FOURTH EDITION

EVIDENCE-BASED PRACTICE

ACROSS THE HEALTH PROFESSIONS

TAMMY HOFFMANN
SALLY BENNETT
CHRIS DEL MAR

ELSEVIER

EVIDENCE-BASED PRACTICE

ACROSS THE HEALTH PROFESSIONS

FOURTH
EDITION

EVIDENCE-BASED PRACTICE

ACROSS THE HEALTH PROFESSIONS

TAMMY **HOFFMANN**
SALLY **BENNETT**
CHRIS **DEL MAR**

ELSEVIER

ELSEVIER

Elsevier Australia. ACN 001 002 357
(a division of Reed International Books Australia Pty Ltd)
Tower 1, 475 Victoria Avenue, Chatswood, NSW 2067

ISBN: 978-0-7295-4443-6

Notice

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

National Library of Australia Cataloguing-in-Publication Data

NATIONAL LIBRARY OF AUSTRALIA

A catalogue record for this
book is available from the
National Library of Australia

Senior Content Strategist: Natalie Hunt and Dorothy Chiu
Content Project Manager: Kritika Kaushik
Copy edited by Robyn Flemming
Proofread by Tim Learner
Cover design by Gopalakrishnan Venkatraman
Index by SPi Global

Typeset by GW Tech

Printed in TBA

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

Dedication

Tammy and Sally dedicate this edition to Professor Chris Del Mar, who passed away shortly before it was published. Chris was renowned locally, nationally and internationally for his superior skills in, and commitment to, teaching evidence-based practice. He was a passionate advocate for evidence, clinical research, patient-centredness, and questioning assumptions in health care. Chris leaves an enormous legacy in his various fields of research. One indelible legacy is the immeasurable contribution that he made towards training and inspiring thousands of students and clinicians to provide evidence-based care. This book has influenced our lives in ways that we never could have imagined, and we are deeply privileged to have been able to learn from and partner with him.

From Tammy: to the most wonderful husband I could have ever wished for. Thank you for every moment. I will be forever grateful to this book for bringing us together. You are so very dearly missed.

FOREWORD

The COVID-19 pandemic highlighted the critical importance of research evidence to inform clinical practice and policy worldwide. Never before have we seen such a pressing demand for evidence from the public, clinicians and policy makers. Unfortunately, alongside the COVID-19 pandemic, we also experienced a massive, global misinformation pandemic. The impact of misinformation is substantial—eroding public trust and causing deaths and increased morbidity worldwide. This impact was particularly harsh for those with compounding vulnerabilities. In turn, the misinformation pandemic 'shone' light on the need for all decision makers to seek and appraise evidence. But how is this done?

This fourth edition of *Evidence-Based Practice Across the Health Professions*, edited by Professors Tammy Hoffmann, Sally Bennett and Chris Del Mar, is well-timed to meet this need! This book provides the foundations of evidence-based practice from asking questions, through to seeking, appraising and using evidence in decision making. Particularly helpful are the tips on embedding evidence into routine clinical care, which targets individual clinicians and organisations. At the individual clinician level there is much needed discussion of shared decision making and the book provides practical approaches to incorporating this into clinical care. Some of the commonly reported barriers to evidence-based practice are identified and strategies for overcoming these provided.

Optimising research evidence use in decision making must happen at all levels within the health care system and include the public, clinicians, patients, managers and policy makers. Without this focus, research will be wasted and its impact on patients and the health care (and public health) system will not be realised. We have an ethical and moral imperative to avoid research waste and to use high-quality evidence in health decision making. The latest edition of this book provides the way forward.

Finally, this work is a wonderful legacy of the amazing Professor Del Mar. His commitment to evidence-based practice in his own practice and teaching, while also advancing its methods, served as an exemplar for all of us. This book ensures that his legacy and impact on patients and clinicians continues into the future.

Sharon E Straus, CM, MD, MSc, FRCPC, CAHS, FRSC
Professor, Department of Medicine
University of Toronto

AUTHORS

Tammy Hoffmann OAM, BOccThy (Hons 1), PhD, FOTARA, FAHMS

Professor of Clinical Epidemiology, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Tammy has been teaching and researching about evidence-based practice and shared decision making for over 20 years and has an international reputation in her various areas of research. Her research spans many aspects of shared decision making, evidence-based practice, informed health decisions, improving the reporting and uptake of effective interventions, reporting guidelines, knowledge translation, minimising waste in research and the teaching of evidence-based practice.

Sally Bennett BOccThy (Hons), PhD, FOTARA

Professor in Occupational Therapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

Sally has extensive experience in teaching and research about evidence-based practice and knowledge translation. Her research interests are about building capacity for knowledge translation and translating knowledge for care of people living with dementia. She was one of the leaders of the internationally recognised OTseeker database that provided evidence relevant to occupational therapy. She has been actively involved at the professional level both nationally and internationally, including having been associate editor on a number of occupational therapy journals.

Chris Del Mar AM, BSc, MA, MB BChir, MD, FRACGP, FAFPHM, FAHMS

Professor of Public Health and academic General Practitioner, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Chris worked as a full-time general practitioner for many years before becoming Professor of General Practice at the University of Queensland. He was invited to become the Dean of a new Health Sciences and Medicine Faculty and to develop a new medical program at Bond University. He was also Pro-Vice Chancellor (Research) at Bond University. After overseeing the graduation of the first cohort of medical students, he stepped back from those roles so that he could return his focus to research and teaching and was Professor of Public Health until 2022. Chris's international reputation is in the management of acute respiratory infections (for 20 years, he was the coordinating editor of the Cochrane Acute Respiratory Infections Group); general practice research; evidence-based medicine and systematic reviews; and randomised controlled trials, in both clinical medicine and health services research.

CONTRIBUTORS

Bridget Abell BAppSc (HMS-Exercise Science), MSc, PhD
Implementation Scientist and Early Career Health Services Researcher, Australian Centre for Health Services Innovation, Queensland University of Technology (QUT), Brisbane, Australia

Loai Albarqouni MD, MSc, PhD
Assistant Professor, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Mina Bahkit BMedSurg, MA, PhD
Research Fellow, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Lauren Ball BAppSc, Grad Cert (Higher Ed), Grad Dip Health Economics and Health Policy, M Nutrition & Dietetics (Honours), PhD
Professor of Community Health and Wellbeing, The University of Queensland, Brisbane, Australia

John Bennett BMedSc, MBBS, BA (Hons), PhD, FRACGP, FACHI
General Practitioner, UQ Healthcare, The University of Queensland, Brisbane, Australia

Fiona Bogossian RN, RM, DipAppSci (NEd), BAppSci, MPH, PhD, FACM
Professor, Practice Education in Health, Academic Lead USC Clinical School, University of the Sunshine Coast, Sunshine Coast, Australia

Malcolm Boyle ADipBus, ADipHSc (Amb Off), MICA Cert, BInfoTech, MClinEpi, PhD
Associate Professor and Academic Lead in Paramedicine Education, School of Medicine and Dentistry, Griffith University, Gold Coast, Australia

Mary Bushell BPharm (Hons), AACPA, GCTLHE, AFACP, MPS, PhD
Clinical Assistant Professor, Discipline of Pharmacy, School of Health Sciences, Faculty of Health, University of Canberra, Canberra, Australia

Ryan Causby B Podiatry, M Podiatry, PhD
Program Director Podiatry, Allied Health and Human Performance Unit, University of South Australia, Adelaide, Australia

Justin Clark BA
Senior Research Information Specialist, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Jeff Coombes BEd (Hons), BAppSc, MEd, PhD
Professor, School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, Australia

Scott Devenish MaVEdT, BNur, Dip Para Sc, RN, RPara (Aus), FACP, FHEA, PhD
Associate Professor in Paramedicine and Head of Discipline, School of Nursing, Midwifery and Paramedicine, Faculty of Health, Australian Catholic University, Brisbane, Australia

Fiona Dobson BAppSc (Physiotherapy), Postgraduate Diploma (Health Research Methods), PhD
Associate Professor, Department of Physiotherapy, Melbourne School of Health Sciences, The University of Melbourne, Melbourne, Australia

Jenny Doust BEcons, BMBS, Grad Dip Clin Epi, FRACGP, PhD
Clinical Professor Research Fellow, Australian Women and Girls' Health Research (AWaGHR) Centre, School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Australia

Carolyn Ee MBBS, FRACGP, BAppSci (Chinese Med), MMed, GradCert Med Acup, PhD
Senior Research Fellow, NICM Health Research Institute, Western Sydney University, Penrith, Australia

Roma Forbes BHSc (Physiotherapy), MHSc (Hons), Grad Cert (HigherEd), PhD
Senior Lecturer in Physiotherapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

Elizabeth Gibson BOccThy, PhD
Senior Research Fellow, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Paul Glasziou AO, MBBS, FRACGP, MRCGP, FAHMS, PhD
Professor of Evidence-Based Medicine and Director of the Institute for Evidence-Based Healthcare, Bond University, Gold Coast, Australia

Ian Graham FCAHS, FNYAM, FRSC, PhD
Distinguished Professor, Senior Scientist, Centre for Practice-Changing Research, Ottawa Hospital Research Institute, Ottawa, Canada

Romi Haas BPhysio (Hons), MPH, PhD
Research Fellow, Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University; and Monash-Cabrini Department of Musculoskeletal Health and Clinical Epidemiology, Cabrini Health, Melbourne, Australia

Karin Hannes MSc Edu, MSc Med, PhD
Professor in Transdisciplinary Studies, Creative Research Methodology and Meta-Synthesis at Research Group SoMeTHin'K (Social, Methodological and Theoretical Innovation / Kreative), Faculty of Social Sciences, University of Leuven, Leuven, Belgium

Joanna Harnett MHSc, BHSc (Complementary Medicines), Grad Cert Educational Studies (HigherEd), PhD
Senior Lecturer, School of Pharmacy, The University of Sydney, Sydney, Australia

Joy Higgs AM, BSc, MHPEd, PhD, PFHEA
Emeritus Professor in Higher Education, Charles Sturt University, Sydney, Australia

Kylie Hill BSc (Physiotherapy), PhD
Associate Professor, Curtin School of Allied Health, Faculty of Health Sciences, Curtin University, Perth, Australia

Isabelle Jalbert OD, MPH, PhD
Associate Professor, School of Optometry and Vision Science, Faculty of Medicine and Health, The University of New South Wales, Sydney, Australia

Jacqueline Jauncey-Cooke RN, MN, Grad Dip Crit Care, Grad Cert Health Prof Educ, PhD
Senior Lecturer, School of Nursing, Midwifery and Social Work, The University of Queensland, Brisbane, Australia

Sohil Khan MPharm (Clin Pharm), MBA, PhD
Faculty of Pharmacotherapeutics and Evidence Based Practice, School of Pharmacy and Medical Sciences, Griffith University, Gold Coast, Australia

Nerida Klupp BAppScPod (Hons), PhD
Senior Lecturer, School of Health Sciences, Western Sydney University, Penrith, Australia

Karl Landorf Dip App Sc, Grad Cert Clin Instr, Grad Dip Ed, PhD
Professor of Podiatry, Associate Dean, Research and Industry Engagement, School of Allied Health and La Trobe Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, Australia

David Long Adv Dip Paramed Sc (Amb), BEd (Hab), BHlthSc (Pre-Hosp Care), GCertAcadPrac, PhD
Senior Lecturer (Paramedicine), School of Health and Medical Sciences, University of Southern Queensland, Ipswich, Australia

Amary Mey BPharm (Hons), PhD
Lecturer, School of Pharmacy and Medical Sciences, Griffith University, Gold Coast, Australia

Zachary Munn Grad Dip (Health Sciences), B Med Radiation (Nuclear Medicine), PhD
Director of Evidence-based Healthcare Research, JBI, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia

Natalie Munro BAppSc (Speech Pathology) (Hons I), Grad Cert (HigherEd), CPSP, SFHEA, PhD
Associate Professor, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Shannon Munteanu BPod (Hons), PhD
Professor of Podiatry, Discipline of Podiatry, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Australia

Rebecca Packer BSpPath (Hons), GCHEd, PhD
Lecturer in Speech Pathology, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

Denise O'Connor BAppScOT (Hons), PhD
Associate Professor (Research), School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Matthew Page BBSc (Hons), PhD
Senior Research Fellow, Deputy Head of the Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

Toby Pavey BSc, MSc, PhD
Associate Professor in Physical Activity, Sedentary Behaviour and Health, School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Australia

John Pierce BSpPath, Postgraduate Diploma (Health Research Methodology), PhD
Postdoctoral Research Fellow, Centre of Research Excellence in Aphasia Recovery and Rehabilitation, School of Allied Heath, Human Services and Sport, La Trobe University, Melbourne, Australia

Emma Power BAppSc (Speech Path, Hons 1), PhD
Associate Professor, Speech Pathology, Graduate School of Health, University of Technology Sydney, Sydney, Australia

Claire Rickard BN, RN, GradDip (CriticalCare), FAHMS, FACN, PhD
Professor of Infection Prevention and Vascular Access, Metro North Health and School of Nursing, Midwifery and Social Work, The University of Queensland, Brisbane, Australia

Sharon Sanders BSc (Pod), MPH, PhD
Assistant Professor, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia

Katrina Schmid BAppSc (Opt) (Hons), Grad Cert (Ocular Therapeutics), Therapeutically Endorsed Optometrist, Grad Cert (HigherEd), SFHEA, AFHEA (Indigenous), PhD
Associate Professor, School of Optometry and Vision Science, Faculty of Health, Queensland University of Technology, Brisbane, Australia

Michal Schneider BSc, Grad Dip Ed, M Rep Sc, Grad Cert Health Prof Edu, PhD
Professor, Department of Medical Imaging and Radiation Sciences, Monash University, Melbourne, Australia

Ian Scott FRACP, MHA, MEd
Professor and Director of Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, Australia and Faculty of Medicine, The University of Queensland, Brisbane, Australia

Nichola Shelton BA (Hons), MA, GDip, MSLP, CPSP
PhD candidate, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Rachel Thompson BPsySci (Hons), PhD
Senior Lecturer, School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Leigh Tooth BOccThy (Hons), PhD
Associate Professor, Principal Research Fellow and Deputy Director of the Australian Longitudinal Study on Women's Health, School of Public Health, The University of Queensland, Brisbane, Australia

Adrian Traeger MPhty, BSc (Hons I), PhD
Research Fellow, School of Public Health, The University of Sydney, Sydney, Australia

Merrill Turpin BOccThy, Grad Dip Counsel, PhD
Senior Lecturer, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

Adam P Vogel BA, MSc (SpPth), PhD
Professor, Head of Speech Pathology, School of Health Sciences, The University of Melbourne, Melbourne, Australia; and Redenlab Inc., Australia

Cynthia Wensley RN, MHSc, PGDip Health Systems Management, BA Social Sciences (Nursing), PhD
Lecturer, Faculty of Medical and Health Sciences, Nursing, University of Auckland, New Zealand

Shelley Wilkinson BSc (Hons) (Psyc), Grad Dip Nut & Diet, PhD
Associate Professor, School of Human Movement and Nutrition Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, Brisbane, Australia

Kylie Williams BPharm, Grad Dip Hosp Pharm, PhD
Professor and Head, Discipline of Pharmacy, University of Technology Sydney, Sydney, Australia

Caroline Wright BSc (Hons), MSc, DCR Therapy, PGCE, PhD
Associate Professor, Department of Medical Imaging and Radiation Sciences, Monash University, Melbourne, Australia

Joshua R Zadro PhD, BAppSc (Phty) (Hons 1)
Research Fellow, Institute for Musculoskeletal Health, Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

REVIEWERS

Leanne Bisset MPhty (Manipulative), MPhty (Sports), BPhty, PhD
Griffith University, Gold Coast, Queensland, Australia

Melissa Carey BN, MN, MAP (HCR)
University of Southern Queensland, Queensland, Australia

Anne Cusick MA (Psych), MA (Interdisc Stud), Grad Dip Beh Sc, Grad Cert Bus Admin, BAppSc (OT), Dip AICD, PhD
Professor, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia; Adjunct Professor, College of Health & Human Sciences, Charles Darwin University, Casuarina, Northern Territory, Australia; Emeritus Professor, Western Sydney University, Campbelltown, New South Wales, Australia

Thanya Pathirana MPH, MBBS, PhD
Senior Lecturer in Medical Education, Associate Lead in Doctor and Health in the Community theme, MD program, School of Medicine and Dentistry, Griffith University, Sunshine Coast, Queensland, Australia

Cynthia Wensley MHSc, BA, RN, PhD
School of Nursing, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand

PREFACE

Each time we work on a new edition of this book, there are various methodological developments and new resources and literature to incorporate. The field of evidence-based practice continues to mature and, gratifyingly, it is becoming expected and commonplace in more health settings, disciplines and curricula. The COVID-19 pandemic accentuated the importance of being able to rapidly generate, appraise and disseminate quality evidence for decision making (at individual, health system and global policy levels).

An interdisciplinary approach is best in evidence-based practice and health care, and we are delighted that there are now 16 disciplines represented in the book. We are very appreciative of the 56 contributors to this edition, who are national or international experts in their fields and readily prepared the chapters and worked examples for this edition. Thank you for partnering with us to help teach the skills of evidence-based practice and further its uptake.

Since the book's previous edition, the three of us have had numerous interactions with the health system as patients and family for very serious and minor health conditions. In some of these encounters, we have experienced evidence-based *and* patient-centred care. What a difference it has made when this has occurred. We urge all health professionals, and soon-to-be health professionals, who use this book to learn skills in evidence-based practice to not underestimate the impact that *your* interaction can have on a patient and their family. Providing advice and care that is based on the best available evidence, and carefully considering the way in which you discuss the options with your patient and involve them in the decision making, are powerful strategies that are at your disposal. We hope that you can use this book to learn or refine these skills so that your patients receive the best care possible.

Tammy Hoffmann, Sally Bennett, Chris Del Mar

ACKNOWLEDGMENTS

Thank you to the reviewers who provided useful suggestions. We are particularly grateful to our colleagues, especially Dr Libby Gibson, Dr Sharon Sanders and Dr Romi Haas, who willingly stepped in to assist with finalising content for this edition when its timely completion was threatened.

The publisher would like to remember Professor Chris Del Mar who passed away shortly before the publication of this text. Alongside his extensive contribution to the field of evidence-based practice, Chris's invaluable work in all editions of this text is greatly acknowledged with much thanks.

CONTENTS

6

c0006

Evidence about Diagnosis

Sharon Sanders and Jenny Doust

LEARNING OBJECTIVES

After reading this chapter, you should be able to:

- Generate a structured clinical question about diagnosis for a clinical scenario
- Appraise the risk of bias (validity) in studies of diagnostic test accuracy
- Understand how to interpret the results from diagnostic test accuracy studies and calculate additional results (such as positive and negative predictive values and likelihood ratios) where possible
- Describe how evidence from diagnostic test accuracy studies can be used to inform practice
- Understand what diagnostic clinical prediction rules are, how and why they are developed, and how to determine the readiness of published diagnostic clinical prediction rules for clinical practice

p0210 This chapter will begin with describing and explaining how to appraise and interpret research studies that are concerned with determining the accuracy of diagnostic tests. Later sections of the chapter will touch on other types of diagnostic test research, including research that assesses the impact or utility of diagnostic tests and their reliability.

p0215 Let us consider a clinical scenario that will be useful for illustrating the concepts of diagnostic test accuracy that are the focus of this chapter. The scenario relates to testing for the virus (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2) that causes COVID-19. Given the evolving nature of this coronavirus and the rapid, ongoing development of testing technologies, the clinical scenario and test referred to in this chapter should be considered indicative of the situation at the time the chapter was written.

b0060

CLINICAL SCENARIO

p0225 You are a clinician working in a ward of a major city hospital. You have just been advised that you will be required to undergo testing for severe acute respiratory syndrome coronavirus (SARS-CoV-2) infection before each shift. SARS-CoV-2 is the coronavirus that has caused the pandemic of acute respiratory disease, named coronavirus disease 2019—or COVID-19. A rapid antigen test, a test that detects virus particles present in a sample from a swab inserted into your nose or pharynx and provides a result within 15 minutes, will be used. You have heard mixed reports about the ability of these tests to detect SARS-CoV-2 infection. You decide to find out about the accuracy of rapid antigen tests for confirming or ruling out SARS-CoV-2 infection.

p0230 Diagnosis classifies an individual as having, or not having, a particular condition and can provide crucial information for clinical decisions that influence health outcomes for an individual. The diagnostic tests we use might lead to a patient being given a broad diagnostic label or sometimes to patients being classified into various categories and stratifications within a diagnostic label that can be used to assist with decisions about management. However, as explained in earlier chapters of the book, and as can be seen in some of the worked scenarios in Chapter 7, studies of 'diagnostic test accuracy' are not just about identifying the presence or absence of a condition. When we refer to 'diagnosis' in this chapter, we are also referring to assessing aspects of body structure, function or task performance.

p0235 Using the COVID-19 scenario, this chapter will examine and explain how to assess the diagnostic accuracy of rapid antigen tests to detect SARS-CoV-2 infection. We will start by defining the components of a structured clinical question about diagnosis. We will then see how to appraise the evidence to determine its likely risk of bias. Subsequent sections of the chapter will review how to understand the results of a study that tells us about the accuracy of a diagnostic test and how to use the evidence to inform practice.

p0240 While the example used in this chapter focuses on a single diagnostic test (in this case, a pathology test), we can also assess the accuracy of combinations of diagnostic tests. Clinical examination is an example of this. Clinical examination is usually an iterative process of data collection that generally begins with the history of the presenting condition, recording an individual's symptoms and signs and then performing physical examination. Each piece of information collected may be viewed as a diagnostic 'test' with a measurable diagnostic power. Though comprised of many individual 'tests', clinical examination itself may be considered a diagnostic 'test', akin to a laboratory test, that helps health professionals to decide whether a patient has a disease, impairment or disability.

s0010 STUDY DESIGNS THAT CAN BE USED FOR ANSWERING QUESTIONS ABOUT DIAGNOSTIC TEST ACCURACY

p0245 Studies of diagnostic tests generally measure how accurately a test can detect the presence or absence of a disease by comparing the test with a reference standard. The reference standard is considered the best available method for finding out whether an individual has the condition and may sometimes be referred to as the 'gold standard'. The reference standard may be a single test or a combination of 'tests'. For example, in studies of the accuracy of ultrasound for the diagnosis of acute appendicitis, in order to ascertain all cases of appendicitis in the tested population, the reference standard not only needs to be the findings at surgery in those who test positive and subsequently have surgery, but also needs to consider any cases of appendicitis in those individuals who test negative and who do not have surgery, at least initially. The reference standard in this case will therefore be a composite of the findings at surgery and clinical follow-up. As we saw in Chapter 2 (Table 2.5), the best type of study to estimate diagnostic accuracy is a study of test accuracy conducted in a consecutive or random sample of individuals suspected of having the condition of interest. Every eligible individual who presents with a similar type of clinical problem in a particular setting (or a random sample of eligible individuals) over a particular time period should be tested with both the test of interest (the index test) and the reference standard. The index test result of each study participant is then compared with the reference standard result. Again, as we saw in Chapter 2, systematic reviews are even better than an individual study or trying to read all the studies that are available. Systematic reviews are discussed further in Chapter 12.

Other study designs are also possible. For example, the study may compare the test results in patients who are selected into the study by convenience or arbitrary methods rather than by selecting consecutive patients. This is a weaker study design because the selection of participants might introduce selection bias. For example, the most conveniently enrolled patients might be those who are generally sicker. Another alternative is the diagnostic 'two-gate' (or two-group) study design, which compares the test results of the index test and reference standard in two separate groups of patients: the first is a group who are known to have the condition of interest (for example, they have tested positive on the reference standard) and a group of patients who are known or assumed not to have the condition of interest, generally because they have no symptoms of the disease in question (often referred to as 'healthy controls').[1] Some studies may enrol several groups of people—for example, a group with no symptoms, a group with symptoms but known to have another disease, and people known to have the disease—and such studies are termed 'multiple-gate' (or multiple-group). Two-gate and multiple-gate studies are also a weak study design because, apart from the participant selection bias, they are unlikely to enrol patients with the whole spectrum of the condition seen in clinical practice. Spectrum bias can result in the test's diagnostic accuracy being overestimated as the patients who are known to have disease and known to not have disease are often patients who are 'easier' to diagnose.[2] p0250

HOW TO STRUCTURE A DIAGNOSTIC TEST ACCURACY QUESTION s0015

CLINICAL SCENARIO (CONTINUED) b0070

Structuring the clinical question

As with clinical questions about the effectiveness of interventions, we can define the clinical question for diagnostic questions using the PICO format that was outlined in Chapter 2. For questions about diagnosis, p0260 the 'I' (Intervention/Issue) component of PICO is the diagnostic test you are interested in, and the Comparison is the reference standard. The Outcome is the diagnosis in question.

p0265 Using the PICO format for questions about test accuracy is sometimes not straightforward, as there may be more than one test of interest and because of the alternative ways a test might be used in practice (for example, as a replacement for an existing test, as an add-on to an existing test or as a triage test before an existing test[3]). Consequently, an alternative format for defining questions about diagnostic test accuracy has been recently proposed by the Cochrane Collaboration.[4] This approach is known as 'PIT', where P describes the people or populations in whom the test may be used, I is the index test, or the test we wish to evaluate, and T is the target condition, or the condition we are trying to diagnose. There may be more than one index test of interest. For example, we might be interested in knowing the accuracy of different urinary biomarkers (separately or in combination) for detecting endometriosis. With the PIT approach, the reference test is not considered to be the comparator in the PICO sense, but rather, something that is used to establish how well the index test performs in detecting the target condition. As the PICO format is still the most widely recognised format for creating a focused clinical question, we have used it in this chapter.

s0020

Patient/population

p0270 In the clinical scenario described at the beginning of this chapter, the population that we are interested in is clinicians working in a hospital ward who have no symptoms of SARS-CoV-2 infection. When considering the patient/population component of PICO, it may be important to specify the setting you are interested in, as diagnostic tests may have different accuracy in different settings. For example, the accuracy of a test may vary in primary and secondary care settings due to the different types of patients that present in these settings. Test accuracy may also vary depending on characteristics of the population (for example, age, gender or ethnicity, or the prior testing individuals may have had). In the clinical scenario above, it may be important to consider if tests perform differently in people who are symptomatic and those who are asymptomatic. If you think the diagnostic test may perform differently in different sub-groups of patients, you may wish to define the population of interest more narrowly. Remember, though, that if you make the population too specific, you may not find any studies.

Intervention

s0025

p0275 For a diagnostic question, this component of PICO relates to the diagnostic test you are interested in. In the clinical scenario discussed in this chapter, the test of interest (or index test) is a single rapid antigen test for detecting infection with the SARS-CoV-2 virus. Antigens are structures on the surface of a virus that are recognised by the body's immune system and can trigger an immune response in an individual. When a sample taken from the nose or throat is mixed with a solution, viral antigens in the solution are unleashed. A small quantity of the solution is placed on a test cartridge and the presence of the antigens can then be 'captured' by the test containing antibodies specific to the viral antigen. In the setting where this test will be used (testing large numbers of healthcare workers), the collection of the sample by the individual being tested will be supervised by another health professional and the sample analysed in an onsite immunofluorescence analyser. The analyser provides a qualitative result (the test is either positive or negative) and the result is documented and reported by the testing supervisor. The result is usually available after about 15–20 minutes.

p0280 In our example, we are considering only a single test, but you may also be interested in the accuracy of a combination of tests, such as a rapid antigen test performed each day for three days. In other diagnostic scenarios, a combination of different tests may be used to identify a condition. For example, if you are interested in the diagnostic accuracy of physical examination for the presence of an anterior cruciate ligament injury of the knee, you may consider a combination of the anterior drawer test, Lachman's test and the pivot shift test. Another form of diagnostic test is a more formal combination of 'tests', such as a clinical prediction rule (which we discuss later in this chapter).

Comparison

s0030

p0285 The comparator test should be the most accurate method of diagnosing the condition of interest. The most accurate test available for detecting SARS-CoV-2 infection is the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test. A PCR test detects the presence of a part of the genome of the virus, and in SARS-CoV-2 tests the test is often trying to detect the part of the viral RNA that codes for the spike protein that allows the virus to enter host cells. Again, the sample is usually taken from the nose and/or throat. The test uses a polymerase chain reaction to produce a reverse transcription of the viral RNA, resulting in a DNA sample. The test is run over several cycles, with each cycle amplifying the quantity of DNA present. When a large quantity of virus is present, the test will only require a few cycles to detect the presence of the virus, but if only a small amount of virus is present, the test requires more cycles to detect the virus. The sample is analysed in a laboratory using complex techniques and by trained technicians and takes longer to obtain a result than the rapid antigen test—generally, several hours.

s0035

Outcome

p0290 This component of PICO relates to the target condition—that is, the disease or condition we want the index test to detect. The study should specify how the condition is defined by the reference standard.

p0295 For SARS-CoV-2 tests, there are a number of possible target conditions that might be relevant for different settings. These include if a person is infected with the virus, if they have COVID-19 disease caused by the virus, if they are infectious, if they have had a past or recent infection with the virus and if they have immunity to infection. To make decisions about the opening of businesses and public gatherings, for example, testing for infectiousness (whether an individual can spread the virus to other people) rather than for the presence of infection may be more useful.[5] Infectiousness may also be of value in the setting of interest; however, as there is currently no reliable reference standard for infectiousness, the outcome we will use is the presence of SARS-CoV-2 infection denoted by the presence of viral RNA.

You decide on the following question: p0300

- **In clinicians working in a hospital ward with no symptoms of COVID-19, how accurate is rapid antigen testing compared to RT-PCR testing as a reference standard for detecting SARS-CoV-2 infection?** u0075

b0080

CLINICAL SCENARIO (CONTINUED)

Finding the evidence to answer your question

p0315 As we saw in Chapter 3, one of the best options for finding diagnostic accuracy studies is PubMed—Clinical Queries. If you are looking for studies on a particular test, you may type in the name of the test and select 'diagnosis' and 'narrow scope'. This may be enough to find what you want. If you do not find anything with a narrow search, you can then look for more studies by selecting 'broad scope'. If the test is used for diagnosing more than one condition, you will also need to type in the name of the condition to narrow the search to only the condition that you are considering (for example, ultrasound AND breast cancer).

p0320 In the current scenario, the test of interest is the rapid antigen test. As these types of tests may be used for detecting other conditions, you will need also to enter terms related to the condition you are trying to detect—infection with the SARS-CoV-2 virus that causes COVID-19 disease. You would need to think about synonyms used to describe the test of interest, including 'RAT', 'RADT', 'antigen test', 'lateral flow' and 'rapid test' and consider how these would be included in your search.

p0325 You go to PubMed's Clinical Queries section and notice there is a filter for COVID-19 studies. You decide to use this filter and the one for 'Diagnosis' studies and enter the terms 'antigen test' OR 'rapid test' OR RAT OR RADT into the search box. Your search retrieves just over 2,400 studies. You know this is too many results to look through, so while trying to think of a way to revise your search, you look through the first dozen results that come up. You notice several systematic reviews of the accuracy of rapid antigen tests in different settings and population groups, including a living systematic review (these are discussed in Chapter 12). After having a quick look at this review, you realise there are many different types of rapid tests available and that their accuracy varies. You recall the name of the particular rapid antigen test proposed for use in your hospital and go back to your PubMed search. Adding the term 'Elecsys' with an AND—for example ('antigen test' OR 'rapid test' OR RAT OR RADT) AND Elecsys with the COVID-19 and Diagnosis filter—retrieves 32 results. You look through these and find a study that appears to be just what you are looking for—a study of the diagnostic accuracy of the Elecsys antigen test for detecting SARS-CoV-2 infection.[6] You obtain the full text of this study for further appraisal.

b0090

CLINICAL SCENARIO (CONTINUED)

Structured abstract of the chosen article

p0335 **Citation:** Adapted from Montalvo Villalba MC, Sosa Glaria E, et al. Performance evaluation of Elecsys SARS-CoV-2 antigen immunoassay for diagnostic of COVID-19. J Med Virol 2021;Oct 2. doi: 10.1002/jmv.27412.[6] The structured abstract is adapted from this reference.

p0340 **Question:** In nasopharyngeal swabs sent to the national laboratory in Cuba, what is the diagnostic performance of the Elecsys SARS-CoV-2 antigen test compared with the SARS-CoV-2 RT-PCR test for detecting SARS-CoV-2 infection?

p0345 **Design:** Nasopharyngeal swabs obtained from individuals meeting 5 different epidemiological definitions of samples used at the laboratory (see 'Participants' section below) are tested with the test of interest – the rapid antigen test, and the reference standard test (RT-PCR) for detecting SARS-CoV-2 infection. As there are more than two groups of study participants (in this case, samples), the study may be described as a 'multiple-group' or 'multiple-gate' diagnostic accuracy study. The study also evaluated the cross-reactivity of the antigen test—that is, whether the test identifying the SARS-CoV-2 virus proteins also detects other viral proteins it is not intended to detect.

p0350 **Setting:** National Reference Laboratory for Respiratory Virus in Havana, Cuba.

p0355 **Participants:** The study included 523 randomly selected nasopharyngeal swab samples received at the laboratory from individuals who were classified as (1) being a case of COVID-19 based on having a positive RT-PCR test for SARS-CoV-2, (2) having contact with a confirmed or suspected COVID-19 case, (3) being a case of COVID-19 5 days after diagnosis, (4) having met the clinical criteria of COVID-19 and contact with a probable case, and (5) being an international traveller arriving in Cuba.

Test: Elecsys SARS-CoV-2 antigen test. p0360

Diagnostic (reference) standard: SARS-CoV-2 RT-PCR test. p0365

Main results: The sensitivity of the Elecsys SARS-CoV-2 antigen test for identifying non-SARS-CoV-2 infection across all samples was 89.7% and specificity was 90.6%. Cross-reactivity to other respiratory viruses was not detected. Sensitivity of the test ranged from 78.4% in samples taken from travellers returning to Cuba (known as the 'surveillance' samples in the study) to 94.2% in samples taken from people tested 5 days after a positive RT-PCR test for SARS-CoV-2. Specificity ranged from 86.8% in surveillance samples to 97.1% in samples collected from people who had had contact with a confirmed or suspected COVID-19 case (referred to as 'contact cases' in the study). p0370

Conclusions: Elecsys SARS-CoV-2 antigen immunoassay may be used as an alternative to RT-PCR testing or in complement with it. p0375

s0040 IS THIS EVIDENCE LIKELY TO BE BIASED?

p0380 As we saw in Chapter 4 for studies about the effectiveness of interventions, it is important to critically appraise the diagnostic test studies that you find to determine whether each study is adequate to inform your clinical practice. As with the other types of study designs, the main elements to consider are: (1) internal validity (that is, the risk of bias); (2) the results (the estimates of diagnostic accuracy); and (3) whether or how the evidence might be applicable to your patient or clinical practice.

p0385 We will use the Critical Appraisal Skills Program (CASP) checklist for appraising a diagnostic test study to explain how to assess the likelihood of bias in this type of study. The key questions to ask when appraising the risk of bias (validity) of a diagnostic study are shown in Box 6.1. The checklist begins with two simple screening criteria (not shown in Box 6.1) that, if not met, indicate that the article is unlikely to be helpful and that further assessment of potential bias is probably unwarranted. The reporting statement for diagnostic accuracy studies is the STARD (STAndards for the Reporting of Diagnostic accuracy studies) statement. Further details are available at: https://www.equator-network.org/reporting-guidelines/stard/.

BOX 6.1 Key questions to ask when appraising the risk of bias (validity) of a diagnostic accuracy study b0030

1. Did all participants get the diagnostic test and the reference standard? o0050
2. Could the results of the test of interest have been influenced by the results of the reference standard, or vice versa? o0055
3. Was there a clear description of the disease/condition status of the tested population? o0060
4. Was there sufficient description of the methods for performing the test? o0065

Was there a clear question for the study to address? s0045

The first screening criterion on the checklist is whether there was a clear question for the study to address. For diagnostic evidence, the study should clearly define the population, the index and comparator tests, the setting and the outcomes considered. p0390 b0100

CLINICAL SCENARIO (CONTINUED)

Did the study address a clearly focused issue?

p0400 The study addressed a clearly focused question. The study's aim of evaluating the accuracy of the Elecsys SARS-CoV-2 rapid antigen test in nasopharyngeal swab samples received by a large national laboratory using RT-PCR as the reference standard is clearly outlined. There are, however, some differences between the study and the clinical scenario that you will need to keep in mind when interpreting the study results. In the study, the 'population' is not people as such, but samples from individuals who have received the index test for various reasons. The test is likely to perform differently in these distinct groups. The setting in which the study was conducted is a large laboratory with highly experienced staff handling and analysing the samples. You would need to consider if the skills and experience of the staff handling and analysing the sample in this study are similar to the situation in your own clinical setting. The target condition, infection with SARS-CoV-2, is the same in the study and the scenario.

s0050

Is the comparison with an appropriate reference standard?

p0405 The second screening criterion is whether there was a comparison with an appropriate reference standard. The reference standard should, in general, be the most accurate method available to diagnose, or test for, the target disorder(s). If the reference test used in the study is not 100% accurate, the diagnostic accuracy of the index test may be either over- or underestimated.

p0410 Sometimes, the reference standard will be a combination of a number of tests. This is often called a 'composite reference standard'. With this type of reference standard, multiple methods/procedures/tests are used and a rule based on these procedures defines who has or does not have the condition of interest. For example, a test for diagnosing heart failure may be assessed against the combined results of clinical examination and echocardiography. If the index test is included in the reference standard (this is called *incorporation bias*), the diagnostic accuracy of the test is likely to be overestimated.

p0415 A reference standard may also include follow-up. Using follow-up to confirm a diagnosis at the time of testing is known as *delayed verification*. An example is a 30-day follow-up of individuals with suspected appendicitis. If a clinical event related to appendicitis (for example, re-presentation in the emergency department and subsequent surgery) occurs during the follow-up period, appendicitis is presumed to have been present at the time of index testing.

b0110

CLINICAL SCENARIO (CONTINUED)

Comparison with an appropriate reference standard

p0425 The reference standard in this study was the SARS-CoV-2 RT-PCR test. This is widely considered the best available test for detecting SARS-CoV-2 infection. However, even RT-PCR tests may not detect infection in the first few days after infection, as the test requires a certain level of viral load to be present in the throat or nose to be positive.[5] It is important to consider how soon after the exposure to potential SARS-CoV-2 infection the swab was taken for the RT-PCR test.

s0055

Did all participants get the diagnostic test and the reference standard?

p0430 As we explained earlier in this chapter, the best type of study to estimate diagnostic accuracy is a study of test accuracy in a consecutive sample of eligible individuals. In a well-designed study, *every* individual who presents with a similar type of clinical problem in a particular setting over a particular time period receives both the test of interest and the reference standard, and the results are compared. In a study looking at the diagnostic accuracy for appendicitis, if we were to only assess the accuracy of the test in those individuals who have surgery, and not those who did not need surgery, we will have a biased estimate of the accuracy of the test. This type of bias is known as *verification bias*.

p0435 A common form of verification bias occurs when the authors of a study use patient records to select patients to include in the study who have had both the index test and the reference test. For example, in a study of the accuracy of physical signs versus the reference standard of arthroscopy or MRI for diagnosing anterior cruciate ligament injury of the knee, individuals were included in the study if they attended an orthopaedic clinic and had *both* a physical examination and an MRI scan. Further, when patient records are used to select patients for a study, study participants are likely to be different from the type of patients who present to a clinic with a particular clinical problem and the sample is therefore likely to provide a biased estimate of the accuracy of the diagnostic test. This is a form of *spectrum bias*. For example, patients who had both a suspected anterior cruciate ligament injury and an MRI scan are likely to be a different spectrum of patients from all patients who present to an orthopaedic clinic with a suspected anterior cruciate ligament injury. Patients who had both a physical examination and an MRI may, for example, have more severe injuries.

p0440 The timing of the reference standard is also important. Ideally, the results of the test of interest and the reference standard are obtained from the same patients at the same time. If this does not occur and the time between the performance of the index and reference standard test is too long (what is too long will depend on the clinical condition), individuals may be misclassified due to spontaneous recovery, response to treatment or progression to a more advanced stage of the condition that can occur during this delay.

b0120 ◎ CLINICAL SCENARIO (CONTINUED)

Did all participants get the diagnostic test and the reference standard?

p0450 For this study, which used nasopharyngeal samples sent to a national laboratory for the diagnosis of SARS-CoV-2 infection, it is likely both the index and the reference standards would have been performed on all the included samples. It is also likely they would have been performed at the same or very similar time, though neither point is clearly reported in this study. The same type of PCR assay was conducted for all samples.

s0060 Could the results of the test of interest have been influenced by the results of the reference standard, or vice versa?

p0455 The results of the index test and the reference test should each be decided without knowledge of the results of the other test. That is, the person who interprets the test should be blinded to the results of the other test. Knowledge of one test result may bias the reading of the other, particularly where the reading is subjective, such as physical examination or the interpretation of imaging results (this is known as *review bias*).

b0130 ◎ CLINICAL SCENARIO (CONTINUED)

Could the results of the test of interest have been influenced by the results of the reference standard, or vice versa?

p0465 No. It is unlikely that the results of one test would have influenced the results of the other.

p0470 This study took place in a laboratory where specialised automated machinery was used to analyse the samples and provide a result based on pre-specified thresholds/cut-offs for positive and negative results on both the rapid test and the reference standard test. It is not likely that laboratory staff knowing the result of the index test would have been able to influence the result of the reference standard, or vice versa. Ideally, though, the study would provide some statement about blinding or reassurance that test results could not influence each other.

Was there a clear description of the disease/condition status of the tested population? s0065

Studies of test accuracy inform us about the behaviour of a test under particular circumstances. As diagnostic accuracy is very much dependent upon the spectrum of included participants, a clear description of the disease stage and severity of the tested population helps us to understand study findings and how they may apply to our situation. The test should be investigated in a clinical setting that is as close as possible to the clinical setting in which it will be used. The spectrum of patients included in the study can affect the sensitivity or specificity, or both, and therefore may affect the observed accuracy of the test. For example, if the study is conducted in a tertiary referral centre (as compared with a general practitioner's office, say), patients may have more severe symptoms, and this may affect the sensitivity and/or the specificity of the physical examination. p0475

◎ CLINICAL SCENARIO (CONTINUED) b0140

Clear description of the disease/condition status of the tested population

Yes. The study describes the samples as being from individuals who were (1) a confirmed case of COVID-19, (2) a contact of a confirmed or suspected COVID-19 case (that is, contact cases), (3) a case of COVID-19 5 days after diagnosis, (4) meeting the clinical criteria of COVID-19 and had contact with a probable case (that is, suspected cases), and (5) an international traveller arriving in Cuba (that is, surveillance cases). The samples used in the study were a random selection of samples received at the laboratory. p0485

Was there sufficient description of the methods for performing the test? s0070

Both the index test and the reference standard test should be described in sufficient detail so that it is possible to: (1) reproduce the test; and (2) determine whether the test was performed adequately and is similar to the test being conducted in your own clinical setting. p0490

◎ CLINICAL SCENARIO (CONTINUED) b0150

Sufficient description of the methods for performing the test

Yes. The study reports specific details of the assays used, the machinery (analysers) models employed and the procedures followed (for example, as per manufacturer's p0500

Continued

CLINICAL SCENARIO (CONTINUED)—CONT'D

instructions for the model). The cut-off values for positive and negative results for both the index and the reference test are provided. Determining the adequacy of the description for performing the test often requires some knowledge of the clinical area, but, as a minimum, studies of laboratory tests should report the type of assay and machinery (including the manufacturer), the process related to the handling and analysis of samples and cut-offs, and categories of results of the index and reference tests.

p0505 If you have got to this point and determined that the article about diagnostic accuracy that you have been appraising is valid, you then proceed to looking at the importance and applicability of the results.

s0075 WHAT ARE THE RESULTS?

p0510 Diagnostic accuracy studies may report the results in a variety of ways. Most studies will report the sensitivity and specificity of the index test. These metrics tell us about the properties of the diagnostic test—that is, its ability to identify who tests positive (in people with the disease or condition) and who tests negative (in those who are disease free)—and help us determine the test's appropriateness. No test is 100% accurate; there will always be false-positive and false-negative results.

p0515 However, the most useful measures for you as a health professional are the post-test probabilities of a positive and negative test. In practice, we do not know whether an individual has the disease or condition of interest (that is why we are doing the test!). When we do the test and receive the results, the post-test probabilities help us interpret that result—how likely it is that the individual actually has the disease when the test is positive and how likely it is that the individual does not have the disease when the test is negative. Post-test probabilities are dependent on the prevalence or pre-test probability of the disease in the population of interest and on the sensitivity and specificity of the test.

p0520 The measures of test accuracy—sensitivity, specificity and post-test probabilities—are obtained from the cross-classification of the reference standard result and the index test result and are commonly presented in the form of a 2 × 2 table. All study participants are classified into one of the four cells of the 2 × 2 table depending on their index and reference test result. An example of a 2 × 2 table showing data from our clinical scenario study is in Table 6.1. The true positives (the upper left cell of the 2 × 2 table) have a positive index test result *and* are classified by the reference standard as having the disease. The true negatives (lower right cell of the 2 × 2 table) have a negative index test result *and* are negative on the reference standard. The false positive (upper right cell) and false negatives (lower left cell) are the number of individuals misclassified by the index test.

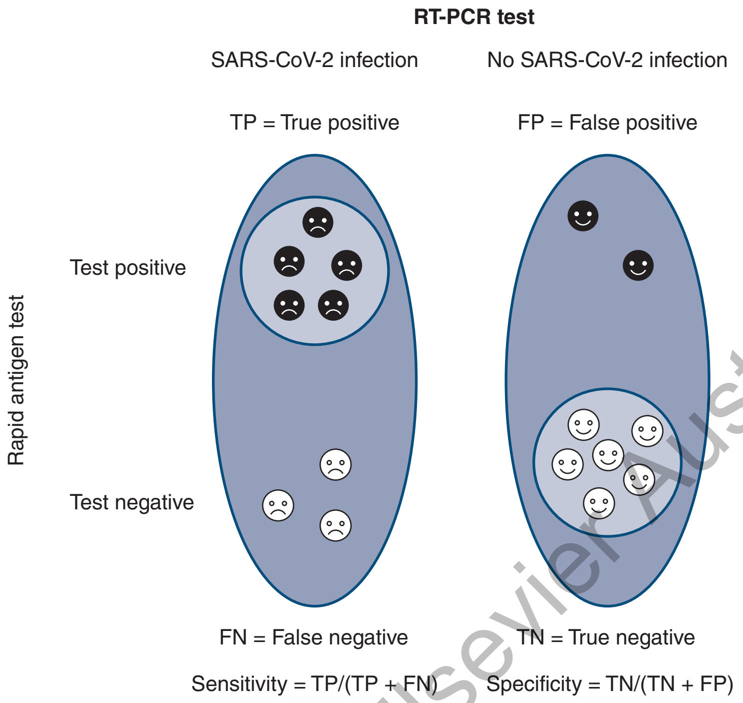
p0525 We will now look at how to interpret and calculate measures of test accuracy.

s0080 Sensitivity and specificity

- u0080 The **sensitivity** of a test measures how well it performs in detecting a condition in people who have the condition. It is the probability that a test is positive in people who have a condition (true positives ÷ [true positives + false negatives]). Using data from our clinical scenario article,[6] this is represented graphically in Figure 6.1.
- u0085 The **specificity** of a test measures how well it performs in determining that a condition is *not* present in people who do not have the condition. It is the probability that

t0010 TABLE 6.1 2 × 2 cross-classification of the index test (the rapid antigen test) for detecting SARS-CoV-2 infection and the reference standard (RT-PCR test) in samples collected from returning international travellers (i.e. the surveillance samples) as reported by Montalvo Villalba et al.[6]

		REFERENCE STANDARD RESULT		
		SARS-CoV-2 infection	No SARS-CoV-2 infection	Total
Index test result	Rapid antigen test positive	True positives 40	False positives 9	49
	Rapid antigen test negative	False negatives 11	True negatives 59	70
	Total	**51**	**68**	**119**



f0010 **Fig 6.1** Graphical representation of sensitivity and specificity

a test is negative in people who do not have the condition (true negatives ÷ [true negatives + false positives]). Using data from our clinical scenario article, this is represented graphically in Figure 6.1.

p0545 Box 6.2 shows how to calculate the sensitivity and specificity of the rapid antigen test being evaluated in the clinical scenario article.

s0085 Post-test probabilities of a positive and a negative test

p0550 These values tell us about the clinical relevance of a test and are the most useful way of interpreting the results of a test accuracy study for you as a health professional:

- u0090 The **post-test probability of a positive test** (also known as **positive predictive value**) tells you the probability that a patient has the condition if they have a positive test result. The closer this number is to 100%, the better the test is at ruling in the condition. Its calculation (true positives ÷ [true positives + false positives]) is represented graphically in Figure 6.2, using data from our clinical scenario article.
- u0095 Conversely, the **post-test probability of a negative test** (which is the *complement* of the **negative predictive value**) tells you the probability that a patient has the condition if they have a negative test result. The closer this number is to 0%, the better the test is at ruling out the condition (as there will be few false negatives). Its calculation (false negatives ÷ [false negatives + true negatives]) is represented graphically in Figure 6.2, using data from our clinical scenario article. The closer a *negative predictive value* approaches 100%, the better the test is at ruling out the condition. Its calculation is true negatives ÷ (false negatives + true negatives).

p0565 The difficulty with post-test probabilities (positive and negative predictive values) is that you need to have an estimate of the pre-test probability of the condition (that is, the likelihood of having the condition before having the test) in order to be able to calculate them. In testing for SARS-CoV-2 infection, for example, the pre-test probability of infection will depend on the rate of new diagnosis of COVID-19 in the community at the time of testing and the level of clinical suspicion that a person has been infected (for example, whether symptoms are present in an individual or there has been a close contact). The pre-test probability of being infected will be higher for a person with a fever, sore throat and exposure to another infected individual, than for a person who has no symptoms and when there is low rate of new cases of infection in the community at the time.

p0570 When more individuals in the study have the condition, the post-test probability of both positive and negative tests will increase.[7] So, if you use post-test probabilities to guide b0020

BOX 6.2 Measuring diagnostic accuracy: sensitivity and specificity

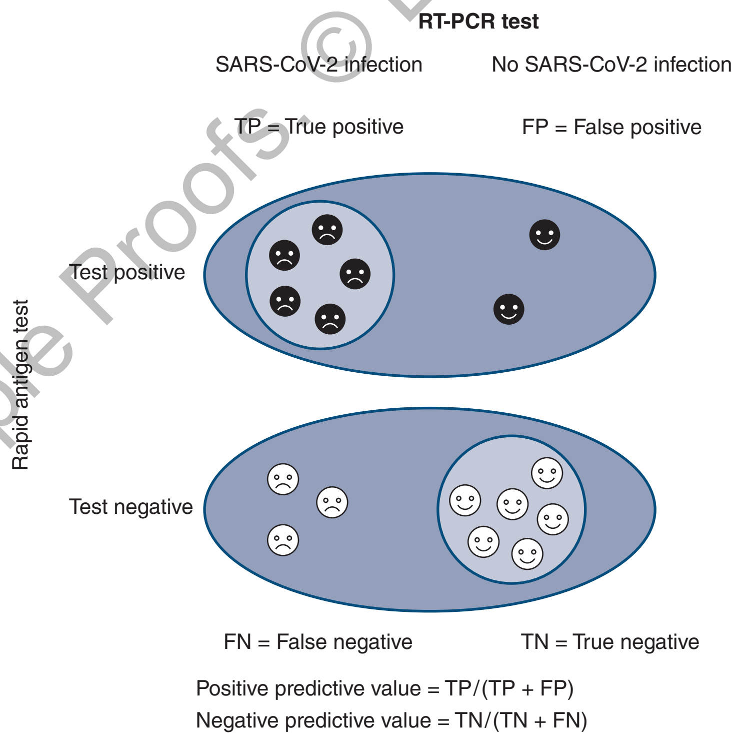
p0055 This box uses data (see Table 6.1) about the diagnostic accuracy of the rapid antigen test for detecting SARS-CoV-2 infection in the samples collected from returned international travellers from our clinical scenario article as an example.

$$\begin{aligned}\text{The sensitivity of the rapid antigen test} &= \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}\\ &= \frac{40}{(40+11)}\\ &= \frac{40}{51}\\ &= 78.4\%\end{aligned}$$

e0010

$$\begin{aligned}\text{The specificity of the rapid antigen test} &= \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}\\ &= \frac{59}{(59+9)}\\ &= \frac{59}{68}\\ &= 86.8\%\end{aligned}$$

e0015



f0020 **Fig 6.2** Graphical representation of post-test probabilities of positive and negative tests b0040 **BOX 6.3 Measuring diagnostic accuracy: post-test probabilities of a positive and a negative test result**

p0085 As a health professional, what you want to know is the probability that a patient has a condition if you receive a positive or a negative test result for them. These values are the post-test probabilities of a positive and a negative test. However, most diagnostic accuracy studies report the sensitivity and the specificity of a diagnostic test. The probability of a condition after a positive or a negative test result requires further calculation, and we also need to consider the *prevalence* (also called the 'pre-test probability') of the condition.

p0090 *Using some of the data from our clinical scenario article as an example:*

p0095 The rapid antigen test in the samples collected from returned international travellers (the surveillance samples) had a sensitivity of 78.4% and a specificity of 86.8%. The study reports that 51 of 119 of these samples had a positive RT-PCT test for SARS-CoV-2 infection. The **prevalence or pre-test probability** of SARS-CoV-2 infection in this population is therefore 51 ÷ 119 = 42.9%.

$$\begin{aligned}
&\text{The post-test probability of a positive test (also known as the 'positive predictive value')}\\
&\quad = \text{the probability of SARS-CoV-2 infection with a positive rapid antigen test}\\
&\quad = \frac{\text{true positives}}{(\text{true positives} + \text{false positives})}\\
&\quad = \frac{40}{(40+9)}\\
&\quad = \frac{40}{49}\\
&\quad = 81.6\%
\end{aligned}$$ e0020

$$\begin{aligned}
&\text{The post-test probability of a negative test (the \textit{complement} of the negative predictive value)}\\
&\quad = \text{the probability of having SARS-CoV-2 infection given a negative rapid antigen test}\\
&\quad = \frac{\text{false negatives}}{(\text{false negatives} + \text{true negatives})}\\
&\quad = \frac{11}{(11+59)}\\
&\quad = 11/70\\
&\quad = 15.7\%
\end{aligned}$$ e0025

p0100 To help people remember whether tests rule in or rule out a condition, the following mnemonics may be helpful:

- u0010 **SpPIn** (Specificity-Positive-In) = if a test has a high specificity and the result is positive, it rules the condition in.
- u0015 **SnNOut** (Sensitivity-Negative-Out) = if a test has a high sensitivity and the result is negative, it rules the condition out.

p0115 Note that this is a generalisation, and that the post-test probability depends on both sensitivity and specificity, and on the prevalence of the condition.[8]

p0120 *Note:* When the pre-test probability is low—for example, in screening programs—even tests with high sensitivity and specificity will have a low positive predictive value; that is, most positive test results will be false positives.

your decision about whether to use a diagnostic test or not, this means it is particularly important that you check the spectrum of patients that were included in the diagnostic accuracy study to ensure they match the sort of patients you see in your practice. Box 6.3 explains how to calculate post-test probabilities of positive and negative test results, as well as the pre-test probability of the condition.

Positive and negative likelihood ratios s0090

p0575 Another pair of values that can be used to report the results of diagnostic test accuracy studies is the **positive and negative likelihood ratios.** Box 6.4 shows how likelihood ratios can be calculated. These results have the advantage of being relatively stable across different clinical settings, but also give an indication of how well the test rules in or rules out a condition.

b0050 **BOX 6.4 Measuring diagnostic accuracy: positive and negative likelihood ratios**

p0125 The *positive likelihood ratio* is the probability that a test is positive in people with the condition divided by the probability that the test is positive in people without the condition.

p0130 The *negative likelihood ratio* is the probability that a test is negative in people with the condition divided by the probability that the test is negative in people without the condition.

p0135 *Using some of the data from our chosen article as an example:*

The positive likelihood ratio for the rapid antigen test

$$= \frac{(\text{true positives / people who have the condition})}{(\text{false positives / people who do not have the condition})}$$

$$= \frac{(40/51)}{(9/68)}$$

$$= 0.784 / 0.132$$

e0030 $$= 5.9$$

The negative likelihood ratio for the rapid antigen test

$$= \frac{(\text{false negatives / people who have the condition})}{(\text{true negatives / people who do not have the condition})}$$

$$= \frac{(11/51)}{(59/68)}$$

$$= 0.216 / 0.867$$

e0035 $$= 0.25$$

p0140 If the article only reports the sensitivity and specificity of the tests, another way to calculate likelihood ratios is:

- u0020 Positive likelihood ratio (LR+) = sensitivity/(100 – specificity)
- u0025 Negative likelihood ratio (LR−) = (100 – sensitivity)/specificity

p0155 When interpreting likelihood ratios, as a rough guide:

- u0030 A positive likelihood ratio >2 indicates a test that helps rule in the condition.
- u0035 A positive likelihood ratio >10 is an extremely good test for ruling in the condition.
- u0040 A negative likelihood ratio of <0.5 indicates a test that helps rule out the condition.
- u0045 A negative likelihood ratio of <0.1 is an extremely good test for ruling out the condition.

b0160 **CLINICAL SCENARIO (CONTINUED)**

What are the results?

p0585 We will focus our attention on the results for the samples taken for screening international travellers returning to Cuba (this group of samples is referred to as ‘surveillance’ samples in the study), as this is the scenario that is closest to our clinical scenario. These samples were from people who were not known to have had contact with other people who had had COVID-19. The study analysed 119 samples collected from returned travellers.

p0590 In this group of samples, RT-PCR identified SARS-CoV-2 infections in 51 of the 119 samples (42.9%). In these samples, the rapid antigen test performed reasonably well for ruling in (positive likelihood ratio of 5.9) and ruling out (negative likelihood ratio of 0.25) SARS-CoV-2 infection. The sensitivity of the rapid antigen test, or the ability of the test to yield a positive result when an individual had SARS-CoV-2, in the samples of returned travellers was 78.4%. Specificity, or the ability of the test to obtain a negative result for an individual who did not have SARS-CoV-2 infection, was 86.8% (Table 6.2). Cross-reactivity with other respiratory viruses was not detected.

t0020 **TABLE 6.2 Estimates of the diagnostic accuracy of the rapid antigen test for the detection of SARS-CoV-2 infection in samples collected from returning international travellers (i.e. the surveillance samples) as reported by Montalvo Villalba et al.[6]**

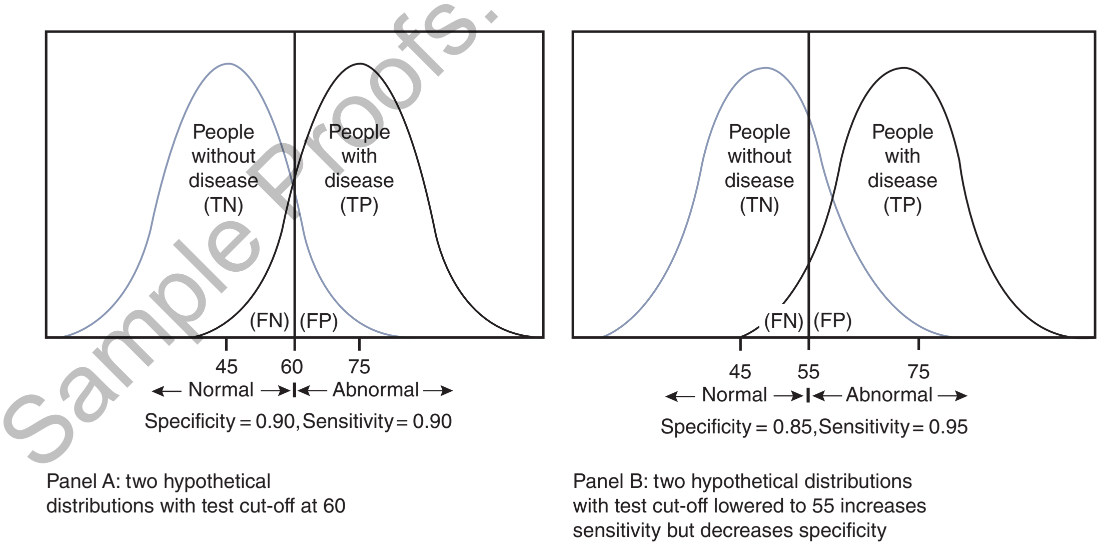
	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio	Positive predictive value	Negative predictive value
Rapid antigen test	78.4%	86.8%	5.9	0.25	81.6%	84.3%

s0095 How changes in the cut-off affect test performance

p0595 For many conditions, there is no clear threshold between the presence and absence of a condition. For example, blood pressure and blood glucose levels exist on a spectrum, and the cut-offs that have been chosen to define hypertension or diabetes are, to some extent, arbitrary. In cases where the cut-off for normal/abnormal levels can be raised or lowered, this will affect the test characteristics, and the choice of cut-off will involve a trade-off between the sensitivity and the specificity of the test. If higher values indicate more-abnormal test results, and the cut-off point is raised, there will be fewer false positives (increased specificity); on the other hand, there will be more false negatives (reduced sensitivity). If the cut-off point is lowered, there are fewer false negatives but more false positives—the test sensitivity increases, but specificity decreases (see Figure 6.3). A receiver operating characteristic (ROC) curve (see Figure 6.4) plots this trade-off between sensitivity and specificity with changes in the cut-off. The curve demonstrates the trade-off between sensitivity and specificity of a test as the cut-off point changes. A test that performs no better than a coin toss would have a ROC curve that traces a straight (diagonal) line from the bottom left to the top right-hand corner of the ROC box. The top left-hand corner of the ROC box is the point where sensitivity = 100% and specificity = 100% (1 – specificity = 0%). This represents a perfect test. The closer the ROC curve is to the top left-hand corner of the ROC box, the better the test overall.

How can we use this evidence to inform practice? s0100

As part of our judgment about whether to use the results of this study in our own practice, we need to think about p0600

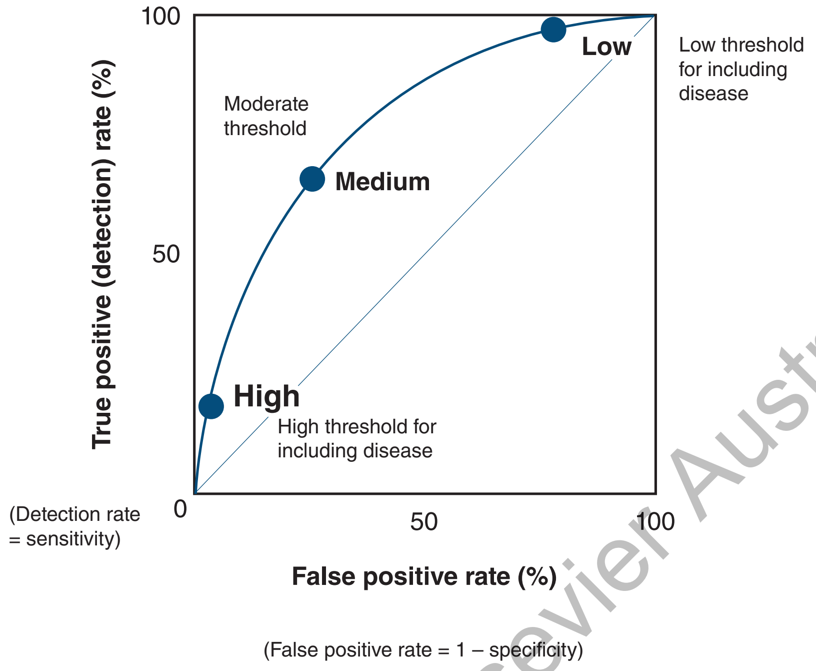


Panel A: two hypothetical distributions with test cut-off at 60

Panel B: two hypothetical distributions with test cut-off lowered to 55 increases sensitivity but decreases specificity

TN = true negative, TP = true positive, FN = false negative, FP = false positive

f0030 **Fig 6.3** The effect of lowering test cut-off on sensitivity and specificity
From Dawson B, Trapp RG. Basic and clinical biostatistics. 4th ed. New York: McGraw-Hill Education; 2004, Ch 12, p 314.



f0040 **Fig 6.4** Receiver operating characteristic (ROC) curve

how likely it is that the test performs in a similar way in our own clinical setting to the diagnostic accuracy in this study.[9] We need to consider:

o0070 1. Is the spectrum of patients in the diagnostic study similar to the spectrum of patients in the clinical setting in which you are working?

o0075 2. Is the prevalence of the condition in the diagnostic study similar to the prevalence of the condition in the clinical setting in which you are working?

3. Is the method for using the index test similar in the diagnostic study and the clinical setting in which you are working? This includes both the method for performing the index test and the person performing the test. o0080

4. Is the method for using the reference test similar in the diagnostic study and the clinical setting in which you are working? o0085

5. Is the study defining the target disorder in the same way as in your own clinical setting? o0090

b0170

CLINICAL SCENARIO (CONTINUED)

Using the evidence to inform practice

p0635 Tests that provide accurate results within a short time frame may facilitate timely decisions around the need for isolation and contact tracing activities, ultimately reducing transmission of SARS-CoV-2 infection. Because of the time delay in receiving results and the resource constraints involved with PCR testing, rapid antigen testing may be a good substitute for PCR testing.

p0640 The findings from the group of returning travellers show that this particular rapid antigen test is reasonably good at detecting the presence and confirming the absence of SARS-CoV-2 infection but will miss some who actually have a SARS-CoV-2 infection and label some who truly do not have SARS-CoV-2 infection as positive.

An important consideration is the timing of the rapid antigen and reference standard tests in relation to when a person may have been exposed to the virus. People who are infected with SARS-CoV-2 will have a positive PCR test earlier than a rapid antigen test (about 2–3 days after exposure for a PCR test and 4–5 days for a rapid antigen test).[5] This is because the rapid antigen test requires a higher viral load before it is able to detect the infection. This may be appropriate in a workplace setting, where it is more important to determine if someone is infectious (which is a risk to other people in the workplace) rather p0645 than if someone is infected. It is important to remember that a person who tests negative one day may be in the early stages of infection, so repeat testing may be required. If someone has symptoms, it is therefore more important to determine if they have the infection and not just if they are possibly infectious, a RT-PCR test may be more appropriate where available.

p0650 As described above, the post-test probability of disease is greatly affected by the pre-test probability of disease. Table 6.3 shows the probabilities of infection given a positive and negative rapid antigen test result when the probability of infection prior to doing the test is low (1%) and higher at 10% and at 43% (the proportion of samples from the returned travellers tested in the study that were positive on RT-PCR testing). As can be seen from the table, when the pre-test probability of infection is low, the positive predictive value will be low and most positive tests will be false positives. When the pre-test probability is 1%, the probability that a person with a positive rapid antigen test is actually infected is only 5.6%. Even when the pre-test probability is higher at 43%, and the positive predictive value is therefore higher at 81.7%, this means approximately 2 in 10 rapid antigen positive results will be falsely positive. In this situation, repeat rapid antigen testing or RT-PCR testing will be necessary to confirm infection and avoid unnecessary exclusion from the workplace.

p0655 False negatives are also of concern. At 43% pre-test probability in the surveillance sample, and negative predictive value of 84.2%, 15.8% of those with negative rapid antigen test results are missed cases of SARS-CoV-2. The potential effect on transmission of infection from missed cases needs to be considered. The role and position of the rapid antigen test in the clinical scenario of interest should be specified and leads to the following further questions. What tolerance is there for false positive and false negative results in the hospital setting of interest? What further testing (if any) happens or what action is taken, given a positive and negative result?

s0105 Could the quality of the study have biased the results?

p0660 Our earlier appraisal of this study suggests that the risk of bias is likely to be low. It is likely all samples were tested with both the rapid antigen test and the reference standard test, and that the results of one would not influence the results of the other. The samples are clearly described according to type of sample, which facilitates understanding of the risk of infection and the methods for performing the tests are clearly described. It is unlikely the quality of the study will have biased the results.

s0110 Could other factors have affected the results—for example, the setting of the study?

p0665 Other factors which may affect the results of the rapid antigen tests include whether the person taking the sample and performing the test is trained, and whether the sample is taken from the nose or the throat. The accuracy of the rapid antigen test may also vary with the procedure used to store and transport the test. When thinking about the possible consequences of testing, it is important to consider how likely it is that the person has been exposed to possible infection, and whether there is potential for repeat testing using either rapid antigen tests or RT-PCR.

s0115 Should I use these tests?

p0670 As a healthcare worker who will be required to undergo rapid antigen testing before each shift, understanding the accuracy of the test being used is important. The rapid time to a result is a key driver for use of the test in this setting, but the trade-off is the possibility of false positive and negative results.

t0030 TABLE 6.3 **Post-test probabilities of rapid antigen testing based on the pre-test probability of infection and sensitivity and specificity of the test**

Pretest probability (the likelihood of infection before having the rapid antigen test)	Sensitivity of the rapid antigen test	Specificity of the rapid antigen test	Positive predictive value/post-test probability of a positive test	Negative predictive value	Post-test probability of a negative test*
1%	78.4%	86.8%	5.6%	99.8%	0.2%
10%	78.4%	86.8%	39.7%	97.3%	2.7%
43%	78.4%	86.8%	81.7%	84.2%	15.8%

*The post-test probability of a negative test is the complement of the negative predictive value. It is the probability of SARS-CoV-2 infection when a negative rapid antigen test result is received.

s0120

OTHER TYPES OF TEST STUDIES

p0675 So far, we have considered diagnostic test accuracy. Studies of diagnostic test accuracy measure how well a test can correctly identify or rule out a disease. But not all studies about tests aim to measure accuracy. Some studies measure the *reliability* of a test; that is, whether you get the same test result when the test is done by different health professionals or by the same health professional at different times. The first are usually called *studies of inter-observer reliability* and the latter *studies of intra-observer reliability*.[10] The agreement between different operators of the test (or different groups of operators) can be assessed using measures of agreement such as Cohen's kappa scores. These scores measure the agreement that is seen beyond that expected by chance.

p0680 Other clinical tests are used for assessing or monitoring patients. For example, haemoglobin A_{1c} (HbA_{1c}) can be used to monitor glycaemic control in patients with diabetes. Other monitoring tests, such as assessments of ability to perform self-care skills or assessments of pain, can be used to monitor a patient's progress, predict the likelihood of their needing further treatment, and/or monitor their response to intervention and whether adjustments to intervention are needed.

p0685 Tests that are used for monitoring need to be reliable, and they are evaluated using measures of reliability such as those described above. Sometimes in clinical practice we use the *average* of several measures to improve the reliability of a test. For example, by taking an average of several blood pressure measurements, we reduce the random error that would be seen in a single measurement. When tests are used to monitor a patient, the most appropriate study design is a randomised controlled trial. In these clinical settings, the test is being used as part of a strategy to intervene in the patient's clinical course. Therefore, these tests should be evaluated in the same way as other interventions (see Chapter 4), and preferably by using outcomes that are clinically relevant to the patient.[11]

p0690 Studies may also be undertaken to determine the clinical utility of tests—that is, the ability of a test to improve clinical outcomes.[12] The availability of an accurate and reliable test does not necessarily translate into better outcomes for people. This is because there are many 'mechanisms' that affect health outcomes once a diagnosis is made, including the degree to which the diagnosis affects treatment plans and the certainty with which a course of treatment is pursued, and the treatment implemented (its timing, efficacy and adherence to it). The test-treatment randomised controlled trial is regarded as the ideal study design for evaluating the clinical utility of a test. In these studies, participants are randomised to the new or existing test, followed by management based on the test results with measurement of patient outcomes.[13]

DIAGNOSTIC CLINICAL PREDICTION RULES

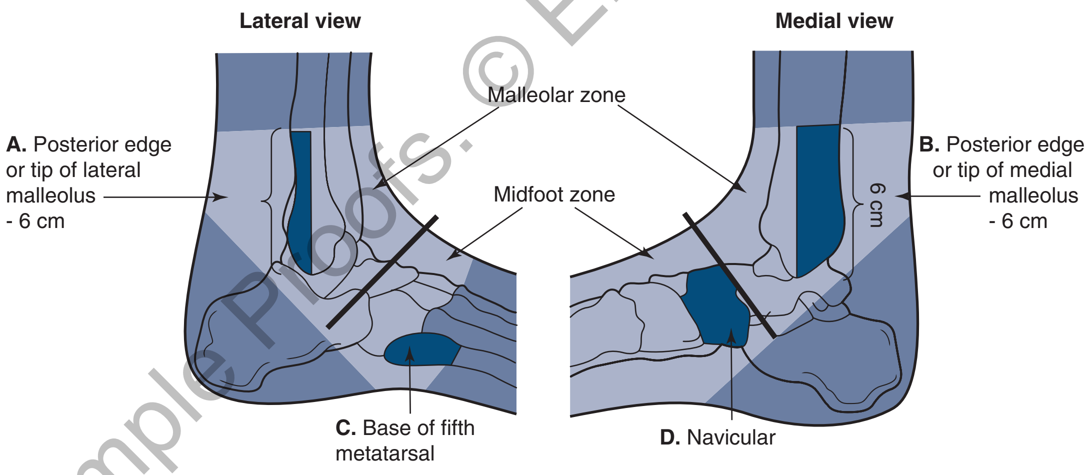
s0125

p0695 Making a diagnosis usually involves the interpretation of multiple 'pieces' of information obtained through questioning the patient about signs and symptoms, performing an examination and/or conducting laboratory or imaging tests. In assessing children with fever for the presence of serious bacterial infection, for example, over 40 clinical signs and symptoms may be considered.[14] However, assimilating and interpreting the often large amount of diagnostic information we collect is challenging and error prone. We may discount some of the diagnostic information we have when the amount of it is overwhelming, or we may misunderstand or misinterpret the 'diagnosticity' or diagnostic value of the information. This may occur because diagnostic test results are often mutually dependent. In other words, the diagnostic information conveyed by the results of different tests is, to varying extents, overlapping and dependent on the information obtained from previous tests. For example, consider two tests—test A and test B—that might be used in the assessment of individuals with chest pain presenting to the emergency department. Both tests measure enzymes found in the blood that are released when there is damage to heart muscle cells. When evaluated on their own (that is, test A is compared to a reference standard and test B is compared to a reference standard), both tests show diagnostic value. However, because the enzymes occur through a related pathological mechanism, the two tests are not independent, and the value of a positive test B is influenced by a positive test A. This means that, in a diagnostic workup, test B has little or no diagnostic value when test A has already been performed. Knowing just which pieces of information have true diagnostic value, and the relative contribution or 'diagnostic power' each test or piece of information makes towards the diagnosis, is a challenge.

p0700 Diagnostic clinical prediction rules are tools that have been developed to assist clinicians to efficiently and objectively combine multiple pieces of diagnostic information. Essentially, clinical prediction rules are combinations of features, or 'predictors' (such as patient characteristics including age and sex, items from the patient's history, physical examination or imaging or laboratory test results), that provide the probability of a 'diagnosis' for an individual. The predicted probabilities assist diagnostic decision making, helping to rule in a condition by identifying individuals very likely to have it (and who thus may require further testing or treatment), or to rule out a condition by identifying those very unlikely to have it (thereby reducing unnecessary testing or treatment).

p0705 Contemporary diagnostic prediction rules are typically developed by applying multivariable statistical techniques (usually, logistic regression) to patient datasets where both the possible predictors and outcome (the presence or absence of the condition of interest) are measured in each participant at the same time (a type of cross-sectional study). The statistical techniques used identify only the predictors that are truly predictive of, or most useful for identifying, the condition of interest in view of other test results (in a way, 'accounting' for the mutual dependency between tests). By entering data from an individual on these predictors, we can obtain an estimate of the probability of a condition (0–100%) for that individual. We may use this objective estimate of the probability in combination with our clinical judgment to assist in reaching a diagnosis when we are uncertain, or as a sort of 'second opinion' when a diagnosis or decision has already been reached. Using a diagnostic prediction rule may also help you to focus on the predictors/features/diagnostic information that are truly useful for identifying the condition of interest, and to give less importance to features that have less predictive power, making the diagnostic process simpler and more efficient. Earlier clinical prediction rules were predominantly developed based on expert opinion. The well-known APGAR score for assessing the health of newborns (developed in the 1950s) is an example of this type of clinical prediction rule.

p0710 Diagnostic clinical prediction rules are often presented as scoring systems, where each predictor in the prediction rule is assigned a point value if present or absent. The points are then summed to give a 'score' which corresponds to a risk probability estimate. Sometimes developers of clinical prediction rules apply cut-offs to the probability estimate provided by the clinical prediction rule, so the prediction rule classifies individuals into risk groups—for example, high, intermediate or low risk of a condition being present. The diagnostic clinical prediction rule may further recommend a course of action based on these groups. This may be a recommendation on further testing or treatment, or both. In some cases, the cut-off may be set at a point that the prediction rule effectively rules out a condition when certain criteria are met (or alternatively rules in a diagnosis, though these are less common). A well-known diagnostic clinical prediction rule using this approach is the Ottawa Ankle Rules (Figure 6.5), which recommends that a foot or ankle X-ray series be performed


Lateral view
Medial view
Malleolar zone
A. Posterior edge or tip of lateral malleolus - 6 cm
Midfoot zone
B. Posterior edge or tip of medial malleolus - 6 cm
6 cm
C. Base of fifth metatarsal
D. Navicular

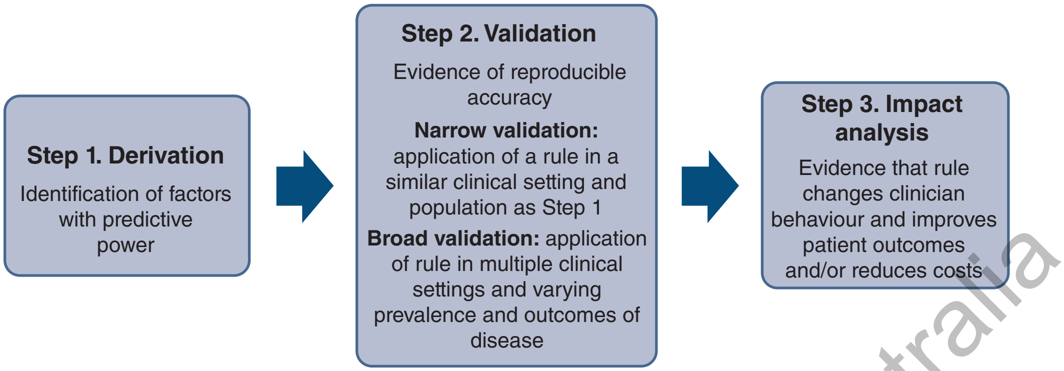
A series of ankle X-ray films is required only if there is any pain in the malleolar zone and any of these findings:

- Bone tenderness at **A**
- Bone tenderness at **B**
- Inability to bear weight both immediately and in emergency department

A series of ankle X-ray films is required only if there is any pain in the mid-foot zone and any of these findings:

- Bone tenderness at **C**
- Bone tenderness at **D**
- Inability to bear weight both immediately and in emergency department

f0050 **Fig 6.5** The Ottawa Ankle Rules
Bachmann LM, Kolb E, Koller MT, et al. Accuracy of Ottawa Ankle Rules to exclude fractures of the ankle and mid-foot: systematic review BMJ 2003;326:417. doi:10.1136/bmj.326.7386.417.


Step 1. Derivation
Identification of factors with predictive power
Step 2. Validation
Evidence of reproducible accuracy
Narrow validation: application of a rule in a similar clinical setting and population as Step 1
Broad validation: application of rule in multiple clinical settings and varying prevalence and outcomes of disease
Step 3. Impact analysis
Evidence that rule changes clinician behaviour and improves patient outcomes and/or reduces costs

f0060 **Fig 6.6** The stages of evaluation of a diagnostic prediction rule

only when certain findings are present. For example, a series of ankle x-rays is only necessary if there is pain near the malleoli, the patient is unable to bear weight and has bone tenderness at specific sites.[15] If one of these findings is not present, an ankle X-ray series is not required.

p0715 Before a diagnostic clinical prediction rule can be considered for use in practice it should have been through three main stages of development and evaluation. It is important to know the stage of evaluation of the rule, as it gives an indication of the 'readiness' of the rule for application in clinical practice. These main stages are: derivation, validation and impact analysis (Figure 6.6).[16] Others have identified a further three stages: the need for a prediction rule, determining the cost-effectiveness of a prediction rule, and long-term dissemination and implementation of the clinical prediction rule.[17] In the derivation phase, the diagnostic clinical prediction rule is developed by applying statistical modelling techniques (often logistic regression modelling) to data obtained from individuals suspected of having the condition and in which the presence of the condition of interest is reported. These statistical techniques identify the predictor variables (characteristics, signs, symptoms or diagnostic tests) statistically related to the presence or absence of the condition of interest. The diagnostic performance of this combination of predictors compared to a reference test is then evaluated. Because the performance of prediction rules is usually overestimated when the rule is used in the same data in which it was developed, its performance should be evaluated in other patient data than was used for the rule derivation. Such validation studies may use participant data collected by the same investigators but at a later time period (temporal or narrow validation) or by other investigators in a different geographical location (geographic or broad validation) or setting (model developed in secondary care and used in a primary care population). A diagnostic prediction rule may also be validated in other types of participants entirely (for example, a model developed for adults but validated in a population of children).

p0720 The performance of prediction rules in validation studies is typically measured in terms of calibration—the agreement between the predictions of the rule and the observed outcomes, and discrimination—the ability of the rule to differentiate between those who do and who do not have the outcome of interest. A widely reported measure of discrimination of a prediction rule is the area under the receiver curve. The AUROC is a measure of the Area Underneath the ROC curve (seen earlier in the chapter) that represents how likely it is that the prediction rule will rank two individuals, one with and one without the condition, in the correct order across all possible thresholds. An AUROC of 1 represents a perfect test, while 0.5 represents a test no better than a coin toss and therefore not worth doing. Sensitivity and specificity and predictive values might also be used to quantify the performance of a diagnostic clinical prediction rule; however, as with other diagnostic tests with continuous outcomes, a cut-off point must be applied to the probability provided by the diagnostic prediction rule to classify individuals as high or low risk.

p0725 The final stage is to evaluate the use of the clinical prediction rule and the effect of its use on patient outcomes. This step generally requires a comparative study. A randomised controlled trial is ideal. At minimum, the diagnostic clinical prediction rule should demonstrate good performance in broad validation studies before being considered for use in practice. But caution should be exercised if incorporating the diagnostic clinical prediction rule into the clinical decision-making process without careful evaluation of its effects on patient outcomes.

p0730 Diagnostic clinical prediction rules are also called prediction models or guides, decision rules or guides, scoring systems, algorithms, decision support systems or risk scores. This can make it difficult to locate studies of clinical prediction rules in the literature. Studies of diagnostic clinical prediction rules may be found in the clinical literature by searching in resources such as syntheses such as Evidence Updates where they are appraised and rated for relevancy and newsworthiness (see Chapter 3). Appraised studies of diagnostic prediction rules may also appear in specific discipline databases such as the Diagnostic Test Accuracy database for physiotherapists (www.dita.org.au), which was described in Chapter 3. If searching for studies of diagnostic clinical prediction rules in bibliographic databases such as Medline via Ovid or PubMed, you can limit your search to studies of clinical prediction guides using the 'Clinical Queries' feature and for PubMed by using the 'Clinical Queries' screen, 'clinical prediction rule' filter.

p0735 When you find a study of a clinical prediction rule, you may need to appraise it (if it has not already been appraised by synopses or pre-appraised sources) to determine the stage of evaluation of the rule and the rigour of the methods used to derive and validate it. The Critical Appraisal Skills Program (CASP) has produced a checklist for appraising a study about a clinical prediction rule. The key questions to ask when appraising the validity of a study of a clinical prediction rule are summarised in Box 6.5. The checklist begins with three simple screening criteria. The first two questions ask about the clinical prediction rule under study: whether it is adequately presented in the paper with regard to how and with whom to use it, and the population in which the rule was derived. If the study you have describes the derivation of a clinical prediction rule, this should be easy to answer. If, instead, the study you have is a validation or impact study, to answer this question you may need to check the references and obtain the derivation study if detail about the derivation is not given in the paper. The third question—Was the rule validated in a different group of patients?—assists in determining the level of evaluation of the clinical prediction rule—that is, whether it is a study of the development of the clinical prediction rule, or a validation study assessing the performance of the rule in a population other than that in which it was developed. If these screening criteria are not met, further assessment of potential bias may not be warranted. If you find a study of the impact of a clinical prediction rule, you may need to use a different appraisal checklist depending on the design of the impact study. The CASP checklist we saw in Chapter 4 would be suitable to appraise a randomised controlled trial of the impact of a clinical prediction rule. If you have located a systematic review of clinical prediction rules, appraise the study using the checklist for systematic reviews presented in Chapter 12.

BOX 6.5 Key questions to ask when appraising the risk of bias (validity) of a clinical prediction rule study b0010

1. Is the clinical prediction rule clearly defined? o0010
2. Did the population from which the rule was derived include an appropriate spectrum of patients? o0015
3. Was the rule validated in a different group of patients? o0020
4. Were the predictor variables and the outcome evaluated in a blinded fashion? o0025
5. Were the predictor variables and the outcome evaluated in the whole sample selected initially? o0030
6. Are the statistical methods used to construct and validate the rule clearly described? o0035
7. Can the performance of the rule be calculated? o0040
8. How precise are the results? o0045

s0130 SUMMARY

- u0100 The diagnostic accuracy of a test is best assessed by a study of the test against a 'gold-standard' reference test in a consecutive series of patients presenting with a clinical problem.
- u0105 The accuracy of single tests (for example, an X-ray) or combinations of tests (for example, clinical examination which usually incorporates information from the patient's history and physical examination) can be assessed.
- u0110 Some of the main risks of bias in a diagnostic accuracy study are: (1) only a selected portion of the patients who receive the index test also receive the reference test (a form of *verification bias*); (2) the study does not include patients with the whole spectrum of the condition that would be seen in clinical practice (*spectrum bias*); (3) not all patients suspected of having the condition are included in the study consecutively or via a process of random selection (*selection bias*); (4) when the results of the test and reference standard are not interpreted independently from the other test or blinded to the results of the other test (*review bias*); and (5) when the test of interest is part of the reference standard (*incorporation bias*).
- u0115 The most common methods for reporting the results of a diagnostic accuracy study are the sensitivity and specificity of a test. However, the most useful results for a health professional are the post-test probabilities

Continued

SUMMARY—cont'd

of a positive and a negative test, or the positive and negative likelihood ratios.

- u0120 Along with the assessment of the risk of bias in a diagnostic accuracy test, it is also necessary to think how the results may be affected by the setting of the study and the types of patients included in it.
- u0125 Using the results of a diagnostic accuracy study can help you to decide whether the test is useful at ruling in or ruling out the diagnosis, or both.
- Tests are also used for assessing and/or monitoring patients, and studies reporting about tests used for this purpose should also be critically appraised. u0130
- Diagnostic clinical prediction rules can be used to assist you during the process of making a diagnosis. Studies of diagnostic clinical prediction rules should be appraised to determine their readiness for clinical use. u0135

REFERENCES

1. Bossuyt PMM. Chapter 3: Understanding the designs of test accuracy studies. In: Deeks JJ, Bossuyt PMM, Leeflant MMG, et al., editors. Cochrane handbook for systematic reviews of diagnostic test accuracy. Draft version for Version 2 (Oct 2022); 2022. Online. Available: https://methods.cochrane.org/sdt/handbook-dta-reviews.
2. Rutjes A, Reitsma J, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006;1744:469–76.
3. Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332(7549):1089–92. doi: 10.1136/bmj.332.7549.1089. Erratum in: BMJ 2006;332(7554):1368.
4. Leeflang MM, Davenport C, Bossuyt PM. Chapter 5: Defining the review question. In: Deeks JJ, Bossuyt PM, Leeflang MM, et al., editors. Cochrane handbook for systematic reviews of diagnostic test accuracy. Draft version for Version 2 (Oct 2022); 2022. Online. Available: https://methods.cochrane.org/sdt/handbook-dta-reviews.
5. Doust JA, Bell KJL, Leeflang MMG, et al. Guidance for the design and reporting of studies evaluating the clinical performance of tests for present or past SARS-CoV-2 infection. BMJ 2021;372:n568. doi: 10.1136/bmj.n568.
6. Montalvo Villalba MC, Sosa Glaria E, Rodriguez Lay L, et al. Performance evaluation of Elecsys SARS-CoV-2 Antigen immunoassay for diagnostic of COVID-19. J Med Virol 2022;94:1001–8. doi:10.1002/jmv.27412.
7. Peat J, Barton B, Elliott E. Statistics workbook for evidence-based healthcare. Chichester, UK: Wiley-Blackwell; 2008.
8. Pewsner D, Battaglia M, Minder C, et al. Ruling a diagnosis in or out with 'SpPIn' and 'SnNOut': a note of caution. BMJ 2004;329:209–13.
9. Deeks J. Using evaluations of diagnostic tests: understanding their limitations and making the most of available evidence. Ann Oncol 1999;10:761–8.
10. Byrt T, Bishop J, Carlin J. Bias, prevalence and kappa. J Clin Epidemiol 1993;46:423–9.
11. Glasziou P, Irwig L, Mant D. Monitoring in chronic disease: a rational approach. BMJ 2005;330(7492):644–8.

12.Bossuyt PM, Reitsma JB, Linnet K, et al. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem 2012;58(12):1636–43.

13. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, et al. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012;344:e686.
14. Craig J, Williams G, Jones M, et al. The accuracy of clinical symptoms and signs for the diagnosis of serious bacterial infection in young febrile children: prospective cohort study of 15,781 febrile illnesses. BMJ 2010;340:c1594.
15. Stiell I, Greenberg G, McKnight R, et al. Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. JAMA 1993;269:1127–30.
16. McGinn T. Putting meaning into meaningful use: a roadmap to successful integration of evidence at the point of care. JMIF Med Inform 2016;4(2):e16.
17. Stiell I, Wells G. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med. 1999;33(4):437–47.