximately 1.3 million people have angina. The death rates from CAD in the UK are amongst the highest in Western Europe (more than 140000 people) but are falling, particularly in younger age groups; in the last 10 years, CAD mortality has fallen by 42% among UK men and women aged 16–64. However, in Eastern Europe and much of Asia, the rates of CAD are rapidly rising.

Disease of the coronary arteries is almost always due to atheroma and its complications, particularly thrombosis (Box 18.49). Occasionally, the coronary arteries are involved in other disorders such as aortitis, polyarteritis and other connective tissue disorders.

**EBM** 18.49 Coronary artery disease: clinical manifestations and pathology

<table>
<thead>
<tr>
<th>Clinical problem</th>
<th>Pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable angina</td>
<td>Ischaemia due to fixed atheromatous stenosis of one or more coronary arteries</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>Ischaemia caused by dynamic obstruction of a coronary artery due to plaque rupture or erosion with superimposed thrombosis</td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>Myocardial necrosis caused by acute occlusion of a coronary artery due to plaque rupture or erosion with superimposed thrombosis</td>
</tr>
<tr>
<td>Heart failure</td>
<td>Myocardial dysfunction due to infarction or ischaemia</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>Altered conduction due to ischaemia or infarction</td>
</tr>
<tr>
<td>Sudden death</td>
<td>Ventricular arrhythmia, asystole or massive myocardial infarction</td>
</tr>
</tbody>
</table>

---

**Stable angina**

Angina pectoris is the symptom complex caused by transient myocardial ischaemia and constitutes a clinical syndrome rather than a disease. It may occur whenever there is an imbalance between myocardial oxygen supply and demand (Box 18.50). Coronary atheroma is by far the most common cause of angina, although the symptom may be a manifestation of other forms of heart disease, particularly aortic valve disease and hypertrophic cardiomyopathy.

**Clinical features**

The history is the most important factor in making the diagnosis (p. 539). Stable angina is characterised by central chest pain, discomfort or breathlessness that is
Cardiovascular disease

18.50 Factors influencing myocardial oxygen supply and demand

<table>
<thead>
<tr>
<th>Oxygen demand: cardiac work</th>
<th>Oxygen supply: coronary blood flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Heart rate</td>
<td>• Duration of diastole</td>
</tr>
<tr>
<td>• BP</td>
<td>• Coronary vasomotor tone</td>
</tr>
<tr>
<td>• Myocardial contractility</td>
<td>• Oxygenation</td>
</tr>
<tr>
<td>• Left ventricular hypertrophy</td>
<td>Haemoglobin</td>
</tr>
<tr>
<td>• Valve disease, e.g. aortic stenosis</td>
<td>Oxygen saturation</td>
</tr>
</tbody>
</table>

N.B. Coronary blood flow occurs mainly in diastole.

18.51 Activities precipitating angina

Common
• Physical exertion
• Cold exposure
• Heavy meals
• Intense emotion

Uncommon
• Lying flat (decubitus angina)
• Vivid dreams (nocturnal angina)

precipitated by exertion or other forms of stress (Box 18.51), and is promptly relieved by rest (see Figs 18.17 and 18.18, pp. 540 and 541). Some patients find the discomfort comes when they start walking, and that later it does not return despite greater effort ('warm-up angina').

Physical examination is frequently unremarkable but should include a careful search for evidence of valve disease (particularly aortic), important risk factors (e.g. hypertension, diabetes mellitus), left ventricular dysfunction (cardiomegaly, gallop rhythm), other manifestations of arterial disease (carotid bruits, peripheral vascular disease) and unrelated conditions that may exacerbate angina (anaemia, thyrotoxicosis).

Investigations

Resting ECG

The ECG may show evidence of previous MI but is often normal, even in patients with severe coronary artery disease. Occasionally, there is T-wave flattening or inversion in some leads, providing non-specific evidence of myocardial ischaemia or damage. The most convincing ECG evidence of myocardial ischaemia is the demonstration of reversible ST segment depression or elevation, with or without T-wave inversion, at the time the patient is experiencing symptoms (whether spontaneous or induced by exercise testing).

Exercise ECG

An exercise tolerance test (ETT) is usually performed using a standard treadmill or bicycle ergometer protocol (p. 534) while monitoring the patient’s ECG, BP and general condition. Planar or down-sloping ST segment depression of 1 mm or more is indicative of ischaemia (Fig. 18.63). Up-sloping ST depression is less specific and often occurs in normal individuals.

Exercise testing is also a useful means of assessing the severity of coronary disease and identifying high-risk individuals (Box 18.52). For example, the amount of exercise that can be tolerated and the extent and degree of any ST segment change (Fig. 18.64) provide a useful guide to the likely extent of coronary disease. Exercise testing is not infallible and may produce false-positive results in the presence of digoxin therapy, left ventricular hypertrophy, bundle branch block or WPW syndrome. The predictive accuracy of exercise testing is

Fig. 18.63 Forms of exercise-induced ST depression. [A] Planar ST depression is usually indicative of myocardial ischaemia. [B] Down-sloping depression also usually indicates myocardial ischaemia. [C] Up-sloping depression may be a normal finding.

Fig. 18.64 A positive exercise test (chest leads only). The resting 12-lead ECG shows some minor T-wave changes in the inferolateral leads but is otherwise normal. After 3 minutes’ exercise on a treadmill, there is marked planar ST depression in leads V4 and V5 (right offset). Subsequent coronary angiography revealed critical three-vessel coronary artery disease.

18.52 Risk stratification in stable angina

<table>
<thead>
<tr>
<th>High risk</th>
<th>Low risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-infarct angina</td>
<td>Predictable exertional angina</td>
</tr>
<tr>
<td>Poor effort tolerance</td>
<td>Good effort tolerance</td>
</tr>
<tr>
<td>Ischaemia at low workload</td>
<td>Ischaemia only at high workload</td>
</tr>
<tr>
<td>Left main or three-vessel disease</td>
<td>Single-vessel or two-vessel disease</td>
</tr>
<tr>
<td>Poor LV function</td>
<td>Good LV function</td>
</tr>
</tbody>
</table>

N.B. Patients may fall between these categories.
lower in women than in men. The test should be classed as inconclusive (rather than negative) if the patient cannot achieve an adequate level of exercise because of locomotor or other non-cardiac problems.

**Other forms of stress testing**

- **Myocardial perfusion scanning.** This may be helpful in the evaluation of patients with an equivocal or uninterpretable exercise test and those who are unable to exercise (p. 539). It entails obtaining scintiscans of the myocardium at rest and during stress (either exercise testing or pharmacological stress, such as a controlled infusion of dobutamine), after the administration of an intravenous radioactive isotope, such as $^{99m}$technetium tetrofosmin. Thallium and tetrofosmin are taken up by viable perfused myocardium. A perfusion defect present during stress but not at rest provides evidence of reversible myocardial ischaemia (Fig. 18.65), whereas a persistent perfusion defect seen during both phases of the study is usually indicative of previous MI.
- **Stress echocardiography.** This is an alternative to myocardial perfusion scanning and can achieve similar predictive accuracy. It uses transthoracic echocardiography to identify ischaemic segments of myocardium and areas of infarction (p. 537). The former characteristically exhibit reversible defects in contractility during exercise or pharmacological stress, and the latter do not contract at rest or during stress.

**Coronary arteriography**

This provides detailed anatomical information about the extent and nature of coronary artery disease (see Fig. 18.15, p. 538), and is usually performed with a view to coronary artery bypass graft (CABG) surgery or percutaneous coronary intervention (PCI) (pp. 587 and 588). In some patients, diagnostic coronary angiography may be indicated when non-invasive tests have failed to establish the cause of atypical chest pain. The procedure is performed under local anaesthesia and requires specialised radiological equipment, cardiac monitoring and an experienced operating team.

**Management: general measures**

The management of angina pectoris involves:

- a careful assessment of the likely extent and severity of arterial disease
- the identification and control of risk factors such as smoking, hypertension and hyperlipidaemia
- the use of measures to control symptoms
- the identification of high-risk patients for treatment to improve life expectancy.

Symptoms alone are a poor guide to the extent of coronary artery disease. Stress testing is therefore advisable in all patients who are potential candidates for revascularisation. An algorithm for the investigation and treatment of patients with stable angina is shown in Figure 18.66.

Management should start with a careful explanation of the problem and a discussion of the potential lifestyle and medical interventions that may relieve symptoms and improve prognosis (Box 18.53). Anxiety and misconceptions often contribute to disability; for example, some patients avoid all forms of exertion because they believe that each attack of angina is a ‘mini heart attack’ that results in permanent damage. Effective management of these psychological factors can make a huge difference to the patient’s quality of life.

**Antiplatelet therapy**

Low-dose (75 mg) aspirin reduces the risk of adverse events such as MI and should be prescribed for all patients with coronary artery disease indefinitely (see Box 18.48). Clopidogrel (75 mg daily) is an equally effective antiplatelet agent that can be prescribed if aspirin causes troublesome dyspepsia or other side-effects.

---

**18.53 Advice to patients with stable angina**

- Do not smoke
- Aim for ideal body weight
- Take regular exercise (exercise up to, but not beyond, the point of chest discomfort is beneficial and may promote collateral vessels)
- Avoid severe unaccustomed exertion, and vigorous exercise after a heavy meal or in very cold weather
- Take sublingual nitrate before undertaking exertion that may induce angina
Anti-anginal drug treatment

Five groups of drug are used to help relieve or prevent the symptoms of angina: nitrates, β-blockers, calcium antagonists, potassium channel activators and an I channel antagonist.

Nitrates

These drugs act directly on vascular smooth muscle to produce venous and arteriolar dilatation. Their beneficial effects are due to a reduction in myocardial oxygen demand (lower preload and afterload) and an increase in myocardial oxygen supply (coronary vasodilatation). Sublingual glyceryl trinitrate (GTN), administered from a metered-dose aerosol (400 µg per spray) or as a tablet (300 or 500 µg), will relieve an attack of angina in 2–3 minutes. Side-effects include headache, symptomatic hypotension and, rarely, syncope.

Patients should be encouraged to use the drug prophylactically before taking exercise that is liable to provoke symptoms. Sublingual GTN has a short duration of action (Box 18.54); however, a variety of alternative nitrate preparations can provide a more prolonged therapeutic effect. GTN can be given transcutaneously as a patch (5–10 mg daily), or as a slow-release buccal tablet (1–5 mg 4 times daily). GTN undergoes extensive first-pass metabolism in the liver and is ineffective when swallowed. Other nitrates, such as isosorbide dinitrate (10–20 mg 3 times daily) and isosorbide mononitrate (20–60 mg once or twice daily), can be given by mouth. Headache is common but tends to diminish if the patient perseveres with the treatment. Continuous nitrate therapy can cause pharmacological tolerance. This can be avoided by a 6–8-hour nitrate-free period, best achieved at night when the patient is inactive. If nocturnal angina is a predominant symptom, long-acting nitrates can be given at the end of the day.

β-blockers

These lower myocardial oxygen demand by reducing heart rate, BP and myocardial contractility, but they may provoke bronchospasm in patients with asthma. The properties and side-effects of β-blockers are discussed on page 599.

In theory, non-selective β-blockers may aggravate coronary vasospasm by blocking the coronary artery β2-adrenoceptors and so a once-daily cardioselective preparation is used (e.g. slow-release metoprolol 50–200 mg daily, bisoprolol 5–15 mg daily). Beta-blockers should not be withdrawn abruptly as rebound effects may precipitate dangerous arrhythmias, worsening angina or MI: the β-blocker withdrawal syndrome.

Calcium channel antagonists

These drugs inhibit the slow inward current caused by the entry of extracellular calcium through the cell membrane of excitable cells, particularly cardiac and arteriolar smooth muscle, and lower myocardial oxygen demand by reducing BP and myocardial contractility.

Dihydropyridine calcium antagonists, such as nifedipine and nicardipine, often cause a reflex tachycardia. This may be counterproductive and it is best to use them in combination with a β-blocker. In contrast, verapamil and diltiazem are particularly suitable for patients who are not receiving a β-blocker (e.g. those with airways obstruction) because they slow SA node firing, inhibit conduction through the AV node and
tend to cause a bradycardia. Calcium channel antagonists reduce myocardial contractility and can aggravate or precipitate heart failure. Other unwanted effects include peripheral oedema, flushing, headache and dizziness (Box 18.55).

**Potassium channel activators**

These have arterial and venous dilating properties but do not exhibit the tolerance seen with nitrates. Nicorandil (10–30 mg twice daily orally) is the only drug in this class currently available for clinical use.

**β-blocker, and rate-limiting calcium antagonists, it does not have other cardiovascular effects. It appears to be safe to use in patients with heart failure.**

Although each of these anti-anginal drugs is superior to placebo in relieving the symptoms of angina, there is little evidence that one group is more effective than another. It is conventional to start therapy with low-dose aspirin, a statin, sublingual GTN and a β-blocker, and then add a calcium channel antagonist or a long-acting nitrates later, if needed. The goal is the control of angina with minimum side-effects and the simplest possible drug regimen. There is little evidence that prescribing multiple anti-anginal drugs is of benefit, and revascularisation should be considered if an appropriate combination of two or more drugs fails to achieve an acceptable symptomatic response.

**Invasive treatment**

**Percutaneous coronary intervention**

Percutaneous coronary intervention (PCI) is performed by passing a fine guidewire across a coronary stenosis under radiographic control and using it to position a balloon, which is then inflated to dilate the stenosis (Fig. 18.67). A coronary stent is a piece of coated metallic ‘scaffolding’ that can be deployed on a balloon and used to maximise and maintain dilatation of a stenosed vessel. The routine use of stents in appropriate vessels reduces both acute complications and the incidence of clinically important re-stenosis (Box 18.56 and Fig. 18.76, p. 595).

PCI provides an effective symptomatic treatment but definitive evidence that it improves survival in patients with chronic stable angina is lacking. It is mainly used in single- or two-vessel disease. Stenoses in bypass grafts can be dilated, as well as those in the native coronary arteries. The technique is often used to provide palliative therapy for patients with recurrent angina after CABG. Coronary surgery is usually the preferred option in patients with three-vessel or left main stem disease, although recent trials have demonstrated that PCI is also feasible in such patients.

The main acute complications of PCI are occlusion of the target vessel or a side branch by thrombus or a loose flap of intima (coronary artery dissection), and consequent myocardial damage. This occurs in about 2–5% of procedures and can often be corrected by deploying a stent; however, emergency CABG is sometimes required. Minor myocardial damage, as indicated by elevation of sensitive intracellular markers (troponins, p. 535), occurs

---

**Box 18.55 Calcium channel antagonists used for the treatment of angina**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Dose</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nifedipine</td>
<td>5–20 mg; 3 times daily*</td>
<td>May cause marked tachycardia</td>
</tr>
<tr>
<td>Nicardipine</td>
<td>20–40 mg; 3 times daily</td>
<td>May cause less myocardial depression than the other calcium antagonists</td>
</tr>
<tr>
<td>Amlodipine</td>
<td>2.5–10 mg daily</td>
<td>Ultra-long-acting</td>
</tr>
<tr>
<td>Verapamil</td>
<td>40–80 mg; 3 times daily*</td>
<td>Commonly causes constipation; useful anti-arrhythmic properties (p. 576)</td>
</tr>
<tr>
<td>Diltiazem</td>
<td>60–120 mg; 3 times daily*</td>
<td>Similar anti-arrhythmic properties to verapamil</td>
</tr>
</tbody>
</table>

*Once- or twice-daily slow-release preparations are available.

---

**Fig. 18.67 Vascular and valvular balloon dilatations.**
in up to 10% of cases. The main long-term complication of PCI is re-stenosis (Box 18.57), in up to one-third of cases. This is due to a combination of elastic recoil and smooth muscle proliferation (neo-intimal hyperplasia) and tends to occur within 3 months. Stenting substantially reduces the risk of re-stenosis, probably because it allows the operator to achieve more complete dilatation in the first place. Drug-eluting stents reduce this risk even further by allowing an antiproliferative drug, e.g. sirolimus or paclitaxel, to elute slowly from the coating and prevent neo-intimal hyperplasia and in-stent re-stenosis. There is an increased risk of late stent thrombosis with drug-eluting stents, although the absolute risk is small (< 0.5%). Recurrent angina (affecting up to 15–20% of patients receiving an intracoronary stent at 6 months) may require further PCI or bypass grafting.

The risk of complications and the likely success of the procedure are closely related to the morphology of the stenoses, the experience of the operator and the presence of important comorbidity, e.g. diabetes, peripheral arterial disease. A good outcome is less likely if the target lesion is complex, long, eccentric or calcified, lies on a bend or within a tortuous vessel, involves a branch or contains acute thrombus.

In combination with aspirin and heparin, adjunctive therapy with potent platelet inhibitors, such as clopidogrel or glycoprotein IIb/IIIa receptor antagonists, improves the outcome of PCI, with lower short- and long-term rates of death and MI.

**Coronary artery bypass grafting**

The internal mammary arteries, radial arteries or reversed segments of the patient’s own saphenous vein can be used to bypass coronary artery stenoses (Fig. 18.68). This usually involves major surgery under cardiopulmonary bypass but, in some cases, grafts can be applied to the beating heart: ‘off-pump’ surgery. The operative mortality is approximately 1.5% but risks are higher in elderly patients, those with poor left ventricular function and those with significant comorbidity, such as renal failure.

Approximately 90% of patients are free of angina 1 year after CABG surgery, but fewer than 60% of patients are asymptomatic after 5 or more years. Early postoperative angina is usually due to graft failure arising from technical problems during the operation, or poor ‘run-off’ due to disease in the distal native coronary vessels. Late recurrence of angina may be due to progressive disease in the native coronary arteries or graft degeneration. Fewer than 50% of vein grafts are patent 10 years after surgery. However, arterial grafts have a much better long-term patency rate, with more than 80% of internal mammary artery grafts patent at 10 years. This has led many surgeons to consider total arterial revascularisation during CABG surgery. Aspirin (75–150 mg daily) and clopidogrel (75 mg daily) both improve graft patency, and one or other should be prescribed indefinitely, if well tolerated. Intensive lipid-lowering therapy slows the progression of disease in the native coronary arteries and bypass grafts, and reduces clinical cardiovascular events. There is substantial excess cardiovascular morbidity and mortality in patients who continue to smoke after bypass grafting. Persistent smokers are twice as likely to die in the 10 years following surgery than those who give up at surgery.

CABG improves survival in symptomatic patients with left main stem stenosis or three-vessel coronary disease (i.e. involving LAD, CX and right coronary arteries, Box 18.58) or two-vessel disease involving the
proximal LAD coronary artery. Improvement in survival is most marked in those with impaired left ventricular function or positive stress testing prior to surgery and in those who have undergone left internal mammary artery grafting.

Neurological complications are common, with a 1–5% risk of peri-operative stroke. Between 30% and 80% of patients develop short-term cognitive impairment that typically resolves within 6 months. There are also reports of long-term cognitive decline that may be evident in more than 30% of patients at 5 years. PCI and CABG are compared in Boxes 18.59 and 18.60.

**Prognosis**

Symptoms are a poor guide to prognosis; nevertheless, the 5-year mortality of patients with severe angina (NYHA class III or IV, p. 539) is nearly double that of patients with mild symptoms. Exercise testing and other forms of stress testing are much more powerful predictors of mortality; for example, in one study, the 4-year mortality of patients with stable angina and a negative exercise test was 1%, compared to more than 20% in those with a strongly positive test.

In general, the prognosis of coronary artery disease is related to the number of diseased vessels and the degree of left ventricular dysfunction. A patient with single-vessel disease and good left ventricular function has an excellent outlook (5-year survival > 90%), whereas a patient with severe left ventricular dysfunction and extensive three-vessel disease has a poor prognosis (5-year survival < 30%) without revascularisation. Spontaneous symptomatic improvement due to the development of collateral vessels is common.

**Angina with normal coronary arteries**

Approximately 10% of patients who report stable angina on effort will have angiographically normal coronary arteries. Many of these patients are women and the mechanism of their symptoms is often difficult to establish. It is important to review the original diagnosis and explore other potential causes.

**Coronary artery spasm**

Vasospasm in coronary arteries may coexist with atheroma, especially in unstable angina (see below); in less than 1% of cases, vasospasm may occur without angiographically detectable atheroma. This is sometimes known as variant angina, and may be accompanied by spontaneous and transient ST elevation on the ECG (Prinzmetal’s angina). Calcium channel antagonists, nitrates and other coronary vasodilators are the most useful therapeutic agents but may be ineffective.

**Syndrome X**

The constellation of typical angina on effort, objective evidence of myocardial ischaemia on stress testing, and angiographically normal coronary arteries is sometimes known as syndrome X. This disorder is poorly understood but carries a good prognosis and may respond to treatment with anti-anginal therapy.

**Acute coronary syndrome**

Acute coronary syndrome is a term that encompasses both unstable angina and myocardial infarction (MI). It is characterised by new-onset or rapidly worsening angina (crescendo angina), angina on minimal exertion or angina at rest in the absence of myocardial damage. In contrast, MI occurs when symptoms occur at rest and there is evidence of myocardial necrosis, as demonstrated by an elevation in cardiac troponin or creatine kinase-MB isoenzyme (Box 18.61).

An acute coronary syndrome may present as a new phenomenon or against a background of chronic stable angina. The culprit lesion is usually a complex
ulcerated or fissured atheromatous plaque with adherent platelet-rich thrombus and local coronary artery spasm (see Fig. 18.61, p. 580). This is a dynamic process whereby the degree of obstruction may either increase, leading to complete vessel occlusion, or regress due to the effects of platelet disaggregation and endogenous fibrinolysis. In acute MI, occlusive thrombus is almost always present at the site of rupture or erosion of an atheromatous plaque. The thrombus may undergo spontaneous lysis over the course of the next few days, although, by this time, irreversible myocardial damage has occurred. Without treatment, the infarct-related artery remains permanently occluded in 20–30% of patients. The process of infarction progresses over several hours (Fig. 18.69) and most patients present when it is still possible to salvage myocardium and improve outcome.

Clinical features

Pain is the cardinal symptom of an acute coronary syndrome but breathlessness, vomiting and collapse are common features (Box 18.62). The pain occurs in the same sites as angina but is usually more severe and lasts longer; it is often described as a tightness, heaviness or constriction in the chest. In acute MI, the pain can be excruciating, and the patient’s expression and pallor may vividly convey the seriousness of the situation.

Most patients are breathless and, in some, this is the only symptom. Indeed, MI may pass unrecognised. Painless or ‘silent’ MI is particularly common in older patients or those with diabetes mellitus. If syncope...
occurs, it is usually due to an arrhythmia or profound hypotension. Vomiting and sinus bradycardia are often due to vagal stimulation and are particularly common in patients with inferior MI. Nausea and vomiting may also be caused or aggravated by opiates given for pain relief. Sometimes infarction occurs in the absence of physical signs.

Sudden death, from ventricular fibrillation or asystole, may occur immediately and often within the first hour. If the patient survives this most critical stage, the liability to dangerous arrhythmias remains, but diminishes as each hour goes by. It is vital that patients know not to delay calling for help if symptoms occur. The development of cardiac failure reflects the extent of myocardial ischaemia and is the major cause of death in those who survive the first few hours.

**Diagnosis and risk stratification**

The differential diagnosis is wide and includes most causes of central chest pain or collapse (pp. 540 and 554). The assessment of acute chest pain depends heavily on evaluation of the character of the pain and its associated features, evaluation of the ECG, and serial measurements of biochemical markers of cardiac damage, such as troponin I and T. A 12-lead ECG is mandatory and defines the initial triage, management and treatment (see Fig. 18.19, p. 542). Patients with ST-segment elevation or new bundle branch block require emergency reperfusion therapy (see below). In patients with acute coronary syndrome without ST-segment elevation, the ECG may show transient or persistent ST-T wave changes, including ST depression and T-wave inversion.

Approximately 12% of patients will die within 1 month and a fifth within 6 months of the index event. The risk markers that are indicative of an adverse prognosis include recurrent ischaemia, extensive ECG changes at rest or during pain, the release of biochemical markers (creatine kinase or troponin), arrhythmias, recurrent ischaemia and haemodynamic complications (e.g. hypotension, mitral regurgitation) during episodes of ischaemia. Risk stratification is important because it guides the use of more complex pharmacological and interventional treatment (Figs 18.17 and 18.19 (p. 542)).

1. **Find points for each predictive factor**

<table>
<thead>
<tr>
<th>Killip class</th>
<th>Points</th>
<th>SBP (mmHg)</th>
<th>Points</th>
<th>Heart rate (beats/min)</th>
<th>Points</th>
<th>Age (years)</th>
<th>Points</th>
<th>Serum creatinine level (μmol/L)</th>
<th>Points</th>
<th>Other risk factors</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>≤ 80</td>
<td>58</td>
<td>≤ 50</td>
<td>0</td>
<td>≤ 30</td>
<td>0</td>
<td>0–34</td>
<td>1</td>
<td>Cardiac arrest at admission</td>
<td>39</td>
</tr>
<tr>
<td>II</td>
<td>20</td>
<td>80–99</td>
<td>53</td>
<td>50–69</td>
<td>3</td>
<td>30–39</td>
<td>8</td>
<td>35–70</td>
<td>4</td>
<td>ST-segment deviation</td>
<td>28</td>
</tr>
<tr>
<td>III</td>
<td>39</td>
<td>100–119</td>
<td>43</td>
<td>70–89</td>
<td>9</td>
<td>40–49</td>
<td>25</td>
<td>71–105</td>
<td>7</td>
<td>Elevated cardiac enzyme levels</td>
<td>14</td>
</tr>
<tr>
<td>IV</td>
<td>59</td>
<td>120–139</td>
<td>34</td>
<td>90–109</td>
<td>15</td>
<td>50–59</td>
<td>41</td>
<td>106–140</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>140–159</td>
<td>24</td>
<td>110–149</td>
<td>24</td>
<td>60–69</td>
<td>58</td>
<td>141–176</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>160–199</td>
<td>10</td>
<td>150–199</td>
<td>38</td>
<td>70–79</td>
<td>75</td>
<td>177–353</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 200</td>
<td>0</td>
<td>≥ 200</td>
<td>46</td>
<td>80–89</td>
<td>91</td>
<td>≥ 353</td>
<td>28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. **Sum points for all predictive factors**

3. **Look up risk corresponding to total points**

<table>
<thead>
<tr>
<th>Total points</th>
<th>≤ 60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
<th>190</th>
<th>200</th>
<th>210</th>
<th>220</th>
<th>230</th>
<th>240</th>
<th>≤ 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of in-hospital death (%)</td>
<td>≤ 0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1.1</td>
<td>1.6</td>
<td>2.1</td>
<td>2.9</td>
<td>3.9</td>
<td>5.4</td>
<td>7.3</td>
<td>9.8</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>29</td>
<td>36</td>
<td>44</td>
<td>≤ 52</td>
</tr>
</tbody>
</table>

**Examples**

A patient has Killip class II, SBP of 99 mmHg, heart rate of 100 beats/min, is 65 years of age, has a serum creatinine level of 76 μmol/L, did not have a cardiac arrest at admission but did have ST-segment deviation and elevated enzyme levels. His score would be: 20 + 53 + 15 + 58 + 7 + 0 + 28 + 14 = 195. This gives about a 16% risk of having an in-hospital death.

Similarly, a patient with Killip class I, SBP of 80 mmHg, heart rate of 60 beats/min, who is 55 years of age, has a serum creatinine level of 30 μmol/L, and no risk factors would have the following score: 0 + 58 + 3 + 41 + 1 = 103. This gives about a 0.9% risk of having an in-hospital death.

**Fig. 18.70 Risk stratification in the acute coronary syndrome: the GRACE score.** Killip class refers to a categorisation of the severity of heart failure based on easily obtained clinical signs. The main clinical features are as follows: class I = no heart failure; class II = crackles audible halfway up the chest; class III = crackles heard in all the lung fields; class IV = cardiogenic shock (SBP = systolic blood pressure). From SIGN 93 – see p. 641.
**Cardiovascular Disease**

**Investigations**

*Electrocardiography*

The ECG is central to confirming the diagnosis but may be difficult to interpret if there is bundle branch block or previous MI. The initial ECG may be normal or non-diagnostic in one-third of cases. Repeated ECGs are important, especially where the diagnosis is uncertain or the patient has recurrent or persistent symptoms.

The earliest ECG change is usually ST-segment deviation. With proximal occlusion of a major coronary artery, ST-segment elevation (or new bundle branch block) is seen initially, with later diminution in the size of the R wave and, in transmural (full-thickness) infarction, development of a Q wave. Subsequently, the T wave becomes inverted because of a change in ventricular repolarisation; this change persists after the ST segment has returned to normal. These sequential features (Fig. 18.71) are sufficiently reliable for the approximate age of the infarct to be deduced.

In non-ST segment elevation acute coronary syndrome, there is partial occlusion of a major vessel or complete occlusion of a minor vessel, causing unstable angina or partial-thickness (subendocardial) MI. This is usually associated with ST-segment depression and T-wave changes. In the presence of infarction, this may be accompanied by some loss of R waves in the absence of Q waves (Fig. 18.72).

The ECG changes are best seen in the leads that ‘face’ the ischaemic or infarcted area. When there has been anteroseptal infarction, abnormalities are found in one or more leads from V1 to V5, while anterolateral infarction produces changes from V4 to V6 in aVL and in lead I. Inferior infarction is best shown in leads II, III and aVF, while, at the same time, leads I, aVL and the anterior chest leads may show ‘reciprocal’ changes of ST depression (Figs 18.73–18.74). Infarction of the posterior...
Coronary artery disease

wall of the LV does not cause ST elevation or Q waves in the standard leads, but can be diagnosed by the presence of reciprocal changes (ST depression and a tall R wave in leads V1–V4). Some infarctions (especially inferior) also involve the RV. This may be identified by recording from additional leads placed over the right precordium.

Plasma cardiac biomarkers

In unstable angina, there is no detectable rise in cardiac biomarkers or enzymes, and the initial diagnosis is made from the clinical history and ECG only. In contrast, MI causes a rise in the plasma concentration of enzymes and proteins that are normally concentrated within cardiac cells. These biochemical markers are creatine kinase (CK), a more sensitive and cardio-specific isoform of this enzyme (CK-MB), and the cardio-specific proteins, troponins T and I (p. 535). Admission and serial (usually daily) estimations are helpful because it is the change in plasma concentrations of these markers that confirms the diagnosis of MI (Fig. 18.74 and Box 18.61).

CK starts to rise at 4–6 hours, peaks at about 12 hours and falls to normal within 48–72 hours. CK is also present in skeletal muscle, and a modest rise in CK (but not CK-MB) may sometimes be due to an intramuscular injection, vigorous physical exercise or, particularly in older people, a fall. Defibrillation causes significant release of CK but not CK-MB or troponins. The most sensitive markers of myocardial cell damage are the cardiac troponins T and I, which are released within 4–6 hours and remain elevated for up to 2 weeks.

Other blood tests

A leucocytosis is usual, reaching a peak on the first day. The erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) are also elevated.

Chest X-ray

This may demonstrate pulmonary oedema that is not evident on clinical examination (see Fig. 18.25, p. 550). The heart size is often normal but there may be cardiomegaly due to pre-existing myocardial damage.

Echocardiography

This is useful for assessing ventricular function and for detecting important complications, such as mural thrombus, cardiac rupture, ventricular septal defect, mitral regurgitation and pericardial effusion.

Immediate management: the first 12 hours

Patients should be admitted urgently to hospital because there is a significant risk of death or recurrent myocardial ischaemia during the early unstable phase, and appropriate medical therapy can reduce the incidence of these by at least 60%. The essentials of the immediate in-hospital management of acute coronary syndrome are shown in Figure 18.19 (p. 542).

Patients are usually managed in a dedicated cardiac unit, where the necessary expertise, monitoring and resuscitation facilities can be concentrated. If there are no complications, the patient can be mobilised from the second day and discharged after 3–5 days.

Analgesia

Adequate analgesia is essential, not only to relieve distress but also to lower adrenergic drive and thereby reduce vascular resistance, BP, infarct size and susceptibility to ventricular arrhythmias. Intravenous opiates (initially, morphine sulphate 5–10 mg or diamorphine 2.5–5 mg) and antiemetics (initially, metoclopramide 10 mg) should be administered, and titrated by giving repeated small aliquots until the patient is comfortable. Intramuscular injections should be avoided because the clinical effect may be delayed by poor skeletal muscle
perfusion, and a painful haematoma may form following thrombolytic or antithrombotic therapy.

**Antithrombotic therapy**

**Antiplaete therapy**

In patients with acute coronary syndrome, oral administration of 75–325 mg aspirin daily improves survival, with a 25% relative risk reduction in mortality. The first tablet (300 mg) should be given orally within the first 12 hours and therapy should be continued indefinitely if there are no side-effects. In combination with aspirin, 12 hours and therapy should be continued indefinitely as patients with diabetes mellitus or an elevated troponin ischaemia and those at particularly high risk, such as patients with acute coronary syndromes who undergo PCI (Box 18.64), those with recurrent ischaemia and those at particularly high risk, such as patients with diabetes mellitus or an elevated troponin concentration.

**Anticoagulants**

Anticoagulation reduces the risk of thromboembolic complications, and prevents re-infarction in the absence of reperfusion therapy or after successful thrombolysis (Box 18.65). Anticoagulation can be achieved using unfractionated heparin, fractioned (low-molecular-weight) heparin or a pentasaccharide. Comparative clinical trials suggest that the pentasaccharides (subcutaneous fondaparinux 2.5 mg daily) have the best safety and efficacy profile, with low-molecular-weight heparin (subcutaneous enoxaparin 1 mg/kg twice daily) being a reasonable alternative. Anticoagulation should be continued for 8 days or until discharge from hospital or coronary revascularisation. A period of treatment with warfarin should be considered if there is persistent atrial fibrillation or evidence of extensive anterior infarction, or if echocardiography shows mobile mural thrombus, because these patients are at increased risk of systemic thromboembolism.

**Anti-anginal therapy**

Sublingual glyceryl trinitrate (300–500 µg) is a valuable first-aid measure in unstable angina or threatened infarction, and intravenous nitrates (glyceryl trinitrate 0.6–1.2 mg/hr or isosorbide dinitrate 1–2 mg/hr) are useful for the treatment of left ventricular failure and the relief of recurrent or persistent ischaemic pain. Intravenous β-blockers (e.g. atenolol 5–10 mg or metoprolol 5–15 mg given over 5 mins) relieve pain, reduce arrhythmias and improve short-term mortality in patients who present within 12 hours of the onset of symptoms (see Fig. 18.19). However, they should be avoided if there is heart failure (pulmonary oedema), hypotension (systolic BP < 105 mmHg) or bradycardia (heart rate < 65/min).

A dihydropyridine calcium channel antagonist (e.g. nifedipine or amlodipine) can be added to the β-blocker if there is persistent chest discomfort but may cause tachycardia if used alone. Because of their rate-limiting action, verapamil and diltiazem are the calcium channel antagonists of choice if a β-blocker is contraindicated.

**Reperfusion therapy**

**Non-ST segment elevation acute coronary syndrome**

Immediate emergency reperfusion therapy has no demonstrable benefit in patients with non-ST segment elevation MI and thrombolytic therapy may be harmful. Selected medium- to high-risk patients do benefit from in-hospital coronary angiography and coronary revascularisation but this does not need to take place in the first 12 hours.
Coronary artery disease

ST segment elevation acute coronary syndrome
Immediate reperfusion therapy restores coronary artery patency, preserves left ventricular function and improves survival. Successful therapy is associated with pain relief, resolution of acute ST elevation and, sometimes, transient arrhythmias (e.g. idioventricular rhythm).

**Primary percutaneous coronary intervention (PCI).** This is the treatment of choice for ST segment elevation MI (Figs 18.19 and 18.76). Outcomes are best when it is used in combination with glycoprotein IIb/IIIa receptor antagonists and intracoronary stent implantation. In comparison to thrombolytic therapy, it is associated with a greater reduction in the risk of death, recurrent MI or stroke (Box 18.67). The universal use of primary PCI has been limited by availability of the necessary resources to provide this highly specialised emergency service. Thus, intravenous thrombolytic therapy remains the first-line reperfusion treatment in many hospitals, especially those in rural or remote areas. When primary PCI cannot be achieved within 2 hours of diagnosis, thrombolytic therapy should be administered.

**Thrombolysis.** The appropriate use of thrombolytic therapy can reduce hospital mortality by 25–50% and this survival advantage is maintained for at least 10 years (Box 18.68). The benefit is greatest in those patients who receive treatment within the first few hours: ‘minutes mean muscle’.

Alteplase (human tissue plasminogen activator, or tPA) is a genetically engineered drug that is given over 90 minutes (bolus dose of 15 mg, followed by 0.75 mg/kg body weight but not exceeding 50 mg, over 30 mins, and then 0.5 mg/kg body weight but not exceeding 35 mg, over 60 mins). Its use is associated with better survival rates than other thrombolytic agents, such as streptokinase, but carries a slightly higher risk of intracerebral bleeding (10 per 1000 increased survival, but 1 per 1000 more non-fatal stroke).

Analogs of tPA, such as tenecteplase (TNK) and reteplase (rPA), have a longer plasma half-life than alteplase and can be given as an intravenous bolus. TNK is as effective as alteplase at reducing death and MI, whilst conferring similar intracerebral bleeding risks. However, other bleeding and transfusion risks are lower and the practical advantages of bolus administration provide opportunities for prompt treatment in the emergency department or in the pre-hospital setting. rPA is administered as a double bolus and also produces a similar outcome to that achieved with alteplase, although some of the bleeding risks appear slightly higher.

An overview of all large randomised trials confirms that thrombolytic therapy reduces short-term mortality in patients with MI if given within 12 hours of the onset...
of symptoms and the ECG shows bundle branch block or characteristic ST segment elevation of more than 1 mm in the limb leads or 2 mm in the chest leads (see Box 18.68). Thrombolysis appears to be of little net benefit and may be harmful in those who present more than 12 hours after the onset of symptoms and in those with a normal ECG or ST depression. In patients with ST elevation or bundle branch block, the absolute benefit of thrombolysis plus aspirin is approximately 50 lives saved per 1000 patients treated within 6 hours, and 40 lives saved per 1000 treated between 7 and 12 hours after the onset of symptoms. The benefit is greatest for patients treated within the first 2 hours.

The major hazard of thrombotic therapy is bleeding. Cerebral haemorrhage causes 4 extra strokes per 1000 patients treated, and the incidence of other major bleeds is between 0.5% and 1%. Accordingly, the treatment should be withheld if there is a significant risk of serious bleeding (Box 18.69).

For some patients, thrombotic therapy is contraindicated or fails to achieve coronary arterial reperfusion (see Fig. 18.19, p. 542). Early emergency PCI may then be considered, particularly where there is evidence of cardiogenic shock.

Complications of acute coronary syndrome
Complications are seen in all forms of acute coronary syndrome, although the frequency and extent vary with the severity of ischaemia and infarction. Major mechanical and structural complications are seen only with significant, often transmural, MI.

Arrhythmias
Many patients with acute coronary syndrome have some form of arrhythmia (Box 18.70). In the majority of cases this is transient and of no haemodynamic or prognostic importance. Pain relief, rest and the correction of hypokalaemia may help prevent arrhythmias. Diagnosis and management of arrhythmias are discussed in detail on pages 562–579.

Ventricular fibrillation
This occurs in 5–10% of patients who reach hospital and is thought to be the major cause of death in those who die before receiving medical attention. Prompt defibrillation restores sinus rhythm and is life-saving.

Atrial fibrillation
This is common but frequently transient, and usually does not require emergency treatment. However, if it causes a rapid ventricular rate with hypotension or circulatory collapse, prompt cardioversion by immediate synchronised DC shock is essential. In other situations, digoxin or a β-blocker is usually the treatment of choice.

Bradycardia
This does not usually require treatment, but if there is hypotension or haemodynamic deterioration, atropine (0.6–1.2 mg IV) may be given. AV block complicating inferior infarction is usually temporary and often resolves following reperfusion therapy. If there is clinical deterioration due to second-degree or complete AV block, a temporary pacemaker should be considered. AV block complicating anterior infarction is more serious because asystole may suddenly supervene; a prophylactic temporary pacemaker should be inserted (p. 578).

Ischaemia
Patients who develop recurrent angina at rest or on minimal exertion following an acute coronary syndrome
are at high risk and should be considered for prompt coronary angiography with a view to revascularisation. Patients with dynamic ECG changes and ongoing pain should be treated with intravenous glycoprotein IIb/IIIa receptor antagonists. Patients with resistant pain or marked haemodynamic changes should be considered for intra-aortic balloon counterpulsation and emergency coronary revascularisation.

Post-infarct angina occurs in up to 50% of patients treated with thrombolysis. Most patients have a residual stenosis in the infarct-related vessel, despite successful thrombolysis, and this may cause angina if there is still viable myocardium downstream. For this reason, all patients who have received successful thrombolysis should be considered for early (within the first 6–24 hours) coronary angiography with a view to coronary revascularisation.

**Acute circulatory failure**

Acute circulatory failure usually reflects extensive myocardial damage and indicates a bad prognosis. All the other complications of MI are more likely to occur when acute heart failure is present. The assessment and management of heart failure complicating acute MI are discussed in detail on page 545.

**Pericarditis**

This only occurs following infarction and is particularly common on the second and third days. The patient may recognise that a different pain has developed, even though it is at the same site, and that it is positional and tends to be worse or sometimes only present on inspiration. A pericardial rub may be audible. Opiate-based analgesia should be used. Non-steroidal (NSAIDs) and steroidal anti-inflammatory drugs may increase the risk of aneurysm formation and myocardial rupture in the early recovery period, and so should be avoided.

The post-MI syndrome (Dressler’s syndrome) is characterised by persistent fever, pericarditis and pleurisy, and is probably due to autoimmunity. The symptoms tend to occur a few weeks or even months after the infarct and often subside after a few days; prolonged or severe symptoms may require treatment with high-dose aspirin, NSAIDs or even corticosteroids.

**Mechanical complications**

Part of the necrotic muscle in a fresh infarct may tear or rupture, with devastating consequences:

- **Rupture of the papillary muscle** can cause acute pulmonary oedema and shock due to the sudden onset of severe mitral regurgitation, which presents with a pansystolic murmur and third heart sound. In the presence of severe regurgitation, the murmur may be quiet or absent. The diagnosis is confirmed by echocardiography and emergency valve replacement may be necessary. Lesser degrees of mitral regurgitation due to papillary muscle dysfunction are common and may be transient.

- **Rupture of the interventricular septum** causes left-to-right shunting through a ventricular septal defect. This usually presents with sudden haemodynamic deterioration accompanied by a new loud pansystolic murmur radiating to the right sternal border, but may be difficult to distinguish from acute mitral regurgitation. However, patients with an acquired ventricular septal defect tend to develop right heart failure rather than pulmonary oedema. Doppler echocardiography and right heart catheterisation will confirm the diagnosis. Without prompt surgery, the condition is usually fatal.

  - **Rupture of the ventricle** may lead to cardiac tamponade and is usually fatal (p. 545), although it may rarely be possible to support a patient with an incomplete rupture until emergency surgery is performed.

**Embolism**

Thrombus often forms on the endocardial surface of freshly infarcted myocardium. This can lead to systemic embolism and occasionally causes a stroke or ischaemic limb. Venous thrombosis and pulmonary embolism may occur but have become less common with the use of prophylactic anticoagulants and early mobilisation.

**Impaired ventricular function, remodelling and ventricular aneurysm**

Acute transmural MI is often followed by thinning and stretching of the infarcted segment (infarct expansion). This leads to an increase in wall stress with progressive dilatation and hypertrophy of the remaining ventricle (ventricular remodelling, Fig. 18.77). As the ventricle dilates, it becomes less efficient and heart failure may supervene. Infarct expansion occurs over a few days and weeks but ventricular remodelling can take years. ACE inhibitor therapy reduces late ventricular remodelling and can prevent the onset of heart failure (p. 551).

A left ventricular aneurysm develops in approximately 10% of patients with MI and is particularly common when there is persistent occlusion of the infarct-related vessel. Heart failure, ventricular arrhythmias, mural thrombus and systemic embolism are all recognised complications of aneurysm formation. Other features include a paradoxical impulse on the chest wall, persistent ST elevation on the ECG, and sometimes an unusual bulge from the cardiac silhouette on the chest X-ray. Echocardiography is diagnostic. Surgical removal
of a left ventricular aneurysm carries a high morbidity and mortality but is sometimes necessary.

**Later in-hospital management**

Late management of MI is summarised in Box 18.71.

**Risk stratification and further investigation**

Simple clinical tools can be used to identify medium- to high-risk patients. The GRACE score (see Fig. 18.70, p. 591) is a simple method of calculating early mortality that can help guide which patients should be selected for intensive therapy, and specifically early inpatient coronary angiography.

The prognosis of patients who have survived an acute coronary syndrome is related to the extent of residual myocardial ischaemia, the degree of myocardial damage, and the presence of ventricular arrhythmias.

**Left ventricular function**

The degree of left ventricular dysfunction can be crudely assessed from physical findings (tachycardia, third heart sound, crackles at the lung bases, elevated venous pressure and so on), ECG changes and chest X-ray (size of the heart and presence of pulmonary oedema). Formal assessment with echocardiography should, however, be undertaken in the early recovery phase.

**Ischaemia**

Patients with early ischaemia following an acute coronary syndrome should undergo coronary angiography with a view to revascularisation. Low-risk patients without spontaneous ischaemia should undergo an exercise tolerance test approximately 4 weeks after the acute coronary syndrome. This will help to identify those individuals with residual myocardial ischaemia who require further investigation, and may help to boost the confidence of the remainder.

If the exercise test is negative and the patient has a good effort tolerance, the outlook is good, with a 1–4% chance of an adverse event in the next 12 months. In contrast, patients with residual ischaemia in the form of chest pain or ECG changes at low exercise levels are at high risk, with a 15–25% chance of suffering a further ischaemic event in the next 12 months.

---

**Arhythmias**

The presence of ventricular arrhythmias during the convalescent phase of acute coronary syndrome may be a marker of poor ventricular function and may herald sudden death. Although empirical anti-arrhythmic treatment is of no value and is even hazardous, selected patients may benefit from electrophysiological testing and specific anti-arrhythmic therapy (including implantable cardiac defibrillators, p. 579).

Recent ventricular arrhythmias are sometimes manifestations of myocardial ischaemia or impaired left ventricular function and may respond to appropriate treatment directed at the underlying problem.

**Lifestyle and risk factor modification**

**Smoking**

The 5-year mortality of patients who continue to smoke cigarettes is double that of those who quit smoking at the time of their acute coronary syndrome. Giving up smoking is the single most effective contribution a patient can make to his or her future. The success of smoking cessation can be increased by supportive advice and pharmacological therapy (p. 100).

**Hyperlipidaemia**

The importance of lowering serum cholesterol following acute coronary syndrome has been demonstrated in large-scale randomised trials. Lipids should be measured within 24 hours of presentation because there is often a transient fall in cholesterol in the 3 months following infarction. HMG CoA reductase enzyme inhibitors (‘statins’, p. 456) can produce marked reductions in total (and LDL) cholesterol and reduce the subsequent risk of death, re-infarction, stroke and the need for revascularisation (see Box 18.46, p. 583). Irrespective of serum cholesterol concentrations, all patients should receive statin therapy after acute coronary syndrome, but those with serum LDL cholesterol concentrations above 3.2 mmol/L (~120 mg/dL) benefit from more intensive therapy, such as atorvastatin 80 mg daily.

**Other risk factors**

Maintaining an ideal body weight, eating a Mediterranean-style diet, taking regular exercise, and achieving good control of hypertension and diabetes mellitus may all improve the long-term outlook.

**Mobilisation and rehabilitation**

The necrotic muscle of an acute myocardial infarct takes 4–6 weeks to be replaced with fibrous tissue and it is conventional to restrict physical activities during this period. When there are no complications, the patient can mobilise on the second day, return home in 3–5 days and gradually increase activity, with the aim of returning to work in 4–6 weeks. The majority of patients may resume driving after 4–6 weeks, although, in most countries, vocational driving licence holders (e.g. heavy goods and public service vehicles) require special assessment.

Emotional problems, such as denial, anxiety and depression, are common and must be addressed. Many patients are severely and even permanently incapacitated as a result of the psychological effects of acute coronary syndrome rather than the physical ones, and all benefit from thoughtful explanation, counselling and
reassurance at every stage of the illness. Many patients mistakenly believe that ‘stress’ was the cause of their heart attack and may restrict their activity inappropriately. The patient’s spouse or partner will also require emotional support, information and counselling. Formal rehabilitation programmes, based on graded exercise protocols with individual and group counselling, are often very successful and, in some cases, have been shown to improve the long-term outcome.

**Secondary prevention drug therapy**

**Aspirin and clopidogrel**

Low-dose aspirin therapy reduces the risk of further infarction and other vascular events by approximately 25% and should be continued indefinitely if there are no unwanted effects. Clopidogrel should be given in combination with aspirin for at least 3 months. If patients are intolerant of long-term aspirin, clopidogrel is a suitable alternative.

**Beta-blockers**

Continuous treatment with an oral β-blocker reduces long-term mortality by approximately 25% among the survivors of acute MI (Box 18.72). Unfortunately, a minority of patients do not tolerate β-blockers because of bradycardia, AV block, hypotension or asthma. Patients with heart failure, irreversible chronic obstructive pulmonary disease or peripheral vascular disease derive similar, if not greater second-ary preventative benefits from β-blocker therapy if they can tolerate it, so it should be tried. The secondary preventative role of β-blockers in patients with unstable angina is unknown.

**ACE inhibitors**

Several clinical trials have shown that long-term treatment with an ACE inhibitor (e.g. enalapril 10 mg twice daily or ramipril 2.5-5 mg twice daily) can counteract ventricular remodelling, prevent the onset of heart failure, improve survival, reduce recurrent MI and avoid rehospitalisation. The benefits are greatest in those with overt heart failure (clinical or radiological) but extend to patients with asymptomatic left ventricular dysfunction and those with preserved left ventricular function. They should therefore be considered in all patients with acute coronary syndrome. Caution must be exercised in hypovolaemic or hypotensive patients because the introduction of an ACE inhibitor may exacerbate hypotension and impair coronary perfusion. In patients intolerant of ACE inhibitors, angiotensin receptor blockers (e.g. valsartan 40–160 mg twice daily or candesartan 4–16 mg daily) are alternatives and are better tolerated.

Patients with acute MI and left ventricular dysfunction (ejection fraction < 35%) and either pulmonary oedema or diabetes mellitus further benefit from additional mineralocorticoid receptor antagonism (e.g. eplerenone 25–50 mg daily).

**Coronary revascularisation**

Most low-risk patients stabilise with aspirin, clopidogrel, anticoagulation and anti-anginal therapy, and can be rapidly mobilised. In the absence of recurrent symptoms, low-risk patients do not benefit from routine coronary angiography. Coronary angiography should be considered with a view to revascularisation in all patients at moderate or high risk, including those who fail to settle on medical therapy, those with extensive ECG changes, those with an elevated plasma troponin and those with severe pre-existing stable angina. This often reveals disease that is amenable to PCI or urgent CABG. In these cases, coronary revascularisation is associated with short- and long-term benefits, including reductions in MI and death.

**Device therapy**

Implantable cardiac defibrillators are of benefit in preventing sudden cardiac death in patients who have severe left ventricular impairment (ejection fraction ≤ 30%) after MI (p. 579).

**Prognosis**

In almost one-quarter of all cases of MI, death occurs within a few minutes without medical care. Half the deaths occur within 24 hours of the onset of symptoms and about 40% of all affected patients die within the first month. The prognosis of those who survive to reach hospital is much better, with a 28-day survival of more than 85%. Patients with unstable angina have a mortality of approximately half that of those patients with MI.

Early death is usually due to an arrhythmia and is independent of the extent of MI. However, late outcomes are determined by the extent of myocardial damage, and unfavourable features include poor left ventricular function, AV block and persistent ventricular arrhythmias. The prognosis is worse for anterior than for inferior infarcts. Bundle branch block and high

---

**18.73 Myocardial infarction in old age**

- **Atypical presentation:** often with anorexia, fatigue or weakness rather than chest pain.
- **Case fatality:** rises steeply. Hospital mortality exceeds 25% in those over 75 yrs old, which is five times greater than that seen in those aged less than 55 yrs.
- **Survival benefit of treatments:** not influenced by age. The absolute benefit of evidence-based treatments may therefore be greatest in older people.
- **Hazards of treatments:** rise with age (e.g. increased risk of intracerebral bleeding after thrombolysis) and are due partly to increased comorbidity.
- **Quality of evidence:** older patients, particularly those with significant comorbidity, were under-represented in many of the randomised controlled clinical trials that helped to establish the treatment of MI. The balance of risk and benefit for many treatments (e.g. thrombolysis, primary percutaneous transluminal coronary angiography) in frail older people is therefore uncertain.
cardiac marker levels both indicate extensive myocardial damage. Old age, depression and social isolation are also associated with a higher mortality.

Of those who survive an acute attack, more than 80% live for a further year, about 75% for 5 years, 50% for 10 years and 25% for 20 years.

Cardiac risk of non-cardiac surgery

Non-cardiac surgery, particularly major vascular, abdominal or thoracic surgery, can precipitate serious peri-operative cardiac complications, such as MI and death, in patients with coronary artery and other forms of heart disease. Careful pre-operative cardiac assessment may help to determine the balance of benefit versus risk on an individual basis, and identify measures that minimise the operative risk (Box 18.74).

A hypercoagulable state is part of the normal physiological response to surgery, and may promote coronary thrombosis leading to an acute coronary syndrome in the early post-operative period. Patients with a history of recent PCI or acute coronary syndrome are at greatest risk and, whenever possible, elective non-cardiac surgery should be avoided for 3 months after such an event. Antiplatelet agents, statins and β-blockers reduce the risk of peri-operative MI in patients with coronary artery disease and, where possible, should be prescribed throughout the peri-operative period.

Careful attention to fluid balance during and after surgery is particularly important in patients with impaired left ventricular function and valvular heart disease because antidiuretic hormone is released as part of the normal physiological response to surgery and, in these circumstances, the overzealous administration of intravenous fluids can easily precipitate heart failure. Patients with severe valvular heart disease, particularly aortic stenosis and mitral stenosis, are also at increased risk because they may not be able to increase their cardiac output in response to the stress of surgery.

Atrial fibrillation may be triggered by hypoxia, myocardial ischaemia or heart failure, and is a common post-operative complication in patients with pre-existing heart disease. It usually terminates spontaneously when the precipitating factors have been eliminated, but digoxin or β-blockers can be prescribed to control the heart rate.

VASCULAR DISEASE

Peripheral arterial disease

In developed countries, almost all peripheral arterial disease (PAD) is due to atherosclerosis (p. 579) and so shares common risk factors with coronary artery disease: namely, smoking, diabetes mellitus, hyperlipidaemia and hypertension. As with coronary artery disease, plaque rupture is responsible for the most serious manifestations of PAD, and not infrequently occurs in a plaque that hitherto has been asymptomatic.

Approximately 20% of middle-aged (55–75 years) people in the UK have PAD but only one-quarter of them will have symptoms. The clinical manifestations depend upon the anatomical site, the presence or absence of a collateral supply, the speed of onset and the mechanism of injury (Box 18.75).

Chronic lower limb arterial disease

PAD affects the leg eight times more often than the arm. The lower limb arterial tree comprises the aorto-iliac
Intermittent claudication

This term describes ischaemic pain affecting the muscles of the leg upon walking. The pain is usually felt in the calf because the disease most commonly affects the superficial femoral artery. However, the pain may be felt in the thigh or buttock if the iliac arteries are involved. Typically, the pain comes on after a reasonably constant ‘claudication distance’ and rapidly subsides on stopping walking. Resumption of walking leads to a return of the pain. Most patients describe a cyclical pattern of exacerbation and resolution due to the progression of disease and the subsequent development of collaterals.

Approximately 5% of middle-aged men report IC. Provided patients comply with ‘best medical therapy’ (BMT, Box 18.77), only 1–2% per year will deteriorate to a point where amputation and/or revascularisation are required. However, the annual mortality rate exceeds 5%, 2–3 times higher than in an equivalent non-claudicant population. This is because IC is nearly always found in association with widespread atherosclerosis, so that most claudicants succumb to MI or stroke. The mainstay of treatment is BMT, including (preferably supervised) exercise therapy. The peripheral vasodilator, cilostazol, has been shown to improve walking distance. Intervention with angioplasty, stenting, endarterectomy or bypass is usually only considered after BMT has been given at least 6 months to effect symptomatic improvement, and then only in patients who are severely disabled or whose livelihood is threatened by their disability.

Critical limb ischaemia

This is defined as rest (night) pain, requiring opiate analgesia, and/or tissue loss (ulceration or gangrene), present for more than 2 weeks, in the presence of an ankle BP of less than 50 mmHg (Fig. 18.78). Rest pain only, with ankle pressures above 50 mmHg, is known as subcritical limb ischaemia (SCLI). The term severe limb ischaemia (SLI) is used to describe both CLI and SCLI. Whereas IC is usually due to single-segment plaque, SLI is always due to multilevel disease.

Many patients with SLI have not previously sought medical advice for IC, principally because they have other comorbidity that prevents them from walking to a point where claudication pain might develop. In contrast to patients with IC, those with SLI are at high risk of losing their limb, and sometimes their life, in a matter of weeks or months without surgical bypass or endovascular revascularisation by angioplasty or stenting. Treatment is difficult, however, because patients

---

**18.76 Clinical features of chronic lower limb ischaemia**

- **Pulses:** diminished or absent
- **Bruit:** denote turbulent flow but bear no relationship to the severity of the underlying disease
- **Reduced skin temperature**
- **Pallor on elevation and rubor on dependency** (Buerger’s sign)
- **Superficial veins that fill sluggishly and empty (‘gutter’) upon minimal elevation**
- **Muscle-wasting**
- **Skin and nails:** dry, thin and brittle
- **Loss of hair**

**18.77 Best medical therapy (BMT) for peripheral arterial disease**

- **Smoking cessation**
- **Regular exercise** (30 mins of walking, three times per week)
- **Antiplatelet agent** (aspirin 75 mg or clopidogrel 75 mg daily)
- **Reduction of cholesterol** (diet and statin therapy)
- **Diagnosis and treatment of diabetes mellitus** (all should have fasting glucose measured)
- **Diagnosis and treatment of frequently associated conditions** (e.g. hypertension, anaemia, heart failure)

*All patients with any manifestation of PAD should be considered for BMT.*
have extensive and severe (often bilateral) end-stage disease, are usually elderly and nearly always have significant multisystem comorbidity. Imaging is performed using duplex ultrasonography, MRI or CT with intravenous injection of contrast agents. Intra-arterial digital subtraction angiography (IA-DSA) is usually reserved for those undergoing endovascular revascularisation.

**Diabetic vascular disease**

Approximately 5–10% of patients with PAD have diabetes but this proportion increases to 30–40% in those with SLI. Diabetes does not cause obstructive microangiopathy at the capillary level, as previously thought, and so is not a contraindication to lower limb revascularisation. Nevertheless, the ‘diabetic foot’ does pose a number of particular problems (Box 18.78 and p. 833). If the blood supply is adequate, then dead tissue can be excised in the expectation that healing will occur, provided infection is controlled and the foot is protected from pressure. However, if significant ischaemia is also present, the priority is to revascularise the foot if possible. Sadly, many diabetic patients present late with extensive tissue loss, which accounts for the high amputation rate.

### 18.78 Diabetic vascular disease: the ‘diabetic foot’

<table>
<thead>
<tr>
<th>Feature</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arterial calcification</td>
<td>Spuriously high ABPI due to incompressible ankle vessels. Inability to clamp arteries for the purposes of bypass surgery. Resistant to angioplasty.</td>
</tr>
<tr>
<td>Immunocompromise</td>
<td>Prone to rapidly spreading cellulitis, gangrene and osteomyelitis.</td>
</tr>
<tr>
<td>Multisystem arterial disease</td>
<td>Coronary and cerebral arterial disease increase the risks of intervention.</td>
</tr>
<tr>
<td>Distal disease</td>
<td>Diabetic vascular disease has a predilection for the calf vessels. Although vessels in the foot are often spared, performing a satisfactory bypass or angioplasty to these small vessels is a technical challenge.</td>
</tr>
<tr>
<td>Sensory neuropathy</td>
<td>Even severe ischaemia and/or tissue loss may be completely painless. Diabetic patients often present late with extensive destruction of the foot. Loss of proprioception leads to abnormal pressure loads and worsens joint destruction (Charcot joints).</td>
</tr>
<tr>
<td>Motor neuropathy</td>
<td>Weakness of the long and short flexors and extensors leads to abnormal foot architecture, abnormal pressure loads, callus formation and ulceration.</td>
</tr>
<tr>
<td>Autonomic neuropathy</td>
<td>Leads to a dry foot deficient in sweat that normally lubricates the skin and is antibacterial. Scaling and fissuring create a portal of entry for bacteria. Abnormal blood flow in the bones of the ankle and foot may also contribute to osteopenia and bony collapse.</td>
</tr>
</tbody>
</table>

(BABPI = ankle–brachial pressure index)

---

**Buerger’s disease (thromboangiitis obliterans)**

This is an inflammatory obliterative arterial disease that is distinct from atherosclerosis and usually presents in young (20–30 years) male smokers. It is most common in those from the Mediterranean and North Africa. It characteristically affects distal arteries, giving rise to claudication in the feet or rest pain in the fingers or toes. Wrist and ankle pulses are absent but brachial and popliteal pulses are present. Disease also affects the veins, giving rise to superficial thrombophlebitis. It often reverts if the patient stops smoking; sympathectomy and prostaglandin infusions may be helpful. Major limb amputation is the most frequent outcome if patients continue to smoke.

**Chronic upper limb arterial disease**

In the arm, the subclavian artery is the most common site of disease, which may manifest as:

- **Arm claudication** (rare).
- **Atherosclerosis** (blue finger syndrome). Small emboli lodge in digital arteries and may be confused with Raynaud’s phenomenon (see below) but, in this case, the symptoms are unilateral. Failure to make the diagnosis may eventually lead to amputation.
  - **Subclavian steal.** When the arm is used, blood is ‘stolen’ from the brain via the vertebral artery. This leads to vertebo-basilar ischaemia, which is characterised by dizziness, cortical blindness and/or collapse. Where possible, subclavian artery disease is treated by means of angioplasty and stenting, as surgery (e.g. carotid–subclavian bypass) can be difficult.

**Raynaud’s phenomenon and Raynaud’s disease**

Cold (and emotional) stimuli may trigger vasospasm, leading to the characteristic sequence of digital pallor due to vasospasm, cyanosis due to deoxygenated blood, and rubor due to reactive hyperaemia.

**Primary Raynaud’s phenomenon (or disease)**

This affects 5–10% of young women aged 15–30 years in temperate climates and may be familial. It does not
Vascular disease

progress to ulceration or infarction, and significant pain is unusual. The underlying cause is unclear. No investigation is necessary. The patient should be reassured and advised to avoid exposure to cold. Long-acting nifedipine may be helpful but sympathectomy is not indicated.

**Secondary Raynaud’s phenomenon (or syndrome)**

This occurs in older people in association with connective tissue disease (most commonly systemic sclerosis or CREST syndrome, p. 1112), vibration-induced injury (from the use of power tools) and thoracic outlet obstruction (e.g. cervical rib). Unlike primary disease, it is often associated with fixed obstruction of the digital arteries, fingertip ulceration, and necrosis and pain. The fingers must be protected from cold and trauma, infection requires treatment with antibiotics, and surgery should be avoided if possible. Vasoactive drugs have no clear benefit. Sympathectomy helps for a year or two. Prostacyclin infusions are sometimes beneficial.

**Acute limb ischaemia**

This is most frequently caused by acute thrombotic occlusion of a pre-existing stenotic arterial segment, thromboembolism, and trauma that may be iatrogenic. Apart from paralysis (inability to wiggle toes/fingers) and paraesthesia (loss of light touch over the dorsum of the foot/hand), the so-called ‘Ps of acute ischaemia’ (Box 18.80) are non-specific for ischaemia and/or inconsistently related to its severity. Pain on squeezing the calf indicates muscle infarction and impending irreversible ischaemia.

All patients with suspected acutely ischaemic limbs must be discussed immediately with a vascular surgeon; a few hours can make the difference between death/amputation and complete recovery of limb function. If there are no contraindications (for example, acute aortic dissection or trauma, particularly head injury), an intravenous bolus of heparin (3000-5000 U) should be administered to limit propagation of thrombus and protect the collateral circulation. Distinguishing thrombosis from embolism is frequently difficult but is important because treatment and prognosis are different (Box 18.81). Acute limb ischaemia due to thrombosis in situ can usually be treated medically in the first instance with intravenous heparin (target activated partial thromboplastin time (APTT) 2.0-3.0), antiplatelet agents, high-dose statins, intravenous fluids to avoid dehydration, correction of anaemia, oxygen and sometimes prostaglandins, such as iloprost. Careful monitoring is required. Embolism will normally result in extensive tissue necrosis within 6 hours unless the limb is revascularised. The indications for thrombolysis, if any, remain controversial. Irreversible ischaemia mandates early amputation or palliative care.

**Cerebrovascular/renovascular disease and ischaemic gut injury**

See Ch. 27 and pp. 494 and 909.

**Diseases of the aorta**

Aneurysm, dissection and aortitis are the main pathologies (Fig. 18.79).

**Aortic aneurysm**

This is an abnormal dilatation of the aortic lumen; a true aneurysm involves all the layers of the wall, whereas a false aneurysm does not.

**Aetiology and types of aneurysm**

- **Non-specific aneurysms**
  - Why some patients develop occlusive vascular disease, some develop atherosclerotic vascular disease and some develop both in response to atherosclerosis risk factors remains unclear. Unlike occlusive disease, atheroscleromal disease tends to run in families and genetic factors are undoubtedly important. The most common site for ‘non-specific’ aneurysm formation is the infrarenal abdominal aorta. The suprarenal abdominal aorta and a variable length of the descending thoracic aorta may be affected in 10–20% of patients but the ascending aorta is usually spared.
  - **Marfan’s syndrome**
    - This disorder of connective tissue is inherited as an autosomal dominant trait and is caused by mutations in the...
Cardiovascular disease

Fibrillin gene on chromosome 15. Affected systems include the skeleton (arachnodactyly, joint hypermobility, scoliosis, chest deformity and high arched palate), the eyes (dislocation of the lens) and the cardiovascular system (aortic disease and mitral regurgitation). Weakening of the aortic media leads to aortic root dilatation, regurgitation and dissection (see below). Pregnancy is particularly hazardous. Chest X-ray, echocardiography, MRI or CT may detect aortic dilatation at an early stage and can be used to monitor the disease.

Treatment with β-blockers reduces the rate of aortic dilatation and the risk of rupture. Elective replacement of the ascending aorta may be considered in patients with evidence of progressive aortic dilatation but carries a mortality of 5–10%.

Aortitis

Syphilis is a rare cause of aortitis that characteristically produces saccular aneurysms of the ascending aorta containing calcification. Other rare conditions associated with aortitis include Takayasu’s disease (p. 1116), Reiter’s syndrome (p. 1107), giant cell arteritis and ankylosing spondylitis (pp. 1105 and 1117).

Thoracic aortic aneurysms

These may produce chest pain, aortic regurgitation, compressive symptoms such as stridor (trachea, bronchus) and hoarseness (recurrent laryngeal nerve), and superior vena cava syndrome (see Fig. 18.79A). If they erode into adjacent structures, e.g. aorto-oesophageal fistula, massive bleeding occurs.

Abdominal aortic aneurysms

Abdominal aortic aneurysms (AAAs) are present in 5% of men aged over 60 years and 80% are confined to the infrarenal segment. Men are affected three times more commonly than women. AAAs can present in a number of ways (Box 18.82). The usual age at presentation is 65–75 years for elective presentations and 75–85 years for emergency presentations. Ultrasound is the best way of establishing the diagnosis and of following up patients with asymptomatic aneurysms that are not yet large.
enough to warrant surgical repair. CT provides more accurate information about the size and extent of the aneurysm, the surrounding structures and whether there is any other intra-abdominal pathology. It is the standard pre-operative investigation but is not suitable for surveillance because of cost and radiation dose.

**Management.** Until an asymptomatic AAA has reached a maximum of 5.5 cm in diameter, the risks of surgery generally outweigh the risks of rupture (Box 18.83). All symptomatic AAAs should be considered for repair, not only to rid the patient of symptoms but also because pain often predates rupture. Distal embolisation is a strong indication for repair, regardless of size, because otherwise limb loss is common. Most patients with a ruptured AAA do not survive to reach hospital, but if they do and surgery is thought to be appropriate, there must be no delay in getting them to the operating theatre to clamp the aorta.

Open AAA repair has been the treatment of choice in both the elective and the emergency settings, and entails replacing the aneurysmal segment with a prosthetic (usually Dacron) graft. The 30-day mortality for this procedure is approximately 5–8% for elective asymptomatic AAA, 10–20% for emergency symptomatic AAA and 50% for ruptured AAA. However, patients who survive the operation to leave hospital have a long-term survival which approaches that of the normal population. Increasingly, endovascular aneurysm repair (EVAR), using a stent-graft introduced via the femoral arteries in the groin, is replacing open surgery. It is cost-effective and likely to become the treatment of choice for infrarenal AAA. It is possible to treat many suprarenal and thoraco-abdominal aneurysms by EVAR too.

In the UK, a national screening programme for men over 65 years of age has been introduced using ultrasound scanning. For every 10 000 men scanned, 65 ruptures are prevented and 52 lives saved.

### Aortic dissection

A breach in the integrity of the aortic wall allows arterial blood to enter the media, which is then split into two layers, creating a ‘false lumen’ alongside the existing or ‘true lumen’ (see Fig. 18.79B). The aortic valve may be damaged and the branches of the aorta may be compromised. Typically, the false lumen eventually re-enters the true lumen, creating a double-barrelled aorta, but it may also rupture into the left pleural space or pericardium with fatal consequences.

The primary event is often a spontaneous or iatrogenic tear in the intima of the aorta; multiple tears or entry points are common. Other dissections are triggered by primary haemorrhage in the media of the aorta, which then ruptures through the intima into the true lumen. This form of spontaneous bleeding from the vasa vasorum is sometimes confined to the aortic wall, when it may present as a painful intramural haematoma.

Aortic disease and hypertension are the most important aetiological factors but other conditions may also be implicated (Box 18.84). Chronic dissections may lead to aneurysmal dilatation of the aorta, and thoracic aneurysms may be complicated by dissection. It can therefore be difficult to identify the primary pathology.

The peak incidence is in the sixth and seventh decades but dissection can occur in younger patients, usually in association with Marfan’s syndrome, pregnancy or trauma; men are twice as frequently affected as women.

Aortic dissection is classified anatomically and for management purposes into type A and type B (see Fig. 18.79B), involving or sparing the ascending aorta, respectively. Type A dissections account for two-thirds of cases and frequently also extend into the descending aorta.

### 18.82 Abdominal aortic aneurysm: common presentations

**Incidental**

- On physical examination, plain X-ray or, most commonly, abdominal ultrasound
- Even large AAAs can be difficult to feel, so many remain undetected until they rupture
- Studies are currently under way to determine whether screening will reduce the number of deaths from rupture (see Box 18.83)

**Pain**

- In the central abdomen, back, loin, iliac fossa or groin

**Thromboembolic complications**

- Thrombus within the aneurysm sac may be a source of emboli to the lower limbs
- Less commonly, the aorta may undergo thrombotic occlusion

**Compression**

- Surrounding structures such as the duodenum (obstruction and vomiting) and the inferior vena cava (oedema and deep vein thrombosis)

**Rupture**

- Into the retroperitoneum, the peritoneal cavity or surrounding structures (most commonly the inferior vena cava, leading to an aortocaval fistula)

### EBM 18.83 Population screening and prevention of ruptured abdominal aortic aneurysm

‘Ultrasound screening for AAA in men aged 65–75 years, with surgical repair of those AAAs that are bigger than 5.5 cm, are rapidly growing or become symptomatic, reduces the community incidence of rupture by approximately 50% and is cost-effective.’


For further information: www.mrc-bsu.cam.ac.uk

### 18.84 Factors that may predispose to aortic dissection

- Hypertension (in 80%)
- Aortic atherosclerosis
- Non-specific aortic aneurysm
- Aortic coarctation (p. 632)
- Collagen disorders (e.g. Marfan’s syndrome, Ehlers–Danlos syndrome)
- Fibromuscular dysplasia
- Previous aortic surgery (e.g. CABG, aortic valve replacement)
- Pregnancy (usually third trimester)
- Trauma
- Iatrogenic (e.g. cardiac catheterisation, intra-aortic balloon pumping)

(CABG = coronary artery bypass grafting)
Clinical features

Involvement of the ascending aorta typically gives rise to anterior chest pain, and involvement of the descending aorta to intrascapular pain. The pain is typically described as ‘tearing’ and very abrupt in onset; collapse is common. Unless there is major haemorrhage, the patient is invariably hypertensive. There may be asymmetry of the brachial, carotid or femoral pulses and signs of aortic regurgitation. Occlusion of aortic branches may cause MI (coronary), stroke (carotid) paraplegia (spinal), mesenteric infarction with an acute abdomen (coeliac and superior mesenteric), renal failure (renal) and acute limb (usually leg) ischaemia.

Investigations

The chest X-ray characteristically shows broadening of the upper mediastinum and distortion of the aortic ‘knuckle’, but these findings are variable and are absent in 10% of cases. A left-sided pleural effusion is common. The ECG may show left ventricular hypertrophy in patients with hypertension, or rarely changes of acute MI (usually inferior). Doppler echocardiography may
show aortic regurgitation, a dilated aortic root and, occasionally, the flap of the dissection. Transoesophageal echocardiography is particularly helpful because trans-thoracic echocardiography can only provide images of the first 3–4 cm of the ascending aorta (Fig. 18.80). CT and MRI angiography (Figs 18.81 and 18.82) are both highly specific and sensitive.

**Management**

The early mortality of acute dissection is approximately 1–5% per hour and so treatment is urgently required. Initial management comprises pain control and antihypertensive treatment. Type A dissections require emergency surgery to replace the ascending aorta. Type B aneurysms are treated medically unless there is actual or impending external rupture, or vital organ (gut, kidneys) or limb ischaemia, as the morbidity and mortality associated with surgery are very high. The aim of medical management is to maintain a mean arterial pressure (MAP) of 60–75 mmHg to reduce the force of the ejection of blood from the LV. First-line therapy is with β-blockers; the additional α-blocking properties of labetalol make it especially useful. Rate-limiting calcium channel blockers, such as verapamil or diltiazem, are used if β-blockers are contraindicated. Sodium nitroprusside may be considered if these fail to control BP adequately.

Percutaneous or minimal access endoluminal repair is sometimes possible and involves either ‘fenestrating’ (perforating) the intimal flap so that blood can return from the false to the true lumen (so decompressing the former), or implanting a stent graft placed from the femoral artery (see Fig. 18.82).

**Hypertension**

Within any population, blood pressure values occur within a continuum, and are determined by mechanical, hormonal and environmental factors. Any definition of hypertension therefore utilises arbitrary threshold values within this continuum. Systemic BP rises with age, and the incidence of cardiovascular disease (particularly stroke and coronary artery disease) is closely related to average BP at all ages, even when BP readings are within the so-called ‘normal range’.

The cardiovascular risks associated with BP depend upon the combination of risk factors in an individual, such as age, gender, weight, physical activity, smoking, family history, serum cholesterol, diabetes mellitus and pre-existing vascular disease. Thus a practical definition of hypertension is ‘the level of BP at which the benefits of treatment outweigh the costs and hazards’. The British Hypertension Society classification is provided in Box 18.85 and is consistent with those defined by the European Society of Hypertension and the World Health Organization–International Society of Hypertension.

**Aetiology**

In more than 95% of cases, a specific underlying cause of hypertension cannot be found. Such patients are said to have essential hypertension. The pathogenesis is not clearly understood. Many factors may contribute to its development, including renal dysfunction, peripheral resistance vessel tone, endothelial dysfunction, autonomic tone, insulin resistance and neurohumoral factors.

### 18.85 Definition of hypertension

<table>
<thead>
<tr>
<th>Category</th>
<th>Systolic BP (mmHg)</th>
<th>Diastolic BP (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal</td>
<td>&lt; 120</td>
<td>&lt; 80</td>
</tr>
<tr>
<td>Normal</td>
<td>&lt; 130</td>
<td>85</td>
</tr>
<tr>
<td>High normal</td>
<td>130–139</td>
<td>85–89</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1 (mild)</td>
<td>140–159</td>
<td>90–99</td>
</tr>
<tr>
<td>Grade 2 (moderate)</td>
<td>160–179</td>
<td>100–109</td>
</tr>
<tr>
<td>Grade 3 (severe)</td>
<td>≥ 180</td>
<td>&gt; 110</td>
</tr>
<tr>
<td><strong>Isolated systolic hypertension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>140–159</td>
<td>&lt; 90</td>
</tr>
<tr>
<td>Grade 2</td>
<td>≥ 160</td>
<td>&lt; 90</td>
</tr>
</tbody>
</table>

**Approach to newly diagnosed hypertension**

Hypertension is predominanty an asymptomatic condition and the diagnosis is usually made at routine examination or when a complication arises. A BP check is advisable every 5 years in adults.
The objectives of the initial evaluation of a patient with high BP readings are:

• to obtain accurate, representative BP measurements
• to identify contributory factors and any underlying cause (secondary hypertension)
• to assess other risk factors and quantify cardiovascular risk
• to detect any complications (target organ damage) that are already present
• to identify comorbidity that may influence the choice of antihypertensive therapy.

These goals are attained by a careful history, clinical examination and some simple investigations.

**Measurement of blood pressure**

A decision to embark upon antihypertensive therapy effectively commits the patient to life-long treatment, so BP readings must be as accurate as possible.

Measurements should be made to the nearest 2 mmHg, in the sitting position with the arm supported, and repeated after 5 minutes’ rest if the first recording is high (Box 18.87). To avoid spuriously high readings in obese subjects, the cuff should contain a bladder that encompasses at least two-thirds of the arm circumference.

**Home and ambulatory BP recordings**

Exercise, anxiety, discomfort and unfamiliar surroundings can all lead to a transient rise in BP. Sphygmomanometry, particularly when performed by a doctor, can cause an unrepresentative surge in BP which has been termed ‘white coat’ hypertension, and as many as 20% of patients with apparent hypertension in the clinic may have a normal BP when it is recorded by automated devices used at home. The risk of cardiovascular disease in these patients is less than that in patients with sustained hypertension but greater than that in normotensive subjects.

A series of automated ambulatory BP measurements obtained over 24 hours or longer provides a better profile than a limited number of clinic readings and correlates more closely with evidence of target organ damage than casual BP measurements. Treatment thresholds and targets (see Box 18.93, p. 611) must be adjusted downwards, however, because ambulatory BP readings are systematically lower (approximately 12/7 mmHg) than clinic measurements. The average ambulatory daytime (not 24-hour or night-time) BP should be used to guide management decisions.

Patients can measure their own BP at home using a range of commercially available semi-automatic devices. The value of such measurements is less well established and is dependent on the environment and timing of the readings measured. Home or ambulatory BP measurements are particularly helpful in patients with unusually labile BP, those with refractory hypertension, those who may have symptomatic hypotension, and those in whom white coat hypertension is suspected.

**History**

Family history, lifestyle (exercise, salt intake, smoking habit) and other risk factors should be recorded. A careful history will identify those patients with drug- or alcohol-induced hypertension and may elicit the symptoms of other causes of secondary hypertension, such as pheochromocytoma (paroxysmal headache, palpitation and sweating, p. 781) or complications such as coronary artery disease (e.g. angina, breathlessness).

**Examination**

Radio-femoral delay (coarctation of the aorta; see Fig. 18.97, p. 632), enlarged kidneys (polycystic kidney disease), abdominal bruits (renal artery stenosis) and the characteristic facies and habitus of Cushing’s syndrome are all examples of physical signs that may help to identify causes of secondary hypertension (see Box 18.86). Examination may also reveal features of important risk factors, such as central obesity and hyperlipidaemia (tendon xanthomas and so on). Most abnormal signs are due to the complications of hypertension.

Non-specific findings may include left ventricular hypertrophy (apical heave), accentuation of the aortic component of the second heart sound, and a fourth heart sound. The optic fundi are often abnormal (see Fig. 18.83 below) and there may be evidence of generalised atheroma or specific complications, such as aortic aneurysm or peripheral vascular disease.

**Target organ damage**

The adverse effects of hypertension on the organs can often be detected clinically.

**Blood vessels**

In larger arteries (> 1 mm in diameter), the internal elastic lamina is thickened, smooth muscle is hypertrophied and fibrous tissue is deposited. The vessels dilate and become tortuous, and their walls become less compliant. In smaller arteries (< 1 mm), hyaline arteriosclerosis occurs in the wall, the lumen narrows and aneurysms may develop. Widespread atheroma develops and may lead to coronary and cerebrovascular disease, particularly if other risk factors (e.g. smoking, hyperlipidaemia, diabetes) are present.

These structural changes in the vasculature often perpetuate and aggravate hypertension by increasing peripheral vascular resistance and reducing renal blood flow, thereby activating the renin–angiotensin–aldosterone axis (p. 547).

Hypertension is a major risk factor in the pathogenesis of aortic aneurysm and aortic dissection.
Central nervous system

Stroke is a common complication of hypertension and may be due to cerebral haemorrhage or infarction. Carotid atheroma and TIAs are more common in hypertensive patients. Subarachnoid haemorrhage is also associated with hypertension.

Hypertensive encephalopathy is a rare condition characterised by high BP and neurological symptoms, including transient disturbances of speech or vision, paraesthesiae, disorientation, fits and loss of consciousness. Papilloedema is common. A CT scan of the brain often shows haemorrhage in and around the basal ganglia; however, the neurological deficit is usually reversible if the hypertension is properly controlled.

Retina

The optic fundi reveal a gradation of changes linked to the severity of hypertension; fundoscopy can, therefore, provide an indication of the arteriolar damage occurring elsewhere (Box 18.88).

‘Cotton wool’ exudates are associated with retinal ischaemia or infarction, and fade in a few weeks (Fig. 18.83A). ‘Hard’ exudates (small, white, dense deposits of lipid) and microaneurysms (‘dot’ haemorrhages) are more characteristic of diabetic retinopathy (see Fig. 21.12, p. 829). Hypertension is also associated with central retinal vein thrombosis (Fig. 18.83B).

Heart

The excess cardiac mortality and morbidity associated with hypertension are largely due to a higher incidence of coronary artery disease. High BP places a pressure load on the heart and may lead to left ventricular hypertrophy with a forceful apex beat and fourth heart sound. ECG or echocardiographic evidence of left ventricular hypertrophy is highly predictive of cardiovascular complications and therefore particularly useful in risk assessment.

Atrial fibrillation is common and may be due to diastolic dysfunction caused by left ventricular hypertrophy or the effects of coronary artery disease.

Severe hypertension can cause left ventricular failure in the absence of coronary artery disease, particularly when renal function, and therefore sodium excretion, are impaired.

Kidneys

Long-standing hypertension may cause proteinuria and progressive renal failure (p. 478) by damaging the renal vasculature.

Fig. 18.83 Retinal changes in hypertension. (A) Grade 4 hypertensive retinopathy showing swollen optic disc, retinal haemorrhages and multiple cotton wool spots (infarcts). (B) Central retinal vein thrombosis showing swollen optic disc and widespread fundal haemorrhage, commonly associated with systemic hypertension.

‘Malignant’ or ‘accelerated’ phase hypertension

This rare condition may complicate hypertension of any aetiology and is characterised by accelerated microvascular damage with necrosis in the walls of small arteries and arterioles (‘fibrinoid necrosis’) and by intravascular thrombosis. The diagnosis is based on evidence of high BP and rapidly progressive end organ damage, such as retinopathy (grade 3 or 4), renal dysfunction (especially proteinuria) and/or hypertensive encephalopathy (see above). Left ventricular failure may occur and, if this is untreated, death occurs within months.

Investigations

All hypertensive patients should undergo a limited number of investigations (Box 18.89). Additional investigations are appropriate in selected patients (Box 18.90).

Management

Quantification of cardiovascular risk

The sole objective of antihypertensive therapy is to reduce the incidence of adverse cardiovascular events, particularly coronary artery disease, stroke and heart failure. Randomised controlled trials have demonstrated that antihypertensive therapy can reduce the incidence of stroke and, to a lesser extent, coronary artery disease (Box 18.91). The relative benefits (approximately 30% reduction in risk of stroke and 20% reduction in risk of coronary artery disease) are similar in all patient groups,
so the absolute benefit of treatment (total number of events prevented) is greatest in those at highest risk. For example, to extrapolate from the Medical Research Council (MRC) Mild Hypertension Trial (1985), 566 young patients would have to be treated with bendrofluamethiazide for 1 year to prevent 1 stroke, while in the MRC trial of antihypertensive treatment in the elderly (1992), 1 stroke was prevented for every 286 patients treated for 1 year.

A formal estimate of absolute cardiovascular risk, which takes account of all the relevant risk factors, may help to determine whether the likely benefits of therapy will outweigh its costs and hazards. A variety of risk algorithms are available for this purpose (see Fig. 18.62, p. 582). Most of the excess morbidity and mortality associated with hypertension is attributable to coronary artery disease and many treatment guidelines are therefore based on estimates of the 10-year coronary artery disease risk. Total cardiovascular risk can be estimated by multiplying coronary artery disease risk by 4/3 (i.e. if coronary artery disease risk is 30%, cardiovascular risk is 40%). The value of this approach can be illustrated by comparing the two hypothetical cases on page 582.

Threshold for intervention

Systolic BP and diastolic BP are both powerful predictors of cardiovascular risk. The British Hypertension Society management guidelines therefore utilise both readings, and treatment should be initiated if they exceed the given threshold (Fig. 18.84).
loop diuretics, such as furosemide (40 mg daily) or bumetanide (1 mg daily), have few advantages over thiazides in the treatment of hypertension, unless there is substantial renal impairment or they are used in conjunction with an ACE inhibitor.

**ACE inhibitors.** ACE inhibitors (e.g. enalapril 20 mg daily, ramipril 5–10 mg daily or lisinopril 10–40 mg daily) inhibit the conversion of angiotensin I to angiotensin II and are usually well tolerated. They should be used with particular care in patients with impaired renal function or renal artery stenosis because they can reduce the filtration pressure in the glomeruli and precipitate renal failure. Electrolytes and creatinine should be checked before and 1–2 weeks after commencing therapy. Side-effects include first-dose hypotension, cough, rash, hyperkalaemia and renal dysfunction.

**Angiotensin receptor blockers.** Angiotensin receptor blockers (e.g. irbesartan 150–300 mg daily, valsartan

---

**18.93 Optimal target blood pressures**

<table>
<thead>
<tr>
<th>Age</th>
<th>Clinic BP (mmHg)</th>
<th>Ambulatory or home BP (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 80 yrs</td>
<td>&lt; 140/90</td>
<td>&lt; 135/85</td>
</tr>
<tr>
<td>≥ 80 yrs</td>
<td>&lt; 150/90</td>
<td>&lt; 140/85</td>
</tr>
</tbody>
</table>

1Both systolic and diastolic values should be attained.
2Average BP during waking hours.

---

**Thiazide and other diuretics.** The mechanism of action of these drugs is incompletely understood and it may take up to a month for the maximum effect to be observed. An appropriate daily dose is 2.5 mg bendroflumethiazide or 0.5 mg cyclopenthiazide. More potent

---

**Antihypertensive drugs**

**Thiazide and other diuretics.** The mechanism of action of these drugs is incompletely understood and it may take up to a month for the maximum effect to be observed. An appropriate daily dose is 2.5 mg bendroflumethiazide or 0.5 mg cyclopenthiazide. More potent
40–160 mg daily) block the angiotensin II type I receptor and have similar effects to ACE inhibitors; however, they do not cause cough and are better tolerated.

**Calcium channel antagonists.** The dihydropyridines (e.g. amlodipine 5–10 mg daily, nifedipine 30–90 mg daily) are effective and usually well-tolerated antihypertensive drugs that are particularly useful in older people. Side-effects include flushing, palpitations and fluid retention. The rate-limiting calcium channel antagonists (e.g. diltiazem 200–300 mg daily, verapamil 240 mg daily) can be useful when hypertension coexists with angina but they may cause bradycardia. The main side-effect of verapamil is constipation.

**Beta-blockers.** These are no longer used as first-line anti hypertensive therapy, except in patients with another indication for the drug (e.g. angina). Metoprolol (100–200 mg daily), atenolol (50–100 mg daily) and bisoprolol (5–10 mg daily) preferentially block cardiac β1-adrenoceptors, as opposed to the β2-adrenoceptors that mediate vasodilatation and bronchodilatation.

**Labetalol and carvedilol.** Labetalol (200 mg–2.4 g daily in divided doses) and carvedilol (6.25–25 mg twice daily) are combined β- and α-adrenoceptor antagonists which are sometimes more effective than pure β-blockers. Labetalol can be used as an infusion in malignant phase hypertension (see below).

**Other drugs.** A variety of vasodilators may be used. These include the α1-adrenoceptor antagonists (α-blockers), such as prazosin (0.5–20 mg daily in divided doses), indoramin (25–100 mg twice daily) and doxazosin (1–16 mg daily), and drugs that act directly on vascular smooth muscle, such as hydralazine (25–100 mg twice daily) and minoxidil (10–50 mg daily). Side-effects include first-dose and postural hypotension, headache, tachycardia and fluid retention. Minoxidil also causes increased facial hair and is therefore unsuitable for female patients.

**Choice of antihypertensive drug**

Trials that have compared thiazides, calcium antagonists, ACE inhibitors and angiotensin receptor blockers have not shown consistent differences in outcome, efficacy, side-effects or quality of life. Beta-blockers, which previously featured as first-line therapy in guidelines, have a weaker evidence base (see Box 18.91). The choice of antihypertensive therapy is initially dictated by the patient’s age and ethnic background, although cost and convenience will influence the exact drug and preparation used. Response to initial therapy and side-effects guides subsequent treatment. Comorbid conditions also have an influence on initial drug selection (Box 18.94); for example, a β-blocker might be the most appropriate treatment for a patient with angina. Thiazide diuretics and dihydropyridine calcium channel antagonists are the most suitable drugs for treatment in older people.

Although some patients can be treated with a single antihypertensive drug, a combination of drugs is often required to achieve optimal BP control (Fig. 18.85). Combination therapy may be desirable for other reasons; for example, low-dose therapy with two drugs may produce fewer unwanted effects than treatment with the maximum dose of a single drug. Some drug combinations have complementary or synergistic actions; for example, thiazides increase activity of the renin-angiotensin system, while ACE inhibitors block it.

---

**Emergency treatment of accelerated phase or malignant hypertension**

In accelerated phase hypertension, lowering BP too quickly may compromise tissue perfusion (due to altered autoregulation) and can cause cerebral damage, including occipital blindness, and precipitate coronary or renal insufficiency. Even in the presence of cardiac failure or hypertensive encephalopathy, a controlled reduction to a level of about 150/90 mmHg over a period of 24–48 hours is ideal.

In most patients, it is possible to avoid parenteral therapy and bring BP under control with bed rest and oral drug therapy. Intravenous or intramuscular labetalol (2 mg/min to a maximum of 200 mg), intravenous glyceryl trinitrate (0.6–1.2 mg/hr), intramuscular hydralazine (5 or 10 mg aliquots repeated at half-hourly intervals) and intravenous sodium nitroprusside (0.3–1.0 µg/kg body weight/min) are all effective but require careful supervision, preferably in a high-dependency unit.

**Refactory hypertension**

The common causes of treatment failure in hypertension are non-adherence to drug therapy, inadequate therapy, and failure to recognise an underlying cause, such as...
renal artery stenosis or phaeochromocytoma; of these, the first is by far the most prevalent. There is no easy solution to compliance problems but simple treatment regimens, attempts to improve rapport with the patient and careful supervision may all help.

**Adjuvant drug therapy**

- **Aspirin.** Antiplatelet therapy is a powerful means of reducing cardiovascular risk but may cause bleeding, particularly intracerebral haemorrhage, in a small number of patients. The benefits are thought to outweigh the risks in hypertensive patients aged 50 years or over who have well-controlled BP and either target organ damage, diabetes or a 10-year coronary artery disease risk of at least 15% (or 10-year cardiovascular disease risk of at least 20%).
- **Statins.** Treating hyperlipidaemia can produce a substantial reduction in cardiovascular risk. These drugs are strongly indicated in patients who have established vascular disease, or hypertension with a high (at least 20% in 10 years) risk of developing cardiovascular disease (p. 583).

### Diseases of the heart valves

A diseased valve may be narrowed (stenosed) or may fail to close adequately, and thus permit regurgitation of blood. ‘Incompetence’ is a less precise term for regurgitation or reflux, and should be avoided. Box 18.95 gives the principal causes of valve disease.

Doppler echocardiography is the most useful technique for assessing valvular heart disease (p. 536) but may also detect minor and even ‘physiological’ abnormalities, e.g. trivial mitral regurgitation. Disease of the heart valves may progress with time and selected patients may also detect minor and even ‘physiological’ abnor-

---

### Table: The influence of comorbidity on the choice of antihypertensive drug therapy

<table>
<thead>
<tr>
<th>Class of drug</th>
<th>Compelling indications</th>
<th>Possible indications</th>
<th>Caution</th>
<th>Compelling contraindications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>α-blockers</strong></td>
<td>Benign prostatic hypertrophy</td>
<td>–</td>
<td>Postural hypotension, heart failure</td>
<td>Urinary incontinence</td>
</tr>
<tr>
<td><strong>ACE inhibitors</strong></td>
<td>Heart failure Left ventricular dysfunction, post-MI or established coronary artery disease Type 1 diabetic nephropathy Secondary stroke prevention</td>
<td>Chronic renal disease Type 2 diabetic nephropathy</td>
<td>Renal impairment Peripheral vascular disease</td>
<td>Pregnancy Renovascular disease</td>
</tr>
<tr>
<td><strong>Angiotensin II receptor blockers</strong></td>
<td>ACE inhibitor intolerance Type 2 diabetic nephropathy Hypertension with left ventricular hypertrophy Heart failure in ACE-intolerant patients, after MI</td>
<td>Left ventricular dysfunction after MI Intolerance of other antihypertensive drugs Proteinuric renal disease, chronic renal disease Heart failure</td>
<td>Renal impairment Peripheral vascular disease</td>
<td>Pregnancy</td>
</tr>
<tr>
<td><strong>β-blockers</strong></td>
<td>MI, angina Heart failure</td>
<td>–</td>
<td>Heart failure Peripheral vascular disease Diabetes (except with coronary artery disease)</td>
<td>Asthma or chronic obstructive pulmonary disease Heart block</td>
</tr>
<tr>
<td><strong>Calcium channel blockers (dihydropyridine)</strong></td>
<td>Older patients, isolated systolic hypertension</td>
<td>Angina</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>Calcium channel blockers (rate-limiting)</strong></td>
<td>Angina</td>
<td>Older patients</td>
<td>Combination with β-blockade</td>
<td>Atrioventricular block, heart failure</td>
</tr>
<tr>
<td><strong>Thiazides or thiazide-like diuretics</strong></td>
<td>Older patients, isolated systolic hypertension, heart failure, secondary stroke prevention</td>
<td>–</td>
<td>–</td>
<td>Gout</td>
</tr>
</tbody>
</table>

1. In heart failure when used as monotherapy. 2. ACE inhibitors or angiotensin II receptor blockers may be beneficial in chronic renal failure and renovascular disease but should be used with caution, close supervision and specialist advice when there is established and significant renal impairment. 3. Caution with ACE inhibitors and angiotensin II receptor blockers in peripheral vascular disease because of association with renovascular disease. 4. In combination with a thiazide or thiazide-like diuretic. 5. β-blockers are used increasingly to treat stable heart failure but may worsen acute heart failure. 6. Thiazides or thiazide-like diuretics may sometimes be necessary to control BP in people with a history of gout.
antibiotic prophylaxis at times of bacteraemia, e.g. dental extraction, is no longer recommended.

**Rheumatic heart disease**

**Acute rheumatic fever**

*Incidence and pathogenesis*

Acute rheumatic fever usually affects children (most commonly between 5 and 15 years) or young adults, and has become very rare in Western Europe and North America. However, it remains endemic in parts of Asia, Africa and South America, with an annual incidence in some countries of more than 100 per 100,000, and is the most common cause of acquired heart disease in childhood and adolescence.

The condition is triggered by an immune-mediated delayed response to infection with specific strains of group A streptococci, which have antigens that may cross-react with cardiac myosin and sarcolemmal membrane protein. Antibodies produced against the streptococcal antigens cause inflammation in the endocardium, myocardium and pericardium, as well as the joints and skin. Histologically, fibrinoid degeneration is seen in the collagen of connective tissues. Aschoff nodules are pathognomonic and occur only in the heart. They are composed of multinucleated giant cells surrounded by macrophages and T lymphocytes, and are not seen until the subacute or chronic phases of rheumatic carditis.

**Clinical features**

Acute rheumatic fever is a multisystem disorder that usually presents with fever, anorexia, lethargy and joint pain, 2–3 weeks after an episode of streptococcal pharyngitis. There may, however, be no history of sore throat. Arthritis occurs in approximately 75% of patients. Other features include rashes, carditis and neurological changes (Fig. 18.86). The diagnosis, made using the revised Jones criteria (Box 18.96), is based upon two or more major manifestations, or one major and two or more minor manifestations, along with evidence of preceding streptococcal infection. Only about 25% of patients will have a positive culture for group A streptococcus at the time of diagnosis because there is a latent period between infection and presentation. Serological evidence of recent infection with a raised antistreptolysin O (ASO) antibody titre is helpful. A presumptive diagnosis of acute rheumatic fever can be made without evidence of preceding streptococcal infection in cases of isolated chorea or pancarditis, if other causes for these have been excluded. In cases of established rheumatic heart disease or prior rheumatic fever, a diagnosis of acute rheumatic fever can be made based only on the presence of multiple minor criteria and evidence of preceding group A streptococcal pharyngitis.

**Carditis**

A ‘pancarditis’ involves the endocardium, myocardium and pericardium to varying degrees. Its incidence declines with increasing age, ranging from 90% at 3 years to around 30% in adolescence. It may manifest as breathlessness (due to heart failure or pericardial effusion), palpitations or chest pain (usually due to pericarditis or pancarditis). Other features include tachycardia,
cardiac enlargement and new or changed murmurs. A soft systolic murmur due to mitral regurgitation is very common. A soft mid-diastolic murmur (the Carey Coombs murmur) is typically due to valvulitis, with nodules forming on the mitral valve leaflets. Aortic regurgitation occurs in 50% of cases but the tricuspid and pulmonary valves are rarely involved. Pericarditis may cause chest pain, a pericardial friction rub and precordial tenderness. Cardiac failure may be due to myocardial dysfunction or valvular regurgitation. ECG changes commonly include ST and T wave changes. Conduction defects sometimes occur and may cause syncope.

Arthritis

This is the most common major manifestation and occurs early when streptococcal antibody titres are high. An acute painful asymmetric and migratory inflammation of the large joints typically affects the knees, ankles, elbows and wrists. The joints are involved in quick succession and are usually red, swollen and tender for between a day and 4 weeks. The pain characteristically responds to aspirin; if not, the diagnosis is in doubt.

Skin lesions

Erythema marginatum occurs in less than 5% of patients. The lesions start as red macules that fade in the centre but remain red at the edges, and occur mainly on the trunk and proximal extremities but not the face. The resulting red rings or ‘margins’ may coalesce or overlap (see Fig. 18.86). Subcutaneous nodules occur in 5–7% of patients. They are small (0.5–2.0 cm), firm and painless, and are best felt over extensor surfaces of bone or tendons. They typically appear more than 3 weeks after the onset of other manifestations and therefore help to confirm rather than make the diagnosis. Other systemic manifestations are rare but include pleurisy, pleural effusion and pneumonia.

Sydenham’s chorea (St Vitus dance)

This is a late neurological manifestation that appears at least 3 months after the episode of acute rheumatic fever, when all the other signs may have disappeared. It occurs in up to one-third of cases and is more common in females. Emotional lability may be the first feature and is typically followed by purposeless, involuntary, choreiform movements of the hands, feet or face. Speech may be explosive and halting. Spontaneous recovery usually occurs within a few months. Approximately one-quarter of affected patients will go on to develop chronic rheumatic valve disease.

Investigations

The ESR and CRP are useful for monitoring progress of the disease (Box 18.97). Positive throat swab cultures are obtained in only 10–25% of cases. ASO titres are normal in one-fifth of adult cases of rheumatic fever and most cases of chorea. Echocardiography typically shows mitral regurgitation with dilatation of the mitral annulus and prolapse of the anterior mitral leaflet, and may also show aortic regurgitation and pericardial effusion.

Management of the acute attack

A single dose of benzyl penicillin (1.2 million U IM) or oral phenoxymethylpenicillin (250 mg 4 times daily for 10 days) should be given on diagnosis to eliminate any residual streptococcal infection. If the patient is penicillin-allergic, erythromycin or a cephalosporin can be used. Treatment is then directed towards limiting cardiac damage and relieving symptoms.

Bed rest and supportive therapy

Bed rest is important, as it lessens joint pain and reduces cardiac workload. The duration should be guided by symptoms, along with temperature, leucocyte count and ESR, and should be continued until these have settled. Patients can then return to normal physical activity but strenuous exercise should be avoided in those who have had carditis.

Cardiac failure should be treated as necessary. Some patients, particularly those in early adolescence, develop a fulminant form of the disease with severe mitral regurgitation and, sometimes, concomitant aortic regurgitation. If heart failure in these cases does not respond to medical treatment, valve replacement may be necessary and is often associated with a dramatic decline in rheumatic activity. AV block is seldom progressive and pacemaker insertion rarely needed.

Aspirin

This usually relieves the symptoms of arthritis rapidly and a response within 24 hours helps confirm the diagnosis. A reasonable starting dose is 60 mg/kg body weight/day, divided into six doses. In adults, 100 mg/kg per day may be needed up to the limits of tolerance or a maximum of 8 g per day. Mild toxicity includes nausea, tinnitus and deafness; vomiting, tachypnoea and acidosis are more serious. Aspirin should be continued until the ESR has fallen, and then gradually tailed off.

Corticosteroids

These produce more rapid symptomatic relief than aspirin and are indicated in cases with carditis or severe arthritis. There is no evidence that long-term steroids are beneficial. Prednisolone (1.0–2.0 mg/kg per day in divided doses) should be continued until the ESR is normal, and then tailed off.

Secondary prevention

Patients are susceptible to further attacks of rheumatic fever if another streptococcal infection occurs, and long-term prophylaxis with penicillin should be given as benzathine penicillin (1.2 million U IM monthly), if
compliance is in doubt, or oral phenoxymethylpenicillin (250 mg twice daily). Sulfadiazine or erythromycin may be used if the patient is allergic to penicillin; sulphonamides prevent infection but are not effective in the eradication of group A streptococci. Further attacks of rheumatic fever are unusual after the age of 21, when treatment may be stopped. However, it should be extended if an attack has occurred in the last 5 years, or if the patient lives in an area of high prevalence or has an occupation (e.g. teaching) with high exposure to streptococcal infection. In those with residual heart disease, prophylaxis should continue until 10 years after the last episode or 40 years of age, whichever is later. Long-term antibiotic prophylaxis prevents another attack of acute rheumatic fever but does not protect against infective endocarditis.

**Chronic rheumatic heart disease**

Chronic valvular heart disease develops in at least half of those affected by rheumatic fever with carditis. Two-thirds of cases occur in women. Some episodes of rheumatic fever pass unrecognised and it is only possible to elicit a history of rheumatic fever or chorea in about half of all patients with chronic rheumatic heart disease.

The mitral valve is affected in more than 90% of cases; the aortic valve is the next most frequently involved, followed by the tricuspid and then the pulmonary valve. Isolated mitral stenosis accounts for about 25% of all cases, and an additional 40% have mixed mitral stenosis and regurgitation. Valve disease may be asymptomatic during fulminant forms of acute rheumatic fever but may remain asymptomatic for many years.

**Pathology**

The main pathological process in chronic rheumatic heart disease is progressive fibrosis. The heart valves are predominantly affected but involvement of the pericardium and myocardium may contribute to heart failure and conduction disorders. Fusion of the mitral valve commissures and shortening of the chordae tendineae may lead to mitral stenosis with or without regurgitation. Similar changes in the aortic and tricuspid valves produce distortion and rigidity of the cusps, leading to stenosis and regurgitation. Once a valve has been damaged, the altered haemodynamic stresses perpetuate and extend the damage, even in the absence of a continuing rheumatic process.

**Mitraal valve disease**

**Mitraal stenosis**

**Aetiology and pathophysiology**

Mitraal stenosis is almost always rheumatic in origin, although in older people it can be caused by heavy calcification of the mitral valve apparatus. There is also a rare form of congenital mitral stenosis.

In rheumatic mitral stenosis, the mitral valve orifice is slowly diminished by progressive fibrosis, calcification of the valve leaflets, and fusion of the cusps and subvalvular apparatus. The flow of blood from LA to LV is restricted and left atrial pressure rises, leading to pulmonary venous congestion and breathlessness. There is dilatation and hypertrophy of the LA, and left ventricular filling becomes more dependent on left atrial contraction.

Any increase in heart rate shortens diastole when the mitral valve is open and produces a further rise in left atrial pressure. Situations that demand an increase in cardiac output also increase left atrial pressure, so exercise and pregnancy are poorly tolerated.

The mitral valve orifice is normally about 5 cm² in diastole and may be reduced to 1 cm² in severe mitral stenosis. Patients usually remain asymptomatic until the stenosis is less than 2 cm². Reduced lung compliance, due to chronic pulmonary venous congestion, contributes to breathlessness, and a low cardiac output may cause fatigue.

Atrial fibrillation due to progressive dilatation of the LA is very common. Its onset often precipitates pulmonary oedema because the accompanying tachycardia and loss of atrial contraction lead to marked haemodynamic deterioration with a rapid rise in left atrial pressure. In contrast, a more gradual rise in left atrial pressure tends to cause an increase in pulmonary vascular resistance, which leads to pulmonary hypertension that may protect the patient from pulmonary oedema. Pulmonary hypertension leads to right ventricular hypertrophy and dilatation, tricuspid regurgitation and right heart failure.

Fewer than 20% of patients remain in sinus rhythm; many of these have a small fibrotic LA and severe pulmonary hypertension.

**Clinical features**

Effort-related dyspnoea is usually the dominant symptom (Box 18.98). Exercise tolerance typically diminishes very slowly over many years and patients often do not appreciate the extent of their disability. Eventually, symptoms occur at rest. Acute pulmonary oedema or pulmonary hypertension can lead to haemoptysis. All patients with mitral stenosis, and particularly those with atrial fibrillation, are at risk from left atrial thrombosis and systemic thromboembolism. Prior to the advent of anticoagulant therapy, emboli caused one-quarter of all deaths.

**18.98 Clinical features (and their causes) in mitral stenosis**

**Symptoms**

- Breathlessness (pulmonary congestion)
- Fatigue (low cardiac output)
- Oedema, ascites (right heart failure)
- Palpitation (atrial fibrillation)
- Haemoptysis (pulmonary congestion, pulmonary embolism)
- Cough (pulmonary congestion)
- Chest pain (pulmonary hypertension)
- Thromboembolic complications (e.g. stroke, ischaemic limb)

**Signs**

- Atrial fibrillation
- Mitral facies
- Auscultation
  - Loud first heart sound, opening snap
  - Mid-diastolic murmur
- Crepitations, pulmonary oedema, effusions (raised pulmonary capillary pressure)
- RV heave, loud P₂ (pulmonary hypertension)
The physical signs of mitral stenosis are often found before symptoms develop and their recognition is of particular importance in pregnancy. The forces that open and close the mitral valve increase as left atrial pressure rises. The first heart sound (S1) is therefore loud and can be palpable (tapping apex beat). An opening snap may be audible and moves closer to the second sound (S2) as the stenosis becomes more severe and left atrial pressure rises. However, the first heart sound and opening snap may be inaudible if the valve is heavily calcified.

Turbulent flow produces the characteristic low-pitched mid-diastolic murmur and sometimes a thrill (Fig. 18.87). The murmur is accentuated by exercise and during atrial systole (pre-systolic accentuation). Early in the disease, a pre-systolic murmur may be the only auscultatory abnormality but, in patients with symptoms, the murmur extends from the opening snap to the first heart sound. Coexisting mitral regurgitation causes a pansystolic murmur that radiates towards the axilla.

Pulmonary hypertension may ultimately lead to right ventricular hypertrophy and dilatation with secondary tricuspid regurgitation, which causes a systolic murmur and giant ‘v waves’ in the venous pulse.

**Investigations**

The ECG may show either atrial fibrillation or bifid P waves (P mitrale) associated with left atrial hypertrophy (Box 18.99). A typical chest X-ray is shown in Figure 18.9 (p. 535). Doppler echocardiography provides the definitive evaluation of mitral stenosis (see Fig. 18.87). Cardiac catheterisation is used to assess coexisting conditions.
**CARDIOVASCULAR DISEASE**

**Management**

Patients with minor symptoms should be treated medically. Intervention by balloon valvuloplasty, mitral valvotomy or mitral valve replacement should be considered if the patient remains symptomatic despite medical treatment or if pulmonary hypertension develops.

**Medical management**

This consists of anticoagulation to reduce the risk of systemic embolism, ventricular rate control (digoxin, β-blockers or rate-limiting calcium antagonists) in atrial fibrillation, and diuretic therapy to control pulmonary congestion. Antibiotic prophylaxis against infective endocarditis is no longer routinely recommended.

*Mitral balloon valvuloplasty and valve replacement*

Valvuloplasty is the treatment of choice if specific criteria are fulfilled (Box 18.100 and Fig. 18.67, p. 587), although surgical closed or open mitral valvotomy is an acceptable alternative. Patients who have undergone mitral valvuloplasty or valvotomy should be followed up at 1–2-yearly intervals because re-stenosis may occur. Clinical symptoms and signs are a guide to the severity of mitral re-stenosis but Doppler echocardiography provides a more accurate assessment.

Valve replacement is indicated if there is substantial mitral reflux or if the valve is rigid and calcified (p. 629).

---

### 18.100 Criteria for mitral valvuloplasty*

- Significant symptoms
- Isolated mitral stenosis
- No (or trivial) mitral regurgitation
- Mobile, non-calcified valve/subvalve apparatus on echo
- LA free of thrombus

*For comprehensive guidelines on valvular heart disease, see www.acc.org

---

**Mitral regurgitation**

**Aetiology and pathophysiology**

Rheumatic disease is the principal cause in countries where rheumatic fever is common, but elsewhere, including in the UK, other causes are more important (Box 18.101). Mitral regurgitation may also follow mitral valvotomy or valvuloplasty.

Chronic mitral regurgitation causes gradual dilatation of the LA with little increase in pressure and therefore relatively few symptoms. Nevertheless, the LV dilates slowly and the left ventricular diastolic and left atrial pressures gradually increase as a result of chronic volume overload of the LV. In contrast, acute mitral regurgitation causes a rapid rise in left atrial pressure (because left atrial compliance is normal) and marked symptomatic deterioration.

*Mitral valve prolapse*

This is also known as ‘floppy’ mitral valve and is one of the more common causes of mild mitral regurgitation (Fig. 18.88). It is caused by congenital anomalies or degenerative myxomatous changes, and is sometimes a feature of connective tissue disorders such as Marfan’s syndrome (p. 603).

In its mildest forms, the valve remains competent but bulges back into the atrium during systole, causing a mid-systolic click but no murmur. In the presence of a regurgitant valve, the click is followed by a late systolic murmur, which lengthens as the regurgitation becomes more severe. A click is not always audible and the physical signs may vary with both posture and respiration. Progressive elongation of the chordae tendineae leads to increasing mitral regurgitation, and if chordal rupture occurs, regurgitation suddenly becomes severe. This is rare before the fifth or sixth decade of life.

Mitral valve prolapse is associated with a variety of typically benign arrhythmias, atypical chest pain and a very small risk of embolic stroke or TIA. Nevertheless, the overall long-term prognosis is good.

*Other causes of mitral regurgitation*

Mitral valve function depends on the chordae tendineae and their papillary muscles; dilatation of the LV distorts the geometry of these and may cause mitral regurgitation (see Box 18.101). Dilated cardiomyopathy and heart failure from coronary artery disease are common causes of so-called ‘functional’ mitral regurgitation. Endocarditis is an important cause of acute mitral regurgitation.

**Clinical features**

Symptoms depend on how suddenly the regurgitation develops (Box 18.102). Chronic mitral regurgitation produces a symptom complex that is similar to that of mitral stenosis, but sudden-onset mitral regurgitation usually presents with acute pulmonary oedema.

The regurgitant jet causes an apical systolic murmur (see Fig. 18.88), which radiates into the axilla and may be accompanied by a thrill. Increased forward flow through the mitral valve causes a loud third heart sound

---

### 18.102 Clinical features (and their causes) in mitral regurgitation

**Symptoms**

- Dyspnoea (pulmonary venous congestion)
- Fatigue (low cardiac output)
- Palpitation (atrial fibrillation, increased stroke volume)
- Oedema, ascites (right heart failure)

**Signs**

- Atrial fibrillation/flutter
- Cardiomegaly: displaced hyperdynamic apex beat
- Apical pansystolic murmur ± thrill
- Soft S1, apical S3
- Signs of pulmonary venous congestion (crepitations, pulmonary oedema, effusions)
- Signs of pulmonary hypertension and right heart failure
Fig. 18.8 Mitral regurgitation: murmur and systolic wave in left atrial pressure. The first sound is normal or soft and merges with a pansystolic murmur (PSM) extending to the second heart sound. A third heart sound occurs with severe regurgitation. A transoesophageal echocardiogram shows mitral valve prolapse, with one leaflet bulging towards the LA (arrow). This results in a jet of mitral regurgitation on colour Doppler (arrow).

and even a short mid-diastolic murmur. The apex beat feels active and rocking due to left ventricular volume overload and is usually displaced to the left as a result of left ventricular dilatation.

**Investigations**

Atrial fibrillation is common, as a consequence of atrial dilatation. At cardiac catheterisation (Box 18.103), the severity of mitral regurgitation can be assessed by left ventriculography and by the size of the v (systolic) waves in the left atrial or pulmonary artery wedge pressure trace.

**Management**

Mitrail regurgitation of moderate severity can be treated medically (Box 18.104). In all patients with mitral regurgitation, high afterload may worsen the degree of regurgitation, and hypertension should be treated with vasodilators, such as ACE inhibitors. Patients should be reviewed at regular intervals because worsening symptoms, progressive cardiomegaly or echocardiographic evidence of deteriorating left ventricular function are indications for mitral valve replacement or repair. Mitral

<table>
<thead>
<tr>
<th>18.103 Investigations in mitral regurgitation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECG</strong></td>
</tr>
<tr>
<td>• Left atrial hypertrophy (if not in atrial fibrillation)</td>
</tr>
<tr>
<td>• Left ventricular hypertrophy</td>
</tr>
<tr>
<td><strong>Chest X-ray</strong></td>
</tr>
<tr>
<td>• Enlarged LA</td>
</tr>
<tr>
<td>• Enlarged LV</td>
</tr>
<tr>
<td>• Pulmonary venous congestion</td>
</tr>
<tr>
<td>• Pulmonary oedema (if acute)</td>
</tr>
<tr>
<td><strong>Echo</strong></td>
</tr>
<tr>
<td>• Dilated LA, LV</td>
</tr>
<tr>
<td>• Dynamic LV (unless myocardial dysfunction predominates)</td>
</tr>
<tr>
<td>• Structural abnormalities of mitral valve (e.g. prolapse)</td>
</tr>
<tr>
<td><strong>Doppler</strong></td>
</tr>
<tr>
<td>• Detects and quantifies regurgitation</td>
</tr>
<tr>
<td><strong>Cardiac catheterisation</strong></td>
</tr>
<tr>
<td>• Dilated LA, dilated LV, mitral regurgitation</td>
</tr>
<tr>
<td>• Pulmonary hypertension</td>
</tr>
<tr>
<td>• Coexisting coronary artery disease</td>
</tr>
</tbody>
</table>
valve repair is used to treat mitral valve prolapse and offers many advantages when compared to mitral valve replacement, such that it is now advocated for severe regurgitation, even in asymptomatic patients, because results are excellent and early repair prevents irreversible left ventricular damage. Mitral regurgitation often accompanies the ventricular dilatation and dysfunction that are concomitants of coronary artery disease. If such patients are to undergo coronary bypass graft surgery, it is common practice to repair the valve and restore mitral valve function by inserting an annuloplasty ring to overcome annular dilatation and to bring the valve leaflets closer together. It can be difficult, however, to determine whether it is the ventricular dilatation or the mitral regurgitation that is the predominant problem. If ventricular dilatation is the underlying cause of mitral regurgitation, then mitral valve repair or replacement may actually worsen ventricular function, as the ventricle can no longer empty into the low-pressure LA.

Aortic valve disease

Aortic stenosis

Aetiology and pathophysiology

The likely aetiology depends on the age of the patient (Box 18.105). In congenital aortic stenosis, obstruction is present from birth or becomes apparent in infancy. With bicuspid aortic valves, obstruction may take years to develop as the valve becomes fibrotic and calcified. The aortic valve is the second most frequently affected by rheumatic fever and, commonly, both the aortic and mitral valves are involved. In older people, a structurally normal tricuspid aortic valve may be affected by fibrosis and calcification, in a process that is histologically similar to that of atherosclerosis affecting the arterial wall. Haemodynamically significant stenosis develops slowly, typically occurring at 30–60 years in those with rheumatic disease, 50–60 in those with bicuspid aortic valves and 70–90 in those with degenerative calcific disease.

Cardiac output is initially maintained at the cost of a steadily increasing pressure gradient across the aortic valve. The LV becomes increasingly hypertrophied and coronary blood flow may then be inadequate; patients may therefore develop angina, even in the absence of concomitant coronary disease. The fixed outflow obstruction limits the increase in cardiac output required on exercise. Eventually, the LV can no longer overcome the outflow tract obstruction and pulmonary oedema supervenes. In contrast to patients with mitral stenosis, which tends to progress very slowly, those with aortic stenosis typically remain asymptomatic for many years but deteriorate rapidly when symptoms develop, and death usually ensues within 3–5 years of these.

Clinical features

Aortic stenosis is commonly picked up in asymptomatic patients at routine clinical examination but the three cardinal symptoms are angina, breathlessness and syncope (Box 18.106). Angina arises because of the increased demands of the hypertrophied LV working against the high-pressure outflow tract obstruction, leading to a mismatch between oxygen demand and supply, but may also be due to coexisting coronary artery disease, especially in old age, when it affects over 50% of patients. Exertional breathlessness suggests cardiac decompensation as a consequence of the excessive pressure overload placed on the LV. Syncope usually occurs on exertion when cardiac output fails to rise to meet demand, leading to a fall in BP.

The characteristic clinical signs of severe aortic stenosis are shown in Box 18.106. A harsh ejection systolic murmur radiates to the neck, with a soft second heart sound, particularly in those with calcific valves. The murmur is often likened to a saw cutting wood and may (especially in older patients) have a musical quality like the ‘mew’ of a seagull (Fig. 18.89). The severity of aortic stenosis may be difficult to gauge clinically, as older patients with a non-compliant ‘stiff’ arterial system may have an apparently normal carotid upstroke in the presence of severe aortic stenosis. Milder degrees of stenosis may be difficult to distinguish from aortic sclerosis, in which the valve is thickened or calcified but not obstructed. A careful examination should be made for other valve lesions, particularly in rheumatic heart disease, when there is frequently concomitant mitral valve disease.

**Cardiovascular Disease**

**18.104 Medical management of mitral regurgitation**

- Diuretics
- Vasodilators, e.g. ACE inhibitors
- Digoxin if atrial fibrillation is present
- Anticoagulants if atrial fibrillation is present

**18.105 Causes of aortic stenosis**

- Infants, children, adolescents
  - Congenital aortic stenosis
  - Congenital subvalvular aortic stenosis
  - Congenital supravalvular aortic stenosis
- Young adults to middle-aged
  - Calcification and fibrosis of congenitally bicuspid aortic valve
  - Rheumatic aortic stenosis
- Middle-aged to elderly
  - Senile degenerative aortic stenosis
  - Calcification of bicuspid valve
  - Rheumatic aortic stenosis

**18.106 Clinical features of aortic stenosis**

**Symptoms**

- Mild or moderate stenosis: usually asymptomatic
- Exertional dyspnoea
- Angina
- Exertional syncope
- Sudden death
- Episodes of acute pulmonary oedema

**Signs**

- Ejection systolic murmur
- Slow-rising carotid pulse
- Thrusting apex beat (LV pressure overload)
- Narrow pulse pressure
- Signs of pulmonary venous congestion (e.g. crepitations)
Echocardiography demonstrates restricted valve opening (Fig. 18.91) and Doppler assessment permits calculation of the systolic gradient across the aortic valve, from which the severity of stenosis can be assessed (see Fig. 18.11, p. 536). In patients with an impaired LV, velocities across the aortic valve may be diminished because of a reduced stroke volume, while when aortic regurgitation is present, velocities are increased because of an increased stroke volume. In these circumstances, aortic valve area calculated from Doppler measurements is a more accurate assessment of severity. CT and MRI are useful in assessing the degree of valve calcification and stenosis, respectively, but are rarely necessary.

Management

Irrespective of the severity of valve stenosis, patients with asymptomatic aortic stenosis have a good immediate prognosis and conservative management is appropriate. Such patients should be kept under review, as the development of angina, syncope, symptoms of low cardiac output or heart failure has a poor prognosis and is an indication for prompt surgery. In practice, patients with moderate or severe stenosis are evaluated every 1–2 years with Doppler echocardiography to detect progression in severity; this is more rapid in older patients with heavily calcified valves.

Patients with symptomatic severe aortic stenosis should have prompt aortic valve replacement. Old age...
Cardiovascular Disease

1.8.108 Aortic stenosis in old age

- **Incidence**: the most common form of valve disease affecting the very old.
- **Symptoms**: a common cause of syncope, angina and heart failure in the very old.
- **Signs**: because of increasing stiffening in the central arteries, low pulse pressure and a slow rising pulse may not be present.
- **Surgery**: can be successful in those aged 80 yrs or more in the absence of comorbidity, but with a higher operative mortality. The prognosis without surgery is poor once symptoms have developed.
- **Valve replacement type**: a biological valve is often preferable to a mechanical one because this obviates the need for anticoagulation, and the durability of biological valves usually exceeds the patient’s anticipated life expectancy.

is not a contraindication to valve replacement and results are very good in experienced centres, even for those in their eighties (Box 18.108). Delay exposes the patient to the risk of sudden death or irreversible deterioration in ventricular function. Some patients with severe aortic stenosis deny symptoms, and if this could be due to a sedentary lifestyle, a careful exercise test may reveal symptoms on modest exertion. Aortic balloon

Fig. 18.90 Left ventricular hypertrophy. QRS complexes in limb leads have increased amplitude with a very large R wave in V₆ and S wave in V₂. There is ST depression and T-wave inversion in leads II, III, aVF, V₅ and V₆, a ‘left ventricular strain’ pattern.

Fig. 18.91 Two-dimensional echocardiogram comparing a normal subject with a patient with calcific aortic stenosis.

A Normal subject in diastole; the aortic leaflets are closed and thin, and a point of coaptation is seen (arrow). B Calcific aortic stenosis in diastole; the aortic leaflets are thick and calcified (arrow). C Normal in systole; the aortic leaflets are open (arrows). D Calcific aortic stenosis in systole; the thickened leaflets have barely moved (arrows). From Newby and Grubb 2005 – see p. 641.
Diseases of the heart valves

Valvuloplasty is useful in congenital aortic stenosis but is of no value in older patients with calcific aortic stenosis.

Anticoagulants are only required in patients who have atrial fibrillation or those who have had a valve replacement with a mechanical prosthesis.

**Aortic regurgitation**

**Aetiology and pathophysiology**

This condition is due to disease of the aortic valve cusps or dilatation of the aortic root (Box 18.109). The LV dilates and hypertrophies to compensate for the regurgitation. The stroke volume of the LV may eventually be doubled or trebled, and the major arteries are then conspicuously pulsatile. As the disease progresses, left ventricular diastolic pressure rises and breathlessness develops.

**Clinical features**

Until the onset of breathlessness, the only symptom may be an awareness of the heart beat (Box 18.110), particularly when lying on the left side, which results from the increased stroke volume. Paroxysmal nocturnal dyspnoea is sometimes the first symptom, and peripheral oedema or angina may occur. The characteristic murmur is best heard to the left of the sternum during held expiration (Fig. 18.92); a thrill is rare. A systolic murmur due to the increased stroke volume is common and does not

---

**18.109 Causes of aortic regurgitation**

**Congenital**

- Bicuspid valve or disproportionate cusps

**Acquired**

- Rheumatic disease
- Infective endocarditis
- Trauma
- Aortic dilatation (Marfan’s syndrome, aneurysm, dissection, syphilis, ankylosing spondylitis)

---

**18.110 Clinical features of aortic regurgitation**

**Symptoms**

**Mild to moderate aortic regurgitation**
- Often asymptomatic
- Awareness of heart beat, ‘palpitations’

**Severe aortic regurgitation**
- Breathlessness
- Angina

**Signs**

**Pulses**
- Large-volume or ‘collapsing’ pulse
- Low diastolic and increased pulse pressure
- Bounding peripheral pulses
- Capillary pulsation in nail beds: Quincke’s sign
- Femoral bruit (‘pistol shot’: Duroziez’s sign
- Head nodding with pulse: de Mussel’s sign

**Murmurs**
- Early diastolic murmur
- Systolic murmur (increased stroke volume)
- Austin Flint murmur (soft mid-diastolic)

**Other signs**
- Displaced, heaving apex beat (volume overload)
- Pre-systolic impulse
- Fourth heart sound
- Crepitations (pulmonary venous congestion)

---

**Fig. 18.92 Aortic regurgitation.** The early diastolic murmur is best heard at the left sternal edge and may be accompanied by an ejection systolic (‘to and fro’) murmur. The aortic arch and LV may become dilated. A Doppler echocardiogram with the regurgitant jet (arrows). Inset (Colour Doppler echo) From Newby and Grubb 2005 – see p. 641.
necessarily indicate stenosis. The regurgitant jet causes fluttering of the mitral valve and, if severe, causes partial closure of the anterior mitral leaflet, leading to functional mitral stenosis and a soft mid-diastolic (Austin Flint) murmur.

In acute severe regurgitation (e.g. perforation of aortic cusp in endocarditis), there may be no time for compensatory left ventricular hypertrophy and dilatation to develop and the features of heart failure may predominate. In this situation, the classical signs of aortic regurgitation may be masked by tachycardia and an abrupt rise in left ventricular end-diastolic pressure; thus, the pulse pressure may be near normal and the diastolic murmur may be short or even absent.

**Investigations**

Regurgitation is detected by Doppler echocardiography (Box 18.111). In severe acute aortic regurgitation, the rapid rise in left ventricular diastolic pressure may cause premature mitral valve closure. Cardiac catheterisation and aortography can help in assessing the severity of regurgitation, and dilatation of the aorta and the presence of coexisting coronary artery disease. MRI is useful in assessing the degree and extent of aortic dilatation.

<table>
<thead>
<tr>
<th>18.111 Investigations in aortic regurgitation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ECG</strong></td>
</tr>
<tr>
<td>• Initially normal, later left ventricular hypertrophy and T-wave inversion</td>
</tr>
<tr>
<td><strong>Chest X-ray</strong></td>
</tr>
<tr>
<td>• Cardiac dilatation, maybe aortic dilatation</td>
</tr>
<tr>
<td>• Features of left heart failure</td>
</tr>
<tr>
<td><strong>Echo</strong></td>
</tr>
<tr>
<td>• Dilated LV</td>
</tr>
<tr>
<td>• Hyperdynamic LV</td>
</tr>
<tr>
<td>• Doppler detects reflux</td>
</tr>
<tr>
<td><strong>Cardiac catheterisation (may not be required)</strong></td>
</tr>
<tr>
<td>• Dilated LV</td>
</tr>
<tr>
<td>• Aortic regurgitation</td>
</tr>
<tr>
<td>• Dilated aortic root</td>
</tr>
</tbody>
</table>

**Management**

Treatment may be required for underlying conditions, such as endocarditis or syphilis. Aortic valve replacement is indicated if aortic regurgitation causes symptoms, and this may need to be combined with aortic root replacement and coronary bypass surgery. Those with chronic aortic regurgitation can remain asymptomatic for many years because compensatory ventricular dilatation and hypertrophy occur, but should be advised to report the development of any symptoms of breathlessness or angina. Asymptomatic patients should also be followed up annually with echocardiography for evidence of increasing ventricular size. If this occurs or if the end-systolic dimension increases to 55 mm or more, then aortic valve replacement should be undertaken. Systolic BP should be controlled with vasodilating drugs, such as nifedipine or ACE inhibitors. There is conflicting evidence regarding the need for aortic valve replacement in asymptomatic patients with severe aortic regurgitation. When aortic root dilatation is the cause of aortic regurgitation (e.g. Marfan’s syndrome), aortic root replacement is usually necessary.

**Tricuspid valve disease**

**Tricuspid stenosis**

**Aetiology**

Tricuspid stenosis is usually rheumatic in origin and is rare in developed countries. Tricuspid disease occurs in fewer than 5% of patients with rheumatic heart disease and nearly always in association with mitral and aortic valve disease. Tricuspid stenosis and regurgitation are features of the carcinoid syndrome (p. 784).

**Clinical features and investigations**

Although the symptoms of mitral and aortic valve disease predominate, tricuspid stenosis may cause symptoms of right heart failure, including hepatic discomfort and peripheral oedema.

The main clinical feature is a raised JVP with a prominent a wave, and a slow y descent due to the loss of normal rapid right ventricular filling (p. 527). There is also a mid-diastolic murmur, best heard at the lower left or right sternal edge. This is generally higher-pitched than the murmur of mitral stenosis and is increased by inspiration. Right heart failure causes hepatomegaly with pre-systolic pulsation (large a wave), ascites and peripheral oedema. On Doppler echocardiography, the valve has similar appearances to those of rheumatic mitral stenosis.

**Management**

In patients who require surgery to other valves, either the tricuspid valve is replaced or valvotomy is performed at surgery. Balloon valvuloplasty can be used to treat rare cases of isolated tricuspid stenosis.

**Tricuspid regurgitation**

**Aetiology, clinical features and investigations**

Tricuspid regurgitation is common, and is most frequently ‘functional’ as a result of right ventricular dilatation (Box 18.112).

Symptoms are usually non-specific, with tiredness related to reduced forward flow, and oedema and hepatic enlargement due to venous congestion. The most prominent sign is a ‘giant’ v wave in the jugular venous pulse (a co wave replaces the normal x descent). Other features include a pansystolic murmur at the left

<table>
<thead>
<tr>
<th>18.112 Causes of tricuspid regurgitation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
</tr>
<tr>
<td>• Rheumatic heart disease</td>
</tr>
<tr>
<td>• Endocarditis, particularly in injection drug-users</td>
</tr>
<tr>
<td>• Ebstein’s congenital anomaly (see Box 18.123, p. 635)</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
</tr>
<tr>
<td>• Right ventricular dilatation due to chronic left heart failure ‘(functional tricuspid regurgitation’)</td>
</tr>
<tr>
<td>• Right ventricular infarction</td>
</tr>
<tr>
<td>• Pulmonary hypertension (e.g. cor pulmonale)</td>
</tr>
</tbody>
</table>
sternal edge and a pulsatile liver. Echocardiography may reveal dilatation of the RV. If the valve has been affected by rheumatic disease, the leaflets will appear thickened and, in endocarditis, vegetations may be seen. Ebstein’s anomaly (see Box 18.123, p. 635) is a congenital abnormality in which the tricuspid valve is displaced towards the right ventricular apex, with consequent enlargement of the RA. It is commonly associated with tricuspid regurgitation.

Management

Tricuspid regurgitation due to right ventricular dilatation often improves when the cause of right ventricular overload is corrected, with diuretic and vasodilator treatment of congestive cardiac failure. Patients with a normal pulmonary artery pressure tolerate isolated tricuspid reflux well, and valves damaged by endocarditis do not usually need to be replaced. Patients undergoing mitral valve replacement, who have tricuspid regurgitation due to marked dilatation of the tricuspid annulus, benefit from valve repair with an annuloplasty ring to bring the leaflets closer together. Those with rheumatic damage may require tricuspid valve replacement.

Pulmonary valve disease

Pulmonary stenosis

This can occur in the carcinoid syndrome but is usually congenital, in which case it may be isolated or associated with other abnormalities, such as Fallot’s tetralogy (p. 634).

The principal finding on examination is an ejection systolic murmur, loudest at the left upper sternum and radiating towards the left shoulder. There may be a thrill, best felt when the patient leans forward and breathes out. The murmur is often preceded by an ejection sound (click). Delay in right ventricular ejection may cause wide splitting of the second heart sound. Severe pulmonary stenosis is characterised by a loud harsh murmur, an inaudible pulmonary closure sound ($P_2$), an increased right ventricular heave, prominent $a$ waves in the jugular pulse, ECG evidence of right ventricular hypertrophy, and post-stenotic dilatation in the pulmonary artery on the chest X-ray. Doppler echocardiography is the definitive investigation.

Mild to moderate isolated pulmonary stenosis is relatively common and does not usually progress or require treatment. Severe pulmonary stenosis (resting gradient > 50 mmHg with a normal cardiac output) is treated by percutaneous pulmonary balloon valvuloplasty or, if this is not available, by surgical valvotomy. Long-term results are very good. Post-operative pulmonary regurgitation is common but benign.

Pulmonary regurgitation

This is rare in isolation and is usually associated with pulmonary artery dilatation due to pulmonary hypertension. It may complicate mitral stenosis, producing an early diastolic decrescendo murmur at the left sternal edge that is difficult to distinguish from aortic regurgitation (Graham Steell murmur). The pulmonary hypertension may be secondary to other disease of the left side of the heart, primary pulmonary vascular disease or Eisenmenger’s syndrome (p. 631). Trivial pulmonary regurgitation is a frequent finding in normal individuals and has no clinical significance.

Infected endocarditis

This is caused by microbial infection of a heart valve (native or prosthetic), the lining of a cardiac chamber or blood vessel, or a congenital anomaly (e.g. septal defect). The causative organism is usually a bacterium, but may be a ricketsia, chlamydia or fungus.

Pathophysiology

Infected endocarditis typically occurs at sites of pre-existing endocardial damage, but infection with particularly virulent or aggressive organisms (e.g. *Staphylococcus aureus*) can cause endocarditis in a previously normal heart; staphylococcal endocarditis of the tricuspid valve is a common complication of intravenous drug misuse. Many acquired and congenital cardiac lesions are vulnerable to endocarditis, particularly areas of endocardial damage caused by a high-pressure jet of blood, such as ventricular septal defect, mitral regurgitation and aortic regurgitation, many of which are haemodynamically insignificant. In contrast, the risk of endocarditis at the site of haemodynamically important low-pressure lesions, such as a large atrial septal defect, is minimal.

Infection tends to occur at sites of endothelial damage because they attract deposits of platelets and fibrin that are vulnerable to colonisation by blood-borne organisms. The atherosclerotic valve tissue and presence of fibrin and platelet aggregates help to protect proliferating organisms from host defence mechanisms. When the infection is established, vegetations composed of organisms, fibrin and platelets grow and may become large enough to cause obstruction or embolism. Adjacent tissues are destroyed and abscesses may form. Valve regurgitation may develop or increase if the affected valve is damaged by tissue distortion, cusp perforation or disruption of chordae. Extracardiac manifestations, such as vasculitis and skin lesions, are due to emboli or immune complex deposition. Mycotic aneurysms may develop in arteries at the site of infected emboli. At autopsy, infarction of the spleen and kidneys and, sometimes, an immune glomerulonephritis are found.

Microbiology

Over three-quarters of cases are caused by streptococci or staphylococcii. The *viridans* group of streptococci (*Streptococcus mitis, Strep. sanguis*) are commensals in the upper respiratory tract that may enter the blood stream on chewing or teeth-brushing, or at the time of dental treatment, and are common causes of subacute endocarditis (Box 18.113). Other organisms, including *Enterococcus faecalis, E. faecium* and *Strep. bovis*, may enter the blood from the bowel or urinary tract. *Strep. milleri* and *Strep. bovis* endocarditis is associated with large-bowel neoplasms.

*Staph. aureus* has now overtaken streptococci as the most common cause of acute endocarditis. It originates from skin infections, abscesses or vascular access sites (e.g. intravenous and central lines), or from intravenous drug use. It is highly virulent and invasive, usually producing florid vegetations, rapid valve destruction and abscess formation. Other causes of acute endocarditis include *Strep. pneumoniae* and *Strep. pyogenes*. 
Post-operative endocarditis after cardiac surgery may affect native or prosthetic heart valves or other prosthetic materials. The most common organism is a coagulase-negative staphylococcus (Staph. epidermidis), a normal skin commensal. There is frequently a history of wound infection with the same organism. *Staph. epidermidis* occasionally causes endocarditis in patients who have not had cardiac surgery, and its presence in blood cultures may be erroneously dismissed as contamination. Another coagulase-negative staphylococcus, *Staph. lugdenensis*, causes a rapidly destructive acute endocarditis that is associated with previously normal valves and multiple emboli. Unless accurately identified, it may also be overlooked as a contaminant.

In Q fever endocarditis due to *Coxiella burnetii*, the patient often has a history of contact with farm animals. The aortic valve is usually affected and there may also be hepatitis, pneumonia and purpura. Life-long antibiotic therapy may be required.

Gram-negative bacteria of the so-called HACEK group (Haemophilus spp., Actinobacillus actinomycetemcomitans, Cardiobacterium hominis, Eikenella spp. and *Kingella kingae*) are slow-growing, fastidious organisms that are only revealed after prolonged culture and may also be overlooked as a contaminant. *Kingella kingae* is associated with a history of contact with farm animals and often affects the aortic valve.

Yeasts and fungi (*Candida, Aspergillus*) may attack previously normal or prosthetic valves, particularly in immunocompromised patients or those with indwelling intravenous lines. Abscesses and emboli are common, therapy is difficult (surgery is often required) and mortality is high. Concomitant bacterial infection may be present.

**Clinical features**

Endocarditis can take either an acute or a more insidious ‘subacute’ form. However, there is considerable overlap because the clinical pattern is influenced not only by the organism, but also by the site of infection, prior antibiotic therapy and the presence of a valve or shunt prosthesis. The subacute form may abruptly develop acute life-threatening complications, such as valve disruption or emboli.

**Subacute endocarditis**

This should be suspected when a patient with congenital or valvular heart disease develops a persistent fever, complains of unusual tiredness, night sweats or weight loss, or develops new signs of valve dysfunction or heart failure. Less often, it presents as an embolic stroke or peripheral arterial embolism. Other features (Fig. 18.93) include purpura and petechial haemorrhages in the skin and mucous membranes, and splinter haemorrhages under the fingernails or toe nails. Osler’s nodes are painful tender swellings at the fingertips that are probably the product of vasculitis; they are rare. Digital clubbing is a late sign. The spleen is frequently palpable; in *Coxiella* infections, the spleen and the liver may be considerably enlarged. Microscopic haematuria is common. The finding of any of these features in a patient with persistent fever or malaise is an indication for re-examination to detect hitherto unrecognised heart disease.

**Acute endocarditis**

This presents as a severe febrile illness with prominent and changing heart murmurs and petechiae. Clinical stigmata of chronic endocarditis are usually absent. Embolic events are common, and cardiac or renal failure may develop rapidly. Abscesses may be detected on echocardiography. Partially treated acute endocarditis behaves like subacute endocarditis.

**Post-operative endocarditis**

This may present as an unexplained fever in a patient who has had heart valve surgery. The infection usually involves the valve ring and may resemble subacute or
Diseases of the heart valves

Acute endocarditis, depending on the virulence of the organism. Morbidity and mortality are high and redo surgery is often required. The range of organisms is similar to that seen in native valve disease, but when endocarditis occurs during the first few weeks after surgery, it is usually due to infection with a coagulase-negative staphylococcus that was introduced during the peri-operative period. A clinical diagnosis of endocarditis can be made on the presence of two major, one major and three minor, or five minor criteria (Box 18.115).

**Investigations**

Blood culture is the crucial investigation because it may identify the infection and guide antibiotic therapy. Three to six sets of blood cultures should be taken prior to commencing therapy and should not wait for episodes of pyrexia. The first two specimens will detect bacteraemia in 90% of culture-positive cases. Aseptic technique is essential and the risk of contaminants should be minimised by sampling from different venepuncture sites. An in-dwelling line should not be used to take cultures. Aerobic and anaerobic cultures are required.

Echocardiography is key for detecting and following the progress of vegetations, for assessing valve damage

**18.115 Diagnosis of infective endocarditis (modified Duke criteria)**

**Major criteria**
- Positive blood culture
  - Typical organism from two cultures
  - Persistent positive blood cultures taken > 12 hrs apart
  - Three or more positive cultures taken over > 1 hr

**Endocardial involvement**
- Positive echocardiographic findings of vegetations
- New valvular regurgitation

**Minor criteria**
- Predisposing valvular or cardiac abnormality
- Intravenous drug misuse
- Pyrexia ≥ 38°C
- Embolic phenomenon
- Vasculitic phenomenon
- Blood cultures suggestive: organism grown but not achieving major criteria
- Suggestive echocardiographic findings

**Definite endocarditis** = two major, or one major and three minor, or five minor
**Possible endocarditis** = one major and one minor, or three minor
and for detecting abscess formation. Vegetations as small as 2–4 mm can be detected by transthoracic echocardiography, and even smaller ones (1–1.5 mm) can be visualised by transoesophageal echocardiography (TOE), which is particularly valuable for identifying abscess formation and investigating patients with prosthetic heart valves. Vegetations may be difficult to distinguish in the presence of an abnormal valve; the sensitivity of transthoracic echo is approximately 65% but that of TOE is more than 90%. Failure to detect vegetations does not exclude the diagnosis.

Elevation of the ESR, a normocytic normochromic anaemia, and leucocytosis are common but not invariable. Measurement of serum CRP is more reliable than the ESR in monitoring progress. Proteinuria may occur and microscopic haematuria is usually present.

The ECG may show the development of AV block (due to aortic root abscess formation) and occasionally infarction due to emboli. The chest X-ray may show evidence of cardiac failure and cardiomegaly.

Management

The case fatality of bacterial endocarditis is approximately 20% and even higher in those with prosthetic valve endocarditis and those infected with antibiotic-resistant organisms. A multidisciplinary approach, with cooperation between the physician, surgeon and microbiologist, increases the chance of a successful outcome. Any source of infection should be removed as soon as possible; for example, a tooth with an apical abscess formation and investigating patients with prosthetic heart valves. Vegetations may be difficult to distinguish in the presence of an abnormal valve; the sensitivity of transthoracic echo is approximately 65% but that of TOE is more than 90%. Failure to detect vegetations does not exclude the diagnosis.

Elevation of the ESR, a normocytic normochromic anaemia, and leucocytosis are common but not invariable. Measurement of serum CRP is more reliable than the ESR in monitoring progress. Proteinuria may occur and microscopic haematuria is usually present.

The ECG may show the development of AV block (due to aortic root abscess formation) and occasionally infarction due to emboli. The chest X-ray may show evidence of cardiac failure and cardiomegaly.

Management

The case fatality of bacterial endocarditis is approximately 20% and even higher in those with prosthetic valve endocarditis and those infected with antibiotic-resistant organisms. A multidisciplinary approach, with cooperation between the physician, surgeon and microbiologist, increases the chance of a successful outcome. Any source of infection should be removed as soon as possible; for example, a tooth with an apical abscess should be extracted.

Empirical treatment depends on the mode of presentation, the suspected organism, and whether the patient has a prosthetic valve or penicillin allergy (Box 18.116). If the presentation is acute, flucloxacillin and gentamicin are preferred.

In those with penicillin allergy, a prosthetic valve or suspected meticillin-resistant *Staph. aureus* (MRSA) infection, triple therapy with vancomycin, gentamicin and oral rifampicin should be considered. Following identification of the causal organism, determination of the minimum inhibitory concentration (MIC) for the organism is essential to guide antibiotic therapy.

A 2-week treatment regimen may be sufficient for fully sensitive strains of *Strep. viridans* and *Strep. bovis*, provided specific conditions are met (Box 18.117). For the empirical treatment of bacterial endocarditis, penicillin plus gentamicin is the regimen of choice for most patients; when staphylococcal infection is suspected, however, vancomycin plus gentamicin is recommended.

Cardiac surgery (débridement of infected material and valve replacement) is advisable in a substantial proportion of patients, particularly those with *Staph. aureus* and fungal infections (Box 18.118). Antimicrobial therapy must be started before surgery.

Prevention

Until recently, antibiotic prophylaxis was routinely given to people at risk of infective endocarditis undergoing interventional procedures. However, as this has not been proven to be effective and the link between

### 18.117 Conditions for the short-course treatment of *Strep. viridans/bovis* endocarditis

- Native valve infection
- MIC ≤ 0.1 mg/L
- No adverse prognostic factors (e.g. heart failure, aortic regurgitation, conduction defect)
- No evidence of thromboembolic disease
- No vegetations > 5 mm diameter
- Clinical response within 7 days

### 18.116 Antimicrobial treatment of common causative organisms in infective endocarditis

<table>
<thead>
<tr>
<th>Organism</th>
<th>Antimicrobial</th>
<th>Dose</th>
<th>Native valve</th>
<th>Prosthetic valve</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Viridans streptococci</em> and <em>Strep. bovis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIC ≤ 0.1 mg/L</td>
<td>Benzyl penicillin IV and gentamicin IV</td>
<td>1.2 g 6 times daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>MIC &gt; 0.1 to &lt; 0.5 mg/L</td>
<td>Benzyl penicillin IV and gentamicin IV</td>
<td>1 mg/kg 2–3 times daily</td>
<td>2 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 wks</td>
</tr>
<tr>
<td>MIC ≥ 0.5 mg/L</td>
<td>Benzyl penicillin IV and gentamicin IV</td>
<td>1.2 g 6 times daily</td>
<td>4 wks</td>
<td>6 wks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 mg/kg 2–3 times daily</td>
<td>2 wks</td>
<td>4–6 wks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2 g 6 times daily</td>
<td>4 wks</td>
<td>6 wks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 mg/kg 2–3 times daily</td>
<td>4 wks</td>
<td>4–6 wks</td>
</tr>
<tr>
<td>Enterococci</td>
<td>Ampicillin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 g 6 times daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>Ampicillin-resistant Staph.</td>
<td>Vancomycin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1 g twice daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>Penicillin-sensitive Staph.</td>
<td>Benzyl penicillin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1 g twice daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>Penicillin-resistant</td>
<td>Fluoroacillin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1 g twice daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>Metillin-sensitive</td>
<td>Penicillin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1 g twice daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
<tr>
<td>Metillin-resistant</td>
<td>Penicillin IV and gentamicin IV&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1 g twice daily</td>
<td>4 wks&lt;sup&gt;1&lt;/sup&gt;</td>
<td>6 wks</td>
</tr>
</tbody>
</table>

<sup>1</sup>When conditions in Box 18.117 are met, 2 wks of benzyl penicillin. <sup>2</sup>High-level gentamicin resistance, consider streptomycin. <sup>3</sup>Consider additional rifampicin 300–600 mg twice daily orally for 2 wks. (MIC = minimum inhibitory concentration)
episodes of infective endocarditis and interventional procedures has not been demonstrated, antibiotic prophylaxis is no longer offered routinely for defined interventional procedures.

**Valve replacement surgery**

Diseased heart valves can be replaced with mechanical or biological prostheses. The three most commonly used types of mechanical prosthesis are the ball and cage, tilting single disc and tilting bi-leaflet valves. All generate prosthetic sounds or clicks on auscultation. Pig or allograft valves mounted on a supporting stent are the most commonly used biological valves. They generate normal heart sounds. All prosthetic valves used in the aortic position produce a systolic flow murmur.

All mechanical valves require long-term anticoagulation because they can cause systemic thromboembolism or may develop valve thrombosis or obstruction (Box 18.119); the prosthetic clicks may become inaudible if the valve malfunctions. Biological valves have the advantage of not requiring anticoagulants to maintain proper function; however, many patients undergoing valve replacement surgery, especially mitral valve replacement, will have atrial fibrillation that requires anticoagulation anyway. Biological valves are less durable than mechanical valves and may degenerate 7 or more years after implantation, particularly when used in the mitral position. They are more durable in the aortic position and in older patients, so are particularly appropriate for patients over 65 undergoing aortic valve replacement.

Symptoms or signs of unexplained heart failure in a patient with a prosthetic heart valve may be due to valve dysfunction, and urgent assessment is required. Biological valve dysfunction is usually associated with the development of a regurgitant murmur.

**The fetal circulation**

Understanding the fetal circulation helps clarify how some forms of congenital heart disease occur. The fetus has only a small flow of blood through the lungs, as it does not breathe in utero. The fetal circulation allows oxygenated blood from the placenta to pass directly to the left side of the heart through the foramen ovale without having to flow through the lungs, and also from the pulmonary artery into the aorta via the ductus arteriosus (Fig. 18.94).

Congenital defects may arise if the changes from fetal circulation to the extrauterine circulation are incomplete. Atrial septal defects occur at the site of the foramen ovale. A patent ductus arteriosus may remain if it fails to close after birth. Failure of the aorta to develop at the point of the aortic isthmus and where the ductus arteriosus attaches can lead to coarctation of the aorta.

In fetal development, the heart develops as a single tube which folds back on itself and then divides into two separate circulations. Failure of septation can lead to some forms of atrial and ventricular septal defect. Failure of alignment of the great vessels with the ventricles contributes to transposition of the great arteries, tetralogy of Fallot and truncus arteriosus.

**Aetiology and incidence**

The incidence of haemodynamically significant congenital cardiac abnormalities is about 0.8% of live births (Box 18.121). Maternal infection or exposure to drugs or toxins may cause congenital heart disease. Maternal rubella infection is associated with persistent ductus arteriosus, pulmonary valvular and/or artery stenosis, and atrial septal defect. Maternal alcohol misuse is associated with septal defects, and maternal lupus erythematosus with congenital complete heart block. Genetic or chromosomal abnormalities, such as Down’s syndrome, may cause septal defects, and gene defects have also
been identified as leading to specific abnormalities, such as Marfan’s syndrome (p. 603) and DiGeorge’s (deletion in chromosome 22q) syndrome.

**Clinical features**

Symptoms may be absent, or the child may be breathless or fail to attain normal growth and development. Some defects are not compatible with extraterine life, or are so only for a short time. Clinical signs vary with the anatomical lesion. Murmurs, thrills or signs of cardiomegaly may be present. In coarctation of the aorta, radiofemoral delay may be noted (Fig. 18.95) and some female patients have the features of Turner’s syndrome (p. 765). Features of other congenital conditions, such as Marfan’s syndrome or Down’s syndrome, may also be apparent. Cerebrovascular accidents and cerebral abscesses may complicate severe cyanotic congenital disease.

Early diagnosis is important because many types of congenital heart disease are amenable to surgery, but

---

**Fig. 18.94** Changes in the circulation at birth. **A** In the fetus, oxygenated blood comes through the umbilical vein where it enters the inferior vena cava via the ductus venosus (red). The oxygenated blood streams from the RA through the open foramen ovale to the LA and via the LV into the aorta. Venous blood from the superior vena cava (blue) crosses under the main blood stream into the RA and then, partly mixed with oxygenated blood (purple), into the RV and pulmonary artery. The pulmonary vasculature has a high resistance and so little blood passes to the lungs; most blood passes through the ductus arteriosus to the descending aorta. The aortic isthmus is a constriction in the aorta that lies in the aortic arch before the junction with the ductus arteriosus and limits the flow of oxygen-rich blood to the descending aorta. This configuration means that less oxygen-rich blood is supplied to organ systems that take up their function mainly after birth, e.g. the kidneys and intestinal tract. **B** At birth, the lungs expand with air and pulmonary vascular resistance falls, so that blood now flows to the lungs and back to the LA. The left atrial pressure rises above right atrial pressure and the flap valve of the foramen ovale closes. The umbilical arteries and the ductus venosus close. In the next few days, the ductus arteriosus closes under the influence of hormonal changes (particularly prostaglandins) and the aortic isthmus expands (IVC = inferior vena cava; LA = left atrium; LV = left ventricle; PA = pulmonary artery; PV = pulmonary vein; RA = right atrium; RV = right ventricle; SVC = superior vena cava). Adapted from Drews 1995 – see p. 641.

**18.121 Incidence and relative frequency of congenital cardiac malformations**

<table>
<thead>
<tr>
<th>Lesion</th>
<th>% of all congenital heart defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventricular septal defect</td>
<td>30</td>
</tr>
<tr>
<td>Atrial septal defect</td>
<td>10</td>
</tr>
<tr>
<td>Patent ductus arteriosus</td>
<td>10</td>
</tr>
<tr>
<td>Pulmonary stenosis</td>
<td>7</td>
</tr>
<tr>
<td>Coarctation of aorta</td>
<td>7</td>
</tr>
<tr>
<td>Aortic stenosis</td>
<td>6</td>
</tr>
<tr>
<td>Tetralogy of Fallot</td>
<td>6</td>
</tr>
<tr>
<td>Complete transposition of great arteries</td>
<td>4</td>
</tr>
<tr>
<td>Others</td>
<td>20</td>
</tr>
</tbody>
</table>

**Fig. 18.95** Radiofemoral delay. The difference in pulse pressures is shown.
this opportunity is lost if secondary changes, such as irreversible pulmonary hypertension, occur.

Central cyanosis and digital clubbing

Central cyanosis of cardiac origin occurs when desaturated blood enters the systemic circulation without passing through the lungs (i.e. a right-to-left shunt). In the neonate, the most common cause is transposition of the great arteries, in which the aorta arises from the RV and the pulmonary artery from the LV in association with a ventricular septal defect. In older children, cyanosis is usually the consequence of a ventricular septal defect combined with severe pulmonary stenosis (tetralogy of Fallot) or with pulmonary vascular disease (Eisenmenger’s syndrome). Prolonged cyanosis is associated with finger and toe clubbing (p. 526).

Growth retardation and learning difficulties

These may occur with large left-to-right shunts at ventricular or great arterial level, and also with other defects, especially if they form part of a genetic syndrome. Major intellectual impairment is uncommon in children with isolated congenital heart disease; however, minor learning difficulties can occur and may complicate cardiac surgery if cerebral perfusion is compromised.

Syncope

In the presence of increased pulmonary vascular resistance or severe or right ventricular outflow obstruction, exercise may provoke syncope as systemic vascular resistance falls but pulmonary vascular resistance rises, worsening right-to-left shunting and cerebral oxygenation. Syncope can also occur because of associated arrhythmias.

Pulmonary hypertension and Eisenmenger’s syndrome

Persistently raised pulmonary flow (e.g. with left-to-right shunt) causes increased pulmonary resistance followed by pulmonary hypertension. Progressive changes, including obliteration of distal vessels, occur and are irreversible. Central cyanosis appears and digital clubbing develops. The chest X-ray shows enlarged central pulmonary arteries and peripheral ‘pruning’ of the pulmonary vessels. The ECG shows right ventricular hypertrophy. If severe pulmonary hypertension develops, a left-to-right shunt may reverse, resulting in right-to-left shunt and marked cyanosis (Eisenmenger’s syndrome), which may be more apparent in the feet and toes than in the upper part of the body: differential cyanosis. This is more common with large ventricular septal defects or persistent ductus arteriosus than with atrial septal defects. Patients with Eisenmenger’s syndrome are at particular risk from abrupt changes in afterload that exacerbate right-to-left shunting, such as vasodilatation, anaesthesia and pregnancy.

Pregnancy

During pregnancy, there is a 50% increase in plasma volume, a 40% increase in whole blood volume and a similar increase in cardiac output, so problems may arise in women with congenital heart disease (Box 18.122). Many with palliated or untreated disease will tolerate pregnancy well, however. Pregnancy is particularly hazardous in the presence of conditions associated with cyanosis or severe pulmonary hypertension; maternal mortality in patients with Eisenmenger’s syndrome is more than 50%.

Persistent ductus arteriosus

Aetiology

During fetal life, before the lungs begin to function, most of the blood from the pulmonary artery passes through the ductus arteriosus into the aorta (see Fig. 18.94). Normally, the ductus closes soon after birth but sometimes fails to do so. Persistence of the ductus is associated with other abnormalities and is more common in females. Since the pressure in the aorta is higher than that in the pulmonary artery, there will be a continuous arteriovenous shunt, the volume of which depends on the size of the ductus. As much as 50% of the left ventricular output is recirculated through the lungs, with a consequent increase in the work of the heart (Fig. 18.96).

Clinical features

With small shunts there may be no symptoms for years but, when the ductus is large, growth and development may be retarded. Usually, there is no disability in infancy.
but cardiac failure may eventually ensue, dyspnoea being the first symptom. A continuous ‘machinery’ murmur is heard with late systolic accentuation, maximal in the second left intercostal space below the clavicle (see Fig. 18.96). It is frequently accompanied by a thrill. Pulses are increased in volume.

A large left-to-right shunt in infancy may cause a considerable rise in pulmonary artery pressure and sometimes this leads to progressive pulmonary vascular damage. Enlargement of the pulmonary artery may be detected radiologically. The ECG is usually normal.

**Persistent ductus with reversed shunting**

If pulmonary vascular resistance increases, pulmonary artery pressure may rise until it equals or exceeds aortic pressure. The shunt through the defect may then reverse, causing Eisenmenger’s syndrome. The murmur becomes quieter, may be confined to systole or may disappear. The ECG shows evidence of right ventricular hypertrophy.

**Management**

A patent ductus is closed at cardiac catheterisation with an implantable occlusive device. Closure should be undertaken in infancy if the shunt is significant and pulmonary resistance not elevated, but this may be delayed until later childhood in those with smaller shunts, for whom closure remains advisable to reduce the risk of endocarditis.

**Pharmacological treatment in the neonatal period**

When the ductus is structurally intact, a prostaglandin synthetase inhibitor (indomethacin or ibuprofen) may be used in the first week of life to induce closure. However, in the presence of a congenital defect with impaired lung perfusion (e.g. severe pulmonary stenosis and left-to-right shunt through the ductus), it may be advisable to improve oxygenation by keeping the ductus open with prostaglandin treatment. Unfortunately, these treatments do not work if the ductus is intrinsically abnormal.

**Coarctation of the aorta**

**Aetiology**

Narrowing of the aorta occurs in the region where the ductus arteriosus joins the aorta, i.e. at the isthmus just below the origin of the left subclavian artery (see Fig. 18.94). The condition is twice as common in males and occurs in 1 in 4000 children. It is associated with other abnormalities, most frequently bicuspid aortic valve and ‘berry’ aneurysms of the cerebral circulation (p. 1246). Acquired coarctation of the aorta is rare but may follow trauma or occur as a complication of a progressive arteritis (Takayasu’s disease, p. 1116).

**Clinical features and investigations**

Aortic coarctation is an important cause of cardiac failure in the newborn but symptoms are often absent when it is detected in older children or adults. Headaches may occur from hypertension proximal to the coarctation, and occasionally weakness or cramps in the legs may result from decreased circulation in the lower part of the body. The BP is raised in the upper body but normal or low in the legs. The femoral pulses are weak and delayed in comparison with the radial pulse (see Fig. 18.97). A systolic murmur is usually heard posteriorly, over the coarctation. There may also be an ejection click and systolic murmur in the aortic area due to a bicuspid aortic valve. As a result of the aortic narrowing, collaterals form; they mainly involve the periscapular, internal mammary and intercostal arteries, and may result in localised bruits.

The chest X-ray in early childhood is often normal but later may show changes in the contour of the aorta (indentation of the descending aorta, ‘3 sign’) and notching of the under-surfaces of the ribs from collaterals. MRI is the best imaging method. (Fig. 18.97). The ECG may show evidence of left ventricular hypertrophy, which can be confirmed by echocardiography.

**Management**

In untreated cases, death may occur from left ventricular failure, dissection of the aorta or cerebral haemorrhage. Surgical correction is advisable in all but the mildest cases. If this is carried out sufficiently early in childhood, persistent hypertension can be avoided. Patients repaired in late childhood or adult life often remain hypertensive or develop recurrent hypertension later on. Recurrence of stenosis may occur as the child grows and this may be managed by balloon dilatation and sometimes stenting. The latter may be used as the primary treatment. Coexistent bicuspid aortic valve, which occurs in over 50% of cases, may lead to progressive aortic stenosis or regurgitation, and also requires long-term follow-up.

**Atrial septal defect**

**Aetiology**

Atrial septal defect is one of the most common congenital heart defects and occurs twice as frequently in females. Most are ‘ostium secundum’ defects, involving the fossa ovalis that, in utero, was the foramen ovale (see Fig. 18.94). ‘Ostium primum’ defects result from a defect
Congenital heart disease

in the atrioventricular septum and are associated with a ‘cleft mitral valve’ (split anterior leaflet).

Since the normal RV is more compliant than the LV, a large volume of blood shunts through the defect from the LA to the RA, and then to the RV and pulmonary arteries (Fig. 18.98). As a result, there is gradual enlargement of the right side of the heart and of the pulmonary arteries. Pulmonary hypertension and shunt reversal sometimes complicate atrial septal defect, but are less common and tend to occur later in life than with other types of left-to-right shunt.

Clinical features and investigations

Most children are asymptomatic for many years and the condition is often detected at routine clinical examination or following a chest X-ray. Dyspnoea, chest infections, cardiac failure and arrhythmias, especially atrial fibrillation, are other possible manifestations. The characteristic physical signs are the result of the volume overload of the RV:

• wide, fixed splitting of the second heart sound: wide because of delay in right ventricular ejection (increased stroke volume and right bundle branch block) and fixed because the septal defect equalises left and right atrial pressures throughout the respiratory cycle
• a systolic flow murmur over the pulmonary valve.

In children with a large shunt, there may be a diastolic flow murmur over the tricuspid valve. Unlike a mitral flow murmur, this is usually high-pitched.

The chest X-ray typically shows enlargement of the heart and the pulmonary artery, as well as pulmonary plethora. The ECG usually shows incomplete right bundle branch block because right ventricular depolarisation is delayed as a result of ventricular dilatation (with a ‘primum’ defect, there is also left axis deviation). Echocardiography can directly demonstrate the defect and typically shows RV dilatation, RV hypertrophy and pulmonary artery dilatation. The precise size and location of the defect can be shown by transoesophageal echocardiography (Fig. 18.99).

Management

Atrial septal defects in which pulmonary flow is increased 50% above systemic flow (i.e. flow ratio of 1.5:1) are often large enough to be clinically recognisable and should be closed surgically. Closure can also be accomplished at cardiac catheterisation using implantable closure devices (Fig. 18.100). The long-term prognosis thereafter is excellent, unless pulmonary hypertension has developed. Severe pulmonary hypertension and shunt reversal are both contraindications to surgery.

Ventricular septal defect

Aetiology

Congenital ventricular septal defect occurs as a result of incomplete septation of the ventricles. Embryologically, the interventricular septum has a membranous and a muscular portion, and the latter is further divided into inflow, trabecular and outflow portions. Most congenital defects are ‘perimembranous’, i.e. at the junction of the membranous and muscular portions.
Ventricular septal defects are the most common congenital cardiac defect, occurring once in 500 live births. The defect may be isolated or part of complex congenital heart disease. Acquired ventricular septal defect may result from rupture as a complication of acute MI or, rarely, from trauma.

**Clinical features**

Flow from the high-pressure LV to the low-pressure RV during systole produces a pansystolic murmur, usually heard best at the left sternal edge but radiating all over the precordium (Fig. 18.101). A small defect often produces a loud murmur (maladie de Roger) in the absence of other haemodynamic disturbance. Conversely, a large defect produces a softer murmur, particularly if pressure in the RV is elevated. This may be found immediately after birth, while pulmonary vascular resistance remains high, or when the shunt is reversed in Eisenmenger’s syndrome.

Congenital ventricular septal defect may present as cardiac failure in infants, as a murmur with only minor haemodynamic disturbance in older children or adults, or, rarely, as Eisenmenger’s syndrome. In a proportion of infants, the murmur gets quieter or disappears due to spontaneous closure of the defect.

If cardiac failure complicates a large defect, it is usually absent in the immediate postnatal period and only becomes apparent in the first 4–6 weeks of life. In addition to the murmur, there is prominent parasternal pulsation, tachypnoea and indrawing of the lower ribs on inspiration. The chest X-ray shows pulmonary plethora and the ECG shows bilateral ventricular hypertrophy.

**Management and prognosis**

Small ventricular septal defects require no specific treatment. Cardiac failure in infancy is initially treated medically with digoxin and diuretics. Persisting failure is an indication for surgical repair of the defect. Percutaneous closure devices are under development.

Doppler echocardiography helps to predict the small septal defects that are likely to close spontaneously. Eisenmenger’s syndrome is avoided by monitoring for signs of rising pulmonary resistance (serial ECG and echocardiography) and carrying out surgical repair, when appropriate. Surgical closure is contraindicated in fully developed Eisenmenger’s syndrome when heart-lung transplantation may be the only effective treatment.

Except in Eisenmenger’s syndrome, long-term prognosis is very good in congenital ventricular septal defect. Many patients with Eisenmenger’s syndrome die in the second or third decade of life, but a few survive to the fifth decade without transplantation.

**Tetralogy of Fallot**

The RV outflow obstruction is most often subvalvular (infundibular) but may be valvular, supravalvular or a combination of these (Fig. 18.102). The ventricular septal defect is usually large and similar in aperture to the aortic orifice. The combination results in elevated right ventricular pressure and right-to-left shunting of cyanotic blood across the ventricular septal defect.

**Aetiology**

The embryological cause is abnormal development of the bulbar septum that separates the ascending aorta from the pulmonary artery, and which normally aligns and fuses with the outflow part of the interventricular septum. The defect occurs in about 1 in 2000 births and is the most common cause of cyanosis in infancy after the first year of life.
**Clinical features**

Children are usually cyanosed but this may not be the case in the neonate because it is only when right ventricular pressure rises to equal or exceed left ventricular pressure that a large right-to-left shunt develops. The subvalvular component of the RV outflow obstruction is dynamic and may increase suddenly under adrenergic stimulation. The affected child suddenly becomes increasingly cyanosed, often after feeding or a crying attack, and may become apnoeic and unconscious. These attacks are called ‘Fallot’s spells’. In older children, Fallot’s spells are uncommon but cyanosis becomes increasingly apparent, with stunting of growth, digital clubbing and polycythaemia. Some children characteristically obtain relief by squatting after exertion, which increases the afterload of the left heart and reduces the right-to-left shunting. Fallot’s sign. The natural history before the development of surgical correction was variable but most patients died in infancy or childhood.

On examination, the most characteristic feature is the combination of cyanosis with a loud ejection systolic murmur in the pulmonary area (as for pulmonary stenosis). However, cyanosis may be absent in the newborn or in patients with only mild right ventricular outflow obstruction (‘acyanotic tetralogy of Fallot’).

**Investigations and management**

The ECG shows right ventricular hypertrophy and the chest X-ray shows an abnormally small pulmonary artery and a ‘boot-shaped’ heart. Echocardiography is diagnostic and demonstrates that the aorta is not continuous with the anterior ventricular septum.

The definitive management is total correction of the defect by surgical relief of the pulmonary stenosis and closure of the ventricular septal defect. Primary surgical correction may be undertaken prior to the age of 5 years. If the pulmonary arteries are too hypoplastic, then palliation in the form of a Blalock-Taussig shunt may be performed, with an anastomosis created between the pulmonary artery and subclavian artery. This improves pulmonary blood flow and pulmonary artery development, and may facilitate later definitive correction.

The prognosis after total correction is good, especially if the operation is performed in childhood. Follow-up is needed to identify residual shunting, recurrent pulmonary stenosis and arrhythmias. An implantable defibrillator is sometimes recommended in adulthood.

**Other causes of cyanotic congenital heart disease**

There are other causes of cyanotic congenital heart disease (Box 18.123) and echocardiography is usually the definitive diagnostic procedure, supplemented, if necessary, by cardiac catheterisation.

**Grown-up congenital heart disease**

There are increasing numbers of children who have had surgical correction of congenital defects and who may have further problems as adults. The transition period between paediatric and adult care needs to be managed in a carefully planned manner, addressing many diverse aspects of care (Box 18.124). Those who have undergone correction of coarctation of the aorta may develop...
hypertension in adult life. Those with transposition of the great arteries who have had a ‘Mustard’ repair, where blood is redirected at atrial level, leaving the RV to aorta, may develop right ventricular failure in adult life. The RV is unsuited for function at systemic pressures and may begin to dilate and fail when patients are in their twenties or thirties.

Those who have had surgery involving the atria may develop atrial arrhythmias, and those who have ventricular scars may develop ventricular arrhythmias and need consideration for implantation of an implantable cardiac defibrillator. Such patients require careful follow-up from the teenage years throughout adult life, so that problems can be identified early and appropriate medical or surgical treatment instituted. The management of patients with grown-up congenital heart disease (‘GUCH’) is complex and has developed as a cardiological subspecialty.

### Diseases of the Myocardium

Although the myocardium is involved in most types of heart disease, the terms ‘myocarditis’ and ‘cardiomyopathy’ are usually reserved for conditions that primarily affect the heart muscle.

#### Myocarditis

This is an acute inflammatory condition that can have an infectious, toxic or autoimmune aetiology. Myocarditis can complicate many infections in which inflammation may be due directly to infection of the myocardium or the effects of circulating toxins. Viral infections are the most common causes, such as Coxsackie (35 cases per 1000 infections) and influenza A and B (25 cases per 1000 infections) viruses. Myocarditis may occur several weeks after the initial viral symptoms and susceptibility is increased by corticosteroid treatment, immunosuppression, radiation, previous myocardial damage and exercise. Some bacterial and protozoal infections may be complicated by myocarditis; for example, approximately 5% of patients with Lyme disease (*Borrelia burgdorferi*, p. 334) develop myopericarditis, which is often associated with AV block. Toxic aetiologies include drugs, which may directly injure the myocardium (e.g. cocaine, lithium and anti-cancer drugs, such as doxorubicin) or which may cause a hypersensitivity reaction and associated myocarditis (e.g. penicillins and sulphonamides, lead and carbon monoxide). Occasionally, autoimmune conditions, such as systemic lupus erythematosus and rheumatoid arthritis, are associated with myocarditis.

**Clinical features and investigations**

Myocarditis can be classified by four distinct clinical presentations:

- **Fulminant myocarditis** follows a viral prodrome or influenza-like illness, and results in severe heart failure or cardiogenic shock.
- **Acute myocarditis** presents over a longer period with heart failure; it can lead to dilated cardiomyopathy.
- **Chronic active myocarditis** is rare and associated with chronic myocardial inflammation.
- **Chronic persistent myocarditis** is characterised by focal myocardial infiltrates and can cause chest pain and arrhythmia without necessarily causing ventricular dysfunction.

In myocarditis, ECG changes are common but non-specific. Biochemical markers of myocardial injury (e.g. troponin I and T, creatine kinase) may be elevated in the early phases. Echocardiography may reveal left ventricular dysfunction that is sometimes regional (due to focal myocarditis), and cardiac MRI may show diagnostic patterns of myocardial inflammation or infiltration. Endomyocardial biopsy is sometimes used to confirm the diagnosis.

#### Cardiomyopathy

Cardiomyopathies are diseases of the myocardium, and are classified according to their structural and functional presentation ([Fig. 18.103](#)). They can be inherited or have infective, toxic or idiopathic aetiologies.

#### Dilated cardiomyopathy

This is characterised by dilatation and impaired contraction of the LV and often the RV. Left ventricular mass is increased but wall thickness is normal or reduced (see [Fig. 18.103](#)). Dilatation of the valve rings can lead to ‘functional’ mitral and tricuspid incompetence. Histological changes are variable but include myofibrillar loss, interstitial fibrosis and T-cell infiltrates. The differential diagnosis includes ventricular dysfunction due to coronary artery disease, and a diagnosis of dilated cardiomyopathy should only be made when this has been excluded.

The pathogenesis is not clear but dilated cardiomyopathy probably encompasses a heterogenous group of conditions. Alcohol may be an important cause in some patients. At least 25% of cases are inherited as an autosomal dominant trait and a variety of single-gene mutations have been identified. Most of these mutations...
Hypertrophic cardiomyopathy

This is the most common form of cardiomyopathy, with a prevalence of approximately 100 per 100,000. It is characterised by inappropriate and elaborate left ventricular hypertrophy with malalignment of the myocardial fibres and myocardial fibrosis. The hypertrophy may be generalised or confined largely to the interventricular septum (asymmetric septal hypertrophy, see Fig. 18.103) or other regions (e.g. apical hypertrophic cardiomyopathy, a variant which is common in the Far East).

Heart failure may develop because the stiff non-compliant ventricles impede diastolic filling. Septal hypertrophy may also cause dynamic left ventricular outflow tract obstruction (hypertrophic obstructive cardiomyopathy, HOCM) and mitral regurgitation due to abnormal systolic anterior motion of the anterior mitral valve leaflet. Effort-related symptoms (angina and breathlessness), arrhythmia and sudden death are the dominant clinical presentations.

Hypertrophic cardiomyopathy is a genetic disorder, usually with autosomal dominant transmission, a high degree of penetrance and variable expression. In most patients, it is due to a single point mutation in one of the genes that encode sarcomeric contractile proteins. There are three common groups of mutation with different phenotypes. Beta-myosin heavy chain mutations are associated with elaborate ventricular hypertrophy and moderate or severe heart failure may be at risk of sudden arrhythmic death. This risk is substantially reduced by rigorous medical therapy with β-blockers and angiotensin receptor antagonists. Some patients may be considered for implantation of a cardiac defibrillator and/or cardiac resynchronisation therapy (p. 579).
present late in life and are often associated with hypertension and arrhythmia.

Symptoms and signs are similar to those of aortic stenosis, except that, in hypertrophic cardiomyopathy, the character of the arterial pulse is jerky (Box 18.125).

The ECG is abnormal and shows features of left ventricular hypertrophy with a wide variety of often bizarre abnormalities (e.g. pseudo-infarct pattern, deep T-wave inversion). Echocardiography is diagnostic, although the diagnosis may be difficult when another cause of left ventricular hypertrophy is present (e.g. physical training – athletes’ heart, hypertension) but the degree of hypertrophy is greater than expected. Genetic testing may facilitate diagnosis and, in some cases, is helpful in screening relatives of affected individuals.

The natural history is variable but clinical deterioration is often slow. The annual mortality from sudden death is 2–3% among adults and 4–6% in children and adolescents (Box 18.126). Sudden death typically occurs during or just after vigorous physical activity; indeed, hypertrophic cardiomyopathy is the most common cause of sudden death in young athletes. Ventricular arrhythmias may be responsible for many of these.

Beta-blockers, rate-limiting calcium antagonists (e.g. verapamil) and disopyramide can help to relieve symptoms and sometimes prevent syncopal attacks; however, there is no pharmacological treatment that is definitely known to improve prognosis. Arrhythmias are common and often respond to treatment with amiodarone. Outflow tract obstruction can be improved by partial surgical resection (myectomy) or by iatrogenic infarction of the basal septum (septal ablation) using a catheter-delivered alcohol solution. An implantable cardiac defibrillator should be considered in patients with clinical risk factors for sudden death (see Box 18.126). Digoxin and vasodilators may increase outflow tract obstruction and should be avoided.

**Arrhythmogenic right ventricular cardiomyopathy**

In this condition, patches of the right ventricular myocardium are replaced with fibrous and fatty tissue (see Fig. 18.102). It is inherited as an autosomal dominant trait and has a prevalence of approximately 10 per 100,000. The dominant clinical problems are ventricular arrhythmias, sudden death and right-sided cardiac failure. The ECG typically shows a slightly broadened QRS complex and inverted T waves in the right precordial leads. MRI is a useful diagnostic tool and is used, along with the 12-lead ECG and ambulatory ECG monitoring, to screen the first-degree relatives of affected individuals. Patients at high risk of sudden death can be offered an implantable cardiac defibrillator.

**Restrictive cardiomyopathy**

In this rare condition, ventricular filling is impaired because the ventricles are ‘stiff’ (see Fig. 18.102). This leads to high atrial pressures with atrial hypertrophy, dilatation and, later, atrial fibrillation. Amyloidosis is the most common cause in the UK, although other forms of infiltration (e.g. glycogen storage diseases), idiopathic perimyocyte fibrosis and a familial form of restrictive cardiomyopathy do occur. Diagnosis can be very difficult and requires complex Doppler echocardiography, CT or MRI, and endomyocardial biopsy. Treatment is symptomatic but the prognosis is usually poor and transplantation may be indicated.

**Obliterative cardiomyopathy**

This is a rare form of restrictive cardiomyopathy, involving the endocardium of one or both ventricles; it is characterised by thrombosis and fibrosis, with gradual obliteration of the ventricular cavities (e.g. endomyocardial fibroelastosis, see Fig. 18.103). The mitral and tricuspid valves become regurgitant. Heart failure and pulmonary and systemic embolism are prominent features. It can sometimes be associated with eosinophilia (e.g. eosinophilic leukaemia, Churg–Strauss syndrome, p. 1118). In tropical countries, the disease can be responsible for up to 10% of cardiac deaths. Mortality is high: 50% at 2 years. Anticoagulation and antiplatelet therapy are used, and diuretics may help symptoms of heart failure. Surgery (tricuspid and/or mitral valve replacement with decortication of the endocardium) may be helpful in selected cases.

**Specific diseases of heart muscle**

Many forms of specific heart muscle disease produce a clinical picture that is indistinguishable from dilated cardiomyopathy (e.g. connective tissue disorders, sarcoidosis, haemochromatosis, alcoholic heart muscle disease, Box 18.127). In contrast, amyloidosis and eosinophilic heart disease produce symptoms and signs similar to those found in restrictive or obliterative cardiomyopathy, whereas the heart disease associated with Friedreich’s ataxia (p. 1199) can mimic hypertrophic cardiomyopathy.
18.127 Specific diseases of heart muscle

- Viral, e.g. Coxsackie A and B, influenza, HIV
- Bacterial, e.g. diphtheria, *Borrelia burgdorferi*
- Protozoal, e.g. trypanosomiasis

18.128 Aetiology of acute pericarditis

**Common**
- Viral (e.g. Coxsackie B but often not identified)
- Acute MI

**Less common**
- Uraemia
- Malignant disease
- Trauma (e.g. blunt chest injury)

**Rare (in UK)**
- Bacterial infection
- Rheumatic fever
- Tuberculosis

---

**Cardiac tumours**

Primary cardiac tumours are rare (< 0.2% of autopsies) but the heart and mediastinum may be the sites of metastases. Most primary tumours are benign (75%) and, of these, the majority are myxomas. The remainder are fibromas, lipomas, fibroelastomas and haemangiomas.

**Atrial myxoma**

Myxomas most commonly arise in the LA as single or multiple polypoid tumours, attached by a pedicle to the interatrial septum. They are usually gelatinous but may be solid and even calcified, with superimposed thrombus.

On examination, the first heart sound is usually loud, and there may be a murmur of mitral regurgitation with a variable diastolic sound (tumour ‘plop’) due to prolapse of the mass through the mitral valve. The tumour can be detected incidentally on echocardiography, or following investigation of pyrexia, syncope, arrhythmias or emboli. Occasionally, the condition presents with malaise and features suggestive of a connective tissue disorder, including a raised ESR.

Treatment is by surgical excision. If the pedicle is removed, fewer than 5% of tumours recur.

---

**Acute pericarditis**

**Aetiology**

Pericardial inflammation may be due to a number of pathologies (Box 18.128) but sometimes remains unexplained. Pericarditis and myocarditis often coexist, and all forms of pericarditis may produce a pericardial effusion (see below) that, depending on the aetiology, may be fibrinous, serous, haemorrhagic or purulent.

A fibrinous exudate may eventually lead to varying degrees of adhesion formation, whereas serous pericarditis often produces a large effusion of turbid, straw-coloured fluid with a high protein content.

A haemorrhagic effusion is often due to malignant disease, particularly carcinoma of the breast or bronchus, and lymphoma.

Purulent pericarditis is rare and may occur as a complication of septicaemia, by direct spread from an intrathoracic infection, or from a penetrating injury.

**Clinical features**

The characteristic pain of pericarditis is retrosternal, radiates to the shoulders and neck, and is typically aggravated by deep breathing, movement, a change of position, exercise and swallowing. A low-grade fever is common. A pericardial friction rub is a high-pitched superficial scratching or crunching noise, produced by movement of the inflamed pericardium, and is diagnostic of pericarditis; it is usually heard in systole but may also be audible in diastole and frequently has a ‘to-and-fro’ quality.

**Investigations and management**

The ECG shows ST elevation with upward concavity (Fig. 18.104) over the affected area, which may be widespread. PR interval depression is a very specific indicator of acute pericarditis. Later, there may be T-wave inversion, particularly if there is a degree of myocarditis.

The pain is usually relieved by aspirin (600 mg 6 times daily) but a more potent anti-inflammatory agent, such as indometacin (25 mg 3 times daily), may be required. Colchicine or corticosteroids may suppress symptoms but there is no evidence that they accelerate cure.

In viral pericarditis, recovery usually occurs within a few days or weeks but there may be recurrences (chronic relapsing pericarditis). Purulent pericarditis requires...
treatment with antimicrobial therapy, pericardiocentesis and, if necessary, surgical drainage.

**Pericardial effusion**

If a pericardial effusion develops, there is sometimes a sensation of retrosternal oppression. An effusion is difficult to detect clinically. The heart sounds may become quieter, although a friction rub is not always abolished.

The QRS voltages on the ECG are often reduced in the presence of a large effusion. The QRS complexes may alternate in amplitude due to a to-and-fro motion of the heart within the fluid-filled pericardial sac (electrical alternans). The chest X-ray may show an increased size of the cardiac silhouette and, when there is a large effusion, this has a globular appearance. Echocardiography is the definitive investigation and is used to monitor the size of the effusion and its effect on cardiac function (Fig. 18.105).

**Cardiac tamponade**

This term is used to describe acute heart failure due to compression of the heart by a large or rapidly developing effusion, and is described in detail on page 545. Typical physical findings are of a markedly raised JVP, hypotension, pulsus paradoxus (p. 532) and oliguria. Atypical presentations may occur when the effusion is loculated as a result of previous pericarditis or cardiac surgery.

**Pericardial aspiration (pericardiocentesis)**

Aspiration of a pericardial effusion is indicated for diagnostic purposes or for the treatment of cardiac tamponade. A needle is inserted under echocardiographic guidance medial to the cardiac apex or below the xiphoid process, directed upwards towards the left shoulder. The route of choice will depend on the experience of the operator, the shape of the patient and the position of the effusion. A few millilitres of fluid aspirated through the needle may be sufficient for diagnostic purposes but pericardial drainage is needed for symptom relief.

Complications of pericardiocentesis include arrhythmias, damage to a coronary artery, and bleeding with exacerbation of tamponade as a result of injury to the RV. When tamponade is due to cardiac rupture or aortic dissection, pericardial aspiration may precipitate further potentially fatal bleeding and, in these situations, emergency surgery is the treatment of choice. A viscous, loculated or recurrent effusion may also require formal surgical drainage.

**Tuberculous pericarditis**

Tuberculous pericarditis may complicate pulmonary tuberculosis but may also be the first manifestation of the infection. In Africa, a tuberculous pericardial effusion is a common feature of AIDS (p. 405).

The condition typically presents with chronic malaise, weight loss and a low-grade fever. An effusion usually develops and the pericardium may become thick and unyielding, leading to pericardial constriction or tamponade. An associated pleural effusion is often present.

The diagnosis may be confirmed by aspiration of the fluid and direct examination or culture for tubercle bacilli. Treatment requires specific antituberculous chemotherapy (p. 693); in addition, a 3-month course of prednisolone (initial dose 60 mg a day, tapering down rapidly) improves outcome.
Chronic constrictive pericarditis

Constrictive pericarditis is due to progressive thickening, fibrosis and calcification of the pericardium. In effect, the heart is encased in a solid shell and cannot fill properly. The calcification may extend into the myocardium, so there may also be impaired myocardial contraction. The condition often follows an attack of tuberculous pericarditis but can also complicate haemopericardium, viral pericarditis, rheumatoid arthritis and purulent pericarditis. It is often impossible to identify the original insult.

Clinical features and management

The symptoms and signs of systemic venous congestion are the hallmarks of constrictive pericarditis. Atrial fibrillation is common and there is often dramatic ascites and hepatomegaly (Box 18.129). Breathlessness is not a prominent symptom because the lungs are seldom congested.

The condition is sometimes overlooked but should be suspected in any patient with unexplained right heart failure and a small heart. A chest X-ray, which may show pericardial calcification (Fig. 18.106), and echocardiography often help to establish the diagnosis. CT scanning is useful for imaging the pericardial calcification.

Constrictive pericarditis is often difficult to distinguish from restrictive cardiomyopathy and the final diagnosis may depend on complex echo–Doppler studies and cardiac catheterisation. Surgical resection of the diseased pericardium can lead to a dramatic improvement but carries a high morbidity with disappointing results in up to 50% of patients.

Further information and acknowledgements

Websites

- www.acc.org American College of Cardiology (ACC): free access to guidelines for the evaluation and management of many cardiac conditions.
- www.americanheart.org American Heart Association (AHA): free access to all the ACC/AHA/ESC guidelines, AHA scientific statements and fact sheets for patients.
- www.escardio.org European Society of Cardiology (ESC): free access to guidelines for the diagnosis and management of many cardiac conditions, and to educational modules.

Figure acknowledgements

Page 526 insets (Splinter haemorrhage, jugular venous pulse, malar flush, tendon xanthomas), Fig. 18.91, 18.92 inset (Doppler echo), 18.93 insets (Petechial rash, nail-fold infarct)

Figs 18.19, 18.70 Scottish Intercollegiate Guidelines Network (SIGN) 93; Feb 2007; pp. 42 (annex 1) and 47 (annex 4).
Figs 18.34, 18.35 Resuscitation Council (UK) guidelines.
Fig. 18.62 Joint British Societies Cardiovascular Risk Prediction Chart, reproduced with permission from the University of Manchester.
Figs 18.84, 18.85 From NICE Clinical Guideline 127, Hypertension; August 2011.
Fig. 18.86 inset (Erythema marginatum) Savin JA, Hunter JAA, Hepburn NC. Skin signs in clinical medicine. London: Mosby–Wolfe; 1997; copyright Elsevier.
Fig. 18.94 Adapted from Drews U. Colour atlas of embryology. Stuttgart: Georg Thieme; 1995 (Fig. 6.9, p. 299).