Chapter 3: Evaluating diagnostic tests

Jonathan J Deeks, Patrick MM Bossuyt

Key Points:

• Medical tests are procedures to assess an individual’s current health state or predict their future health state.

• Commonly used tests include clinical history taking and physical examination, questionnaire-based measurement tools, physiological measurements, in vitro tests of many different forms, radiological imaging, endoscopic and other optical examinations, and risk scores that combine test evidence of different types.

• Medical tests that provide evidence for diagnosis, staging, screening, monitoring, and surveillance of disease are often evaluated using test accuracy studies. However, assessing the ability to determine disease predisposition, prognosis, or treatment response requires longitudinal rather than test accuracy studies.

• Improving test accuracy can improve patient outcomes if more accurate tests lead to more effective interventions being given to the right patients. Evaluating test accuracy is necessary but not sufficient to assess whether tests benefit patients, as tests may also affect patients in other ways.

3.1 Introduction

In this chapter we consider various sorts of medical tests, and briefly introduce the types of test technology most commonly encountered in medical care and the research literature. We then review the multiple purposes for which medical tests are used and identify those that can be evaluated using the test accuracy paradigm and that therefore can be included in Cochrane Reviews of Diagnostic Test Accuracy. For several of these purposes we explain the need for evidence in addition to test accuracy to fully understand a test’s impact on
patients. We describe the types of research that may be undertaken when evaluating medical tests, and identify the place that test accuracy has in the evaluation pathway. Finally, we review the impact of tests on patients, consider the role that test accuracy has in predicting the effect of testing and treatment on patients, and the additional evidence that may be required to decide whether using a test is likely to do more good than harm.

3.2 Types of medical tests

A medical test refers to any procedure performed on a person’s fluids, cells, tissue, or on the person themself, to detect, diagnose or monitor a condition or the course of a condition. Medical tests come in many different forms, from patient history and physical and visual examination to lab tests and imaging, as well as risk scores that combine multiple pieces of information from different sources.

Diagnostic tests are medical tests undertaken in patients who present to health services with signs or symptoms. They are typically used to identify the likely cause of these symptoms, by identifying an underlying disease, and to decide whether treatment is required.

Cochrane places no restrictions on types of tests for inclusion as review topics, and defines testing as anything that records a characteristic, measurement, or observation from an individual, regardless of the technology used to obtain it. The defining feature is that the observation, resulting from the test or procedure, is made to infer something about the health state of the individual at that point in time. Table 3.2.a lists the types of tests and technologies most commonly used for diagnosis with examples from Cochrane Reviews.
### Table 3.2.a Major types of tests

<table>
<thead>
<tr>
<th>Test type</th>
<th>Example Cochrane Review of Diagnostic Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taking a medical history</td>
<td>First rank symptoms of schizophrenia</td>
</tr>
<tr>
<td>Clinical examination</td>
<td>Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within a general practice (primary care) setting</td>
</tr>
<tr>
<td>Physiological measurements</td>
<td>Ankle brachial index for the diagnosis of symptomatic peripheral arterial disease</td>
</tr>
<tr>
<td>In vitro diagnostics</td>
<td>Procalcitonin, C-reactive protein, and erythrocyte sedimentation rate for the diagnosis of acute pyelonephritis in children</td>
</tr>
<tr>
<td>Radiological imaging</td>
<td>Magnetic resonance imaging versus computed tomography to detect acute vascular lesions in patients presenting with stroke symptoms</td>
</tr>
<tr>
<td>Optical imaging</td>
<td>Capsule endoscopy for the diagnosis of oesophageal varices in people with chronic liver disease or portal vein thrombosis</td>
</tr>
<tr>
<td>Risk scores</td>
<td>Second trimester serum tests for Down syndrome screening</td>
</tr>
</tbody>
</table>

### 3.3 Test accuracy

This *Handbook* provides guidance for performing systematic reviews of test accuracy studies. Test accuracy studies evaluate the ability of a test to correctly classify study participants, based on the test results, as having the target condition or not. In most cases, that information will be relevant when the test is considered for diagnostic reasons: to identify the condition that most likely is responsible for the signs and symptoms of the patient. Chapter 4 describes the study designs for such studies.

Because the purpose of testing may not be strictly the detection of disease, we are, more generally, referring to the detection of a condition. The target condition is the condition that the test aims to detect. In test evaluation, the target condition may be a specific disease, such as chronic kidney disease. It may also be a syndrome (e.g. Down syndrome, acute coronary syndrome), or a disease stage (e.g. end-stage kidney disease, also known as renal failure).

The results from a medical test may be qualitative (expressed in a report) or quantitative (expressed in numbers). Essential for evaluations of test accuracy is that the test result can
be classified as pointing to the target condition, or not. The first situation – test result points to the presence of the target condition – is usually called a positive test result; the second situation – test result does not point to the presence of the target condition – is correspondingly called a negative test result. This means that all test results are classified, as either positive test results or negative test results.

In evaluations of the accuracy of medical tests, this classification into positive and negative test results is compared with a similar classification, but made by using the findings from the reference standard. The reference standard is usually the best available method for finding out whether people have the target condition (see Chapter 4). When we evaluate test accuracy, we compare two classifications: the one made based on the test results (positive versus negative) and the one that is made based on the reference standard (target condition present or not). Test positives – those with a positive test result – can then be found out to have a true positive test result (the reference standard confirms the presence of the target condition) or a false positive test result (the reference standard shows that the target condition is not present. Similarly, test negatives can be further classified as true negatives or false negatives. Test accuracy statistics, such as sensitivity and specificity, are statistical summaries of such comparisons. Chapter 5 discusses the various ways in which the results of this comparison of classifications can be summarized.

When the test is aimed to assist in making a diagnosis, the accuracy of that test can be referred to as its diagnostic accuracy. The question about the diagnostic accuracy of a test is a cross-sectional question: it addresses the extent to which a classification, based on the test results, corresponds to a classification of the same people, had they been evaluated by the reference standard, at that specific point in time.

Test accuracy is sometimes referred to in other terms. It is also known as the discriminatory performance or discrimination. Accuracy is but one of the possible ways to evaluate the clinical performance of a medical test. Clinical performance is a more general term, it refers to the ability of a test to yield results that are correlated with a particular clinical condition or a physiological or pathological process or state (Horvath 2014, EU 2017).

### 3.4 How do diagnostic tests affect patient outcomes?

It is now recognized that recommendations about the use of medical tests should be based on the ability of tests or testing strategies to improve outcomes: to improve or preserve the health of those being tested, or to make health care more efficient without detrimental health effects (Lord 2006, Schünemann 2008).

The impact of a test on patient outcomes is best assessed in randomized trials that allocate patients between diagnostic pathways that do and do not contain the new test and compare patient outcomes once all consequent interventions have been undertaken (Bossuyt 2000). Systematic reviews of these studies are included in the Cochrane Library as reviews of interventions, not of test accuracy. An alternative approach is a decision analysis that compares different diagnostic strategies using different sources of information for test accuracy and treatment efficacy, presuming particular models of clinician diagnostic and management behaviour (National Institute for Health and Clinical Excellence 2011, Chang 2012).
Randomized trials of testing strategies have advantages since they can also assess how other differences in the testing pathway affect patient outcomes (as described in Section 3.4.2). However, they are challenging to mount, needing large sample sizes, and requiring clinicians to adhere to diagnostic and treatment protocols (or at least to record them) (Ferrante di Ruffano 2017a).

Whilst these randomized trials are often proposed as the ideal method for judging the value of a diagnostic test, few such trials have been undertaken in diagnostic settings (Ferrante di Ruffano 2012a), and many have been underpowered or methodologically lacking (Ferrante di Ruffano 2017b). In some fields, such as infectious diseases, trials of tests are used (Odaga 2014). More success has been achieved in undertaking randomized trials of screening interventions, several of which have been evaluated in Cochrane reviews of interventions (e.g. screening for prostate cancer, colorectal cancer, and lung cancer) (Hewitson 2007, Gøtzsche 2013, Ilic 2013, Manser 2013).

Evidence of improved test accuracy may not necessarily be sufficient for a test to lead to improved patient outcomes (Lord 2006). In fact, as there are other ways in which tests affect patients which can lead to changes in outcomes, it is not even necessary for a new test to be as accurate as an existing test to benefit patients overall. It is important to be aware of the possible ways in which each new test can act, and to assess the evidence for each effect (Ferrante di Ruffano 2012b).

Cochrane Reviews of Diagnostic Test Accuracy can provide evidence of the accuracy of a test, but often need to be used alongside other evidence to make a holistic judgement of the likely impact on patients. There may even be some circumstances where trade-offs between harms (e.g. reduced accuracy) over benefits (e.g. less use of risky invasive testing) may need to be assessed. This section outlines the main issues that should be considered. These considerations should be mentioned in both the Background and Discussion sections of a Cochrane Review.

Ferrante di Ruffano and colleagues (Ferrante di Ruffano 2012b) proposed using a checklist to compare testing strategies to identify the key differences that might affect patient outcomes. The key aspects in this framework are described below.

### 3.4.1 Direct test effects
Tests may directly harm patients if they are invasive. It is important to be aware of what these harms may be and how frequently they are encountered. Whilst some tests may involve passing discomfort, others may have long-term sequelae, or put patients at risk of serious and sometimes life-threatening complications. Undergoing testing may also have psychological harms and benefits. Regardless of the test results, testing may cause anxiety through raising the possibility of discovering life-changing health states but may also provide reassurance by reducing the uncertainty afforded by a diagnosis.

### 3.4.2 Altering clinical decisions and actions
More accurate tests will improve patient outcomes if the reductions in false positive or false negative results lead to more people receiving appropriate diagnoses and appropriate treatment. The degree to which appropriate treatment can improve patient outcomes depends on its efficacy. However, it is also important to be wary of the harms associated with false positive and false negative diagnoses. People who receive false positive
diagnoses may undergo additional unnecessary testing and interventions, which may have adverse effects. They may also suffer unnecessary anxiety. Those who get false negative results from testing may be denied, or experience delays in receiving, effective treatment.

Although diagnostic yield generally increases with accuracy, it is also affected by a doctor’s confidence in the diagnostic test. Tests inducing greater confidence could benefit patients by reducing the need for further investigations and shortening the time to treatment. Doctors’ confidence can also influence the actual treatment delivered, particularly in surgery.

Tests will not affect clinical decisions if they are prone to technical failure or if they deliver results that are difficult to interpret. Either problem could require additional investigations, increasing the time to diagnosis, incorrect decision making, or poor diagnostic confidence. It is thus important to assess failure rates and non-diagnostic findings, both of which are commonly reported in test accuracy studies.

3.4.3 Changes to time frames and populations
In some clinical problems the speed with which the correct treatment is identified and given can be critical to a patient’s disease course. Thus, tests that can be undertaken earlier or produce results more quickly can improve health outcomes. However, quicker results are beneficial only if they produce earlier diagnosis and treatment, which may depend on factors other than the speed of the test.

Earlier diagnosis can also provide psychological benefit by dispelling anxiety or providing earlier reassurance but can also cause psychological harm, particularly if effective treatments are unavailable.

The value of a test within a health system will also increase with the number of patients who are able to access it. This may depend on the mode by which it is delivered which may improve access, but also on any contraindications to its use which may exclude particular patient groups.

3.4.4 Influencing patient and clinician perceptions
Predicting and measuring patients’ and clinicians’ perceptions of testing and their impact on health outcomes may be difficult, and many of these aspects may vary between patients, clinicians, and health systems, and have the potential to be modifiable.

Patients’ perceptions of testing, their experience of the testing process, and their understanding of the test result can all affect their health. Many studies show social, emotional, cognitive, and behavioural effects of testing across various clinical conditions. Evaluation of the acceptability of a test is important to assess whether patients are willing to undergo a procedure. This is especially important if multiple testing is required; an unpleasant first test can adversely influence patients’ willingness to attend follow-up testing or treatment. The experience of undergoing tests can also influence illness beliefs. Diagnostic placebo effects might occur if the impression of a thorough investigation improves perceptions of health status. Receiving a diagnosis can have behavioural and health consequences – for example, by confirming patients’ negative health beliefs.

Patients’ experiences and perceptions of testing will also affect subsequent health behaviours, such as the willingness or motivation to adhere to medical advice. Negative
perceptions or experiences of testing and clinical diagnosis could cause patients to lose confidence in the diagnosis or management plan, making them reluctant to have subsequent testing or treatment.

Doctors’ emotional, cognitive, social, or behavioural perspectives, although external to objective medical concerns, are nevertheless important in decision making. Referring doctors might modify management to reassure and satisfy patients or to prevent perceived threats of malpractice, often by requesting additional diagnostic information. This defensive medicine tends to raise the diagnostic threshold needed to trigger a change in management, and if additional tests are less accurate, harmful, or lead to treatment delays, patients will be adversely affected.

3.5 Evaluations of test accuracy during test development

Evaluations of the accuracy of a medical test or procedure can be performed during various stages of the development of a medical test. To appreciate the validity and applicability of these evaluations in Cochrane Reviews, we should understand the differences.

3.5.1 Evaluations of accuracy during biomarker discovery

Many medical tests are assays that measure a specific biomarker. A biomarker is a characteristic that serves as an indicator of normal biological or pathogenic processes, or pharmacologic responses to a therapeutic intervention (FDA-NIH Biomarker Working Group 2016). Assays on the other hand are analytic procedures for detecting or measuring the presence, amount, state, or functional activity of a biomarker (Horvath 2014). In vitro medical tests often use laboratory assays of one or more biomarkers, in a specific clinical context and for a specific clinical purpose, in a specific patient population.

Biomarkers may be discovered in different ways, including scientific reasoning and experimentation, and targeted technological development. In most circumstances no formal medical test exists at this point. However, some forms of modern biomarker development generate data on the ability of biomarker measurement to correctly identify patients with a known condition, and express the findings in accuracy terms, for example as the sensitivity and specificity of the biomarker. Very often, several putative biomarkers are evaluated in parallel in marker discovery studies.

These evaluations during biomarker discovery are regularly done in very specific groups, not representative of any clinical population, and often without a specific intended use. It is therefore not always known whether the biomarker will be used for a screening test or for a diagnostic test, for example. As a consequence, such evaluations are not very informative for actual decision-making in clinical practice. They are primarily done to evaluate the hypothesis that a biomarker holds promise, and biomarkers with encouraging findings may eventually lead to useful medical tests.

3.5.2 Early evaluations of test accuracy

Assays, particularly in vitro diagnostics, undergo studies of analytical performance before their accuracy is evaluated. Their ability to detect conditions in artificial laboratory conditions is investigated and their measurement properties established. Occasionally such studies are described as studies of analytical accuracy, and report estimates of
analytical sensitivity (the assay’s ability to detect very low concentrations of a given substance in a biological specimen) and analytical specificity (the assay’s ability to detect the intended target, without concerns about cross-reactivity or interference). Such studies are not appropriate to be considered in a systematic review of test accuracy, as they evaluate measurement properties, not the ability of a test to correctly classify patients in clinical practice.

Early studies of accuracy may be proof of concept studies that challenge the biomarker to discriminate between people with established disease and those who are healthy. Should a biomarker fail to meet this artificial but easy challenge it will not be evaluated further.

Further challenges may then be set to investigate whether the biomarker can discriminate between those with the target condition and others, with similar conditions, and can detect the condition at different points in its natural history. Such studies may be undertaken using easily available samples stored in biobanks. Often the health status of the study participants will already be known, and reference standards need not be undertaken.

As discussed in Chapter 4 and Chapter 8, the estimates of sensitivity and specificity from these studies are unlikely to be representative of the performance of the test in practice, but these studies will identify whether the test has the potential to be clinically useful.

3.5.3 Clinical evaluations of test accuracy
The test accuracy studies most appropriate for inclusion in Cochrane Reviews are those that are undertaken in the patient group where the test would normally be implemented, for the intended use, and at the point in the clinical pathway where it would actually be used. Such studies require evaluation of a reference standard diagnosis in all patients. These studies are most likely to provide estimates of test accuracy that will be applicable to patient care, as discussed in Chapter 4 and Chapter 5.

A further step is to undertake test accuracy studies that allow the accuracy of the new test to be compared with the accuracy of alternative tests, particularly against current testing practice. Comparative test accuracy studies will be useful in assessing whether new tests improve the accuracy of diagnoses and they provide the strongest type of evidence for selecting tests based on test accuracy. The Cochrane process pays particular attention to studies of this nature as they have the greatest ability to inform management decisions.

3.6 Other purposes of medical testing

Test accuracy studies can also be performed when the test is used for other purposes than arriving at a diagnosis. This section provides a more extensive overview of the reasons to consider medical testing, and to what extent test accuracy can be used to express the clinical performance of the test. Depending on the circumstances, the same procedure can be used as a medical test for more than one purpose.

Whilst a single application of a test may provide information for multiple purposes for an individual patient, the evaluation of the ability of a test to perform each of the roles is likely to need a separate research study, as explained in Table 3.6.a.
### Table 3.6.a Purposes of medical testing

<table>
<thead>
<tr>
<th>Test use</th>
<th>Clinical question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predisposition</td>
<td>Is this healthy person at risk of developing the condition in the future?</td>
</tr>
<tr>
<td>Risk stratification</td>
<td>How likely is this person to develop a condition in the future?</td>
</tr>
<tr>
<td>Screening</td>
<td>Does this asymptomatic person have early disease or a precursor of disease?</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>What condition is causing the symptoms in this person?</td>
</tr>
<tr>
<td>Staging</td>
<td>How severe is the condition in this person?</td>
</tr>
<tr>
<td>Prognosis</td>
<td>What is likely to be the outcome of the condition in this person?</td>
</tr>
<tr>
<td>Treatment selection</td>
<td>Is this person likely to benefit from the treatment considered?</td>
</tr>
<tr>
<td>Treatment monitoring</td>
<td>Is the condition changing under treatment?</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Is this chronic condition under control?</td>
</tr>
<tr>
<td>Surveillance</td>
<td>Has the condition progressed or recurred?</td>
</tr>
</tbody>
</table>

### 3.6.1 Predisposition
Tests of predisposition are undertaken in healthy individuals to identify those at increased risk of developing a particular disease. Knowledge of being at high risk may prompt individuals to modify health-related behaviours, to undergo enhanced surveillance for developing disease, or to undergo prophylactic interventions.

Tests for predisposition are based on how a person’s fixed characteristics, such as their family or medical history or genetic make-up, relate to the incidence of a disease. Such tests require information from longitudinal studies that compare disease incidence in people with and without each characteristic. Studies may be done either prospectively or retrospectively, but due to their longitudinal nature they are not suited for evaluation within the test accuracy framework.

### 3.6.2 Risk stratification
Tests can be used to calculate the chances of developing an event or a condition in the future, based on a combination of fixed and variable characteristics. Well-known examples are the cardiovascular risk scores, such as QRISK3, which are based on sex, age, total cholesterol, systolic blood pressure, smoking status, and other risk factors.

Like tests for predisposition, the development and validation of tools for risk stratification require longitudinal cohort studies and, as such, they are not suited for evaluation within the test accuracy framework.

### 3.6.3 Screening
Screening involves looking for early-stage disease, precursors of disease, or emerging risk factors in people who are asymptomatic and apparently healthy. It aims to identify disease at an early stage when it can successfully be treated, rather than to wait until it becomes symptomatic and less amenable to treatment.
It may be desirable to evaluate the accuracy of a screening test in detecting the early stage disease, by comparing the screening tests results against the reference standard. This would fit the test accuracy framework, even though screening tests are performed in asymptomatic people, in contrast with diagnostic tests. Systematic reviews of the test accuracy of screening tests may be undertaken. Systematic reviews of test accuracy may also be useful to compare the accuracy of a new screening test with an existing one.

However, assessing whether a screening programme creates more benefit than harm requires consideration of the outcomes of the consequent interventions, which is best undertaken in randomized trials. Cochrane Reviews of these trials are published as reviews of interventions and not of test accuracy.

3.6.4 Staging
Tests may be used to stage disease to find out how advanced or severe it is. Staging tests are undertaken only in patients found to have the disease, and may help identify the most appropriate choice of intervention depending on the severity of the disease. Staging thus classifies patients with a known condition into subcategories.

Evaluating the accuracy of a staging test is similar to evaluating a diagnostic test, in that results of the new test are compared with results of a reference standard. However, staging tests often classify patients into more than two categories, which make the application of accuracy statistics such as sensitivity and specificity difficult, as these are designed for binary classifications of data.

Staging tests also often function as tests of prognosis, in that patients with severe disease often have poorer outcomes. However, longitudinal studies are required to evaluate the relationship between disease stage and future likely outcomes (the prognostic value of the test) and to assess the accuracy of a staging test.

3.6.5 Prognosis
Prognostic tests are used to predict what the future state of a disease is likely to be. Knowledge of likely future health states may lead to modification of a patient’s treatment, or enable the patient to make appropriate plans.

Tests for prognosis are similar to those for predisposition, in that they aim to predict what the future holds for a patient, and their evaluation requires longitudinal follow-up. Unlike predisposition studies, prognosis studies examine the relationship between test results and a health outcome, such as death or recovery, rather than the onset of disease. Prognosis may be investigated in a sample of all those who have the disease, while predisposition is investigated in healthy people who do not yet have the disease. Prognosis studies may be done either prospectively or retrospectively, but due to their longitudinal nature they are typically not suited for evaluation within the test accuracy framework. However, some clinical scenarios where outcomes are observed over a short standard time-period, such as predicting the outcome of pregnancy, have strong commonalities with test accuracy studies.

3.6.6 Treatment selection
Stratified or precision medicine focuses on identifying subgroups of patients who will benefit from particular interventions, often through pairing a new test or biomarker with an intervention. The greatest advances are being made where the molecular understanding of
a disease leads to development of both a biomarker to detect the molecular target and a targeted therapy. Typically, this involves the use of molecular pathology tests such as genomics, proteomics, and metabolomics, which have had greatest successes in cancer.

Benefit from treatment is defined as a difference in outcomes: a better outcome after treatment compared with the counterfactual outcome when not being treated. Such a difference in outcome can never be directly observed, and a reference standard for benefit may be difficult to define. Hence, the evaluation of treatment selection tests rarely fits the test accuracy paradigm.

Evaluation of test-treatment combinations is best achieved through randomized trials. Ideally randomized trials randomly allocate participants to treatment or a comparator, and investigate whether the observed treatment effect differs between patients who are biomarker positive and those who are biomarker negative, via a test of the interaction. Such evaluations are more suited to the format of Cochrane Reviews of interventions than those of test accuracy.

3.6.7 Treatment efficacy
Medical tests may be used in both clinical practice and research to predict whether a treatment has worked or not, particularly to get an early indication of a response to a new treatment when long-term follow-up would otherwise be required. In oncology, such tests are called predictive tests, to distinguish them from prognostic tests, which would predict the future health state under no or under conventional treatment.

Such trials often evaluate the presence or absence of a response after treatment in studies in which all participants receive treatment and compare the proportion of responders between biomarker positives and biomarker negatives. Though the results can be expressed in terms of accuracy, the longitudinal nature of such treatment studies makes them unfit for the test accuracy paradigm.

3.6.8 Therapeutic monitoring
Therapeutic monitoring involves using a test repeatedly in a group of patients to see whether their disease is being appropriately controlled by a therapy. For example, patients who are prescribed anti-hypertensive drugs will have their blood pressure regularly measured to assess whether the drug and its dosage are appropriate. If their blood pressure remains high, the dose may be increased, or further anti-hypertensive agents may be added. If their blood pressure is too low, the drug dosage may be reduced.

To inform treatment decisions reliably, monitoring tests need to have good test accuracy, to correctly classify individuals. However, monitoring tests must also respond to changes in patients’ health state quickly, and their clinical value will also depend on how often they are used, and whether the associated changes in management actually improve patient health. Evaluation of whether using a monitoring strategy leads to patient benefit is also best assessed using a randomized trial, as for screening interventions, and is amenable to systematic review using the Cochrane methods for reviewing intervention studies.

3.6.9 Surveillance for progression or recurrence
Tests are also used repeatedly in patients who have a progressive disease or have recently undergone an intervention to check for disease progression or recurrence, which may
prompt changes in patient management. This is sometimes called surveillance, a form of monitoring.

Surveillance has many similarities to screening, but screening is performed in asymptomatic people while surveillance is planned in individuals known to have a disease or known to have undergone an intervention.

Surveillance, like monitoring, requires tests that are accurate but also respond to changes in patient status. Thus, systematic reviews of the accuracy of tests for surveillance are informative, but whether or not surveillance testing benefits patients is also best assessed in randomized trials.

3.7 Chapter information

Authors: Jonathan J Deeks (Institute of Applied Health Research, University of Birmingham, UK); Patrick MM Bossuyt (Department of Epidemiology and Data Science, Amsterdam University Medical Centers, University of Amsterdam, The Netherlands).

Sources of support: The authors declare no external sources of support for writing this chapter.

Declarations of interest: Jonathan Deeks is a member of Cochrane’s DTA Editorial Team. The authors declare no other potential conflicts of interest relevant to the topic of their chapter.

Acknowledgements: The authors would like to thank Jenny Doust and Matthew McInnes for helpful peer review comments.

3.8 References


Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of
diagnostic tests on patient outcomes are rare. *Journal of Clinical Epidemiology* 2012a; 65: 282-287.


Ferrante di Ruffano L, Dinnes J, Sitch AJ, Hyde C, Deeks JJ. Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests. *BMC Medical Research Methodology* 2017b; 17: 35-35.


Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane Database of Systematic Reviews* 2013: CD004720.


