



*National Institute for
Health and Clinical Excellence*

Issue date: [December 2011](#)

Diagnostics Assessment Programme manual

National Institute for Health and Clinical Excellence

Level 1A

City Tower

Piccadilly Plaza

Manchester M1 4BT

© National Institute for Health and Clinical Excellence, 2011. All rights reserved. This material may be freely reproduced for educational and not-for-profit purposes. No reproduction by or for commercial organisations, or for commercial purposes, is allowed without the express written permission of NICE.

Contents

PART I: INTRODUCTION TO THE PROGRAMME AND TO DIAGNOSTIC TECHNOLOGIES.....	5
List of abbreviations	6
Foreword.....	7
1. Introduction to Diagnostics Assessment Programme	8
1.1 General description of NICE and the Diagnostics Assessment Programme	8
1.2 Aims of the Programme.....	10
1.3 Key activities of the Programme.....	10
1.4 Key audiences.....	10
1.5 Participants in the Diagnostics Assessment Programme process.....	11
1.6 Information disclosure	15
1.7 Equality considerations	19
2. Characteristics of diagnostic tests.....	21
2.1 Introduction	21
2.2 Types of diagnostic tests.....	22
2.3 Uses of diagnostic tests	24
2.4 Outcome measures.....	27
PART II: PROGRAMME PROCESSES	29
3. Selection of diagnostic technologies	30
4. Initiation of the evaluation.....	32
4.1 Initiation date of an evaluation.....	32
4.2 Activities undertaken when an evaluation is initiated	32
5. Developing the scope.....	35
5.1 Understanding the care pathway.....	35
5.2 Contents of the scope	36
5.3 Contributors to the development of the draft scope.....	37
5.4 The scoping workshop	38
5.5 The assessment subgroup	39
5.6 Stopping evaluations at scoping stage	40
5.7 Scoping steps and timings	40
6. Assessing clinical outcomes and the cost effectiveness of diagnostic technologies	42
6.1 Impact of the scope on the assessment structure	42
6.2 Assessment methods used by the External Assessment Group	42
6.3 Diagnostics assessment report	43
6.4 Assessment steps and timings.....	44
7. Evaluation by the Diagnostics Advisory Committee	46
7.1 Consideration of evidence and development of draft recommendations.....	46
7.2 Development of and consultation on the diagnostics consultation document	49
7.3 Review of the diagnostics consultation document in light of consultation comments.....	50
7.4 Preparation of the diagnostics guidance document.....	51
8. Resolution and publication of guidance.....	52
8.1 Resolution grounds	52
8.2 Eligibility to make a resolution request	53

8.3	Resolution requests	53
8.4	Initial scrutiny of resolution requests	53
8.5	The resolution panel.....	54
8.6	Publication of diagnostics guidance	56
8.7	Steps and timings for evaluation, resolution and publication.....	57
9.	Guidance reviews.....	59
10.	Updating this manual	60
PART III: METHODS USED FOR DECISION-MAKING		61
11.	Introduction to Programme methods	62
12.	Developing the scope.....	63
12.1	Patient population	63
12.2	Intervention (technology or test) to be evaluated and comparators...	64
12.3	Care pathway	66
12.4	Outcomes and costs	67
12.5	Other considerations	69
13.	Evidence for assessment and evaluation	70
13.1	Introduction	70
13.2	Types of evidence	71
13.3	Identifying and synthesising evidence on diagnostic test accuracy...	72
13.4	Identifying and synthesising evidence for health outcomes including test side effects	80
13.5	Identifying evidence for cost effectiveness	81
14.	Modelling clinical outcomes and the cost effectiveness of diagnostic technologies	84
14.1	Simplified analyses	84
14.2	Structuring the assessment.....	85
14.3	Considerations for assessment and modelling.....	87
14.4	Identifying future research needs from the evidence.....	89
15.	Cost-effectiveness evaluation – the reference case.....	91
15.1	The concept and structure of the reference case	91
15.2	Modelling methods	97
15.3	Characterisation of potential bias and uncertainty.....	98
15.4	Presenting data and results.....	100
15.5	Analysis of data for patient subgroups	102
15.6	Reflecting equity considerations in cost-effectiveness analysis	104
16.	Development of recommendations by the Diagnostics Advisory Committee.....	105
16.1	Committee review of the evidence	105
16.2	Developing recommendations.....	109
16.3	Types of recommendation.....	111
16.4	Framework for research recommendations.....	112
PART IV: APPENDICES AND REFERENCES		113
Appendix A: Glossary		114
Appendix B: Members of diagnostics methods working group.....		123
Appendix C: Process timeline		126
Appendix D: References		129

**PART I: INTRODUCTION TO THE
PROGRAMME AND TO
DIAGNOSTIC TECHNOLOGIES**

List of abbreviations

DAC	Diagnostics Advisory Committee
DAP	Diagnostics Assessment Programme
DAR	Diagnostics assessment report
DCD	Diagnostics consultation document
DGD	Diagnostics guidance document
EAG	External Assessment Group
HRQL	Health-related quality of life
MTAC	Medical Technologies Advisory Committee
NETSCC	NIHR Evaluation, Trials and Studies Coordinating Centre
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NIHR	National Institute for Health Research
PSS	Personal social services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial

Foreword

This programme manual describes how the NICE Diagnostics Assessment Programme develops guidance. The programme is designed to ensure that robust guidance is developed for the NHS in an open, transparent and timely way, allowing appropriate input from stakeholders.

The manual is in four sections:

- a. Introduction to the programme and to [diagnostic technologies](#)
- b. Programme processes
- c. Methods used for decision-making
- d. Appendices and references.

Nothing in this document will restrict any disclosure of information by NICE that is required by law (including, in particular but without limitation, the Freedom of Information Act 2000).

Terms in this document, indicated in [bold text](#) at their first mention, are listed in the glossary (appendix A).

1. Introduction to Diagnostics Assessment Programme

This section includes:

- A general description of NICE and the Diagnostics Assessment Programme (section 1.1)
- The aims and key activities of the Programme (sections 1.2 and 1.3)
- Key audiences for the Programme (section 1.4)
- Participants in the process (section 1.5)
- Information disclosure (section 1.6)
- Equality considerations (section 1.7).

1.1 *General description of NICE and the Diagnostics Assessment Programme*

The National Institute for Health and Clinical Excellence (NICE) provides guidance, sets quality standards and manages a national database to improve people's health and prevent and treat ill health. Further details about NICE and its work programmes are available from the [NICE website](#)

The Diagnostics Assessment Programme (DAP) is part of NICE's activities on evaluating [medical technologies](#). NICE has two programmes in which diagnostic technologies may be evaluated: the Medical Technologies Evaluation Programme and the Diagnostics Assessment Programme. The DAP is suitable for evaluating diagnostic tests and technologies where such evaluation is complex, for example, where recommendations can only be made on the basis of clinical utility and [cost-effectiveness analysis](#) or where meaningful assessment requires the consideration of multiple technologies or [indications](#). The DAP evaluates diagnostic technologies that have the potential to improve health outcomes but whose introduction is likely to be associated with an overall increase in cost to the NHS. Diagnostic technologies that may offer similar health outcomes at less cost, or improved health outcomes at the same cost as current NHS practice, are likely to be more suitable for evaluation by the Medical Technologies Evaluation Programme.

Both Programmes evaluate diagnostic technologies as defined in EU directives 93/42/EEC (concerning medical devices), 98/79/EC (concerning in vitro diagnostic medical devices) and 90/385/EEC (concerning active implantable medical devices), as amended. Genetic tests are covered by the Programmes provided they have a medical purpose and fall within the scope of EU directive 98/79/EC.

There are various types of diagnostic tests and technologies, and DAP concentrates on pathological tests, imaging, endoscopy, algorithms or test

1 - Introduction

combinations, and physiological measurement, because these represent most of the investigations performed on patients. It does not include tests based on clinical sign detection (as part of a 'bedside' clinical examination not involving use of instruments or devices).

Diagnostic technologies may be used for various purposes: diagnosis, clinical monitoring, screening, treatment triage, assessing stages of disease progression, risk stratification, etc. All of these uses of diagnostic technologies fall within the remit of the Programme. Unless specifically stated otherwise, the use in this document of the term technology or technologies should be interpreted as diagnostic technologies.

A [companion diagnostic technology](#), where the primary purpose of the technology is to identify patients who respond best to new drugs, may be suitable for [evaluation](#) in the NICE Technology Appraisal Programme in the context of an appraisal of the drug to which it is linked. In other cases, companion diagnostic technologies may be suitable for evaluation in the DAP. This may include, for example, new companion diagnostics for established drugs.

The Programme evaluates diagnostics that are intended for use within the NHS in England and are paid for by the NHS with public funds, either in part or in whole.

Assessing population screening programmes and making recommendations on their introduction, modification or withdrawal is undertaken by the [UK National Screening Committee](#) and is beyond the scope of our Programme. The NICE DAP evaluates screening tests that are applied to patients who are already suspected of having a disease.

1.1.1 Differences between the Diagnostics Assessment Programme and the NICE Technology Appraisal Programme

The DAP evaluates diagnostic technologies using cost-effectiveness analysis but it differs in various significant ways from the NICE Technology Appraisal Programme, which also uses cost-effectiveness analysis. The differences are outlined below:

- Evidence about [patient outcomes](#) for diagnostic technologies is typically lower in quantity and quality than evidence for pharmaceutical products.
- Because most benefit to the patient arises from treatment based on the result of the diagnostic test, the value of the test or technology is best understood in the context of its effect on the pathway of care.
- Diagnostic technologies, particularly those based on electronics, often change rapidly as new methods, upgrades and capabilities are added.

1 - Introduction

- It is often not obvious where in the [care pathway](#) the [diagnostic technology](#) is best placed, so different options are evaluated.

1.2 Aims of the Programme

The aims of the Programme are:

- to promote the rapid and consistent adoption of innovative clinically and cost-effective diagnostic technologies in the NHS
- to improve treatment choice or the length and quality of life by evaluating diagnostic technologies that have the potential to improve key clinical decisions
- to improve the efficient use of NHS resources by evaluating diagnostic technologies that have the potential to improve systems and processes for the delivery of health and social care.

1.3 Key activities of the Programme

The key activities of the Programme are:

- undertaking evaluations of diagnostic technologies that require complex analysis because of the involvement of multiple technologies or indications; or because cost-effectiveness analysis is required for meaningful evaluation
- developing and publishing [diagnostics guidance](#) on selected diagnostic technologies for the NHS in England and its social care partners
- recommending research into the clinical utility and/or healthcare system benefits of diagnostic technologies
- reviewing and updating diagnostics guidance when required.

1.4 Key audiences

The DAP has several audiences that are expected to take note of NICE's diagnostics guidance:

- NHS commissioners – for example, when specifying services that incorporate use of diagnostic technologies
- practitioners, including clinicians, who use diagnostic technologies in clinical or research settings
- healthcare operational managers in primary and secondary care settings, particularly when planning services or facilities in which diagnostic technologies are used
- purchasing and procurement organisations, when planning procurement of diagnostic technologies.

Patients and carers of people who may be affected by the technologies are an important audience for the Programme because diagnostics guidance can help them make informed decisions about their treatment, in consultation with their clinicians.

1.5 *Participants in the Diagnostics Assessment Programme process*

Table 1 Participants in the Diagnostics Assessment Programme process

<p>Diagnosics Advisory Committee</p>	<p>The Diagnostics Advisory Committee (DAC or ‘the Committee’) is an independent Committee consisting of 22 standing members and additional specialist members.</p> <p>The role of the Committee is:</p> <ul style="list-style-type: none"> • to consider evidence • to make draft recommendations • to consider public consultation comments • to make final recommendations for publication in NICE guidance. <p>Standing Committee members have a range of expertise, and include clinicians who develop and use diagnostic technologies, people who can provide a lay perspective on the issues affecting patients and the NHS, experts in regulation and evaluation of healthcare technologies, people with commissioning experience in the NHS, and people with experience of the diagnostic technologies industry. Standing Committee members are recruited through an open advertisement posted on the NICE website. They are appointed for a period of up to 3 years by a panel including an Executive or Centre Director, a Non-Executive Director and the Chair of the Committee.</p> <p>Specialist Committee members are recruited for their expertise in the diagnostic technology under consideration and/or the care of patients in the pathway in which the results of the test are used. They are recruited for each topic; their term of office is for the duration of the topic (approximately 10 months) and their involvement is for that topic only. Numbers may vary from topic to topic. They typically include clinicians or researchers using the diagnostic technology or involved in the care pathway, as well as lay persons with a perspective on the condition being diagnosed. Specialist Committee members have the same decision-making status as standing members of the</p>
--------------------------------------	---

	<p>Committee. See section 4.2.2 for details on how specialist Committee members are appointed.</p> <p>NICE is committed to the values of equality and diversity and welcomes applications for membership of the Committee from all sectors of the community.</p>
<p>Registered stakeholders</p>	<p><u>Registered stakeholders:</u></p> <ul style="list-style-type: none"> • are invited to attend the scoping workshop • receive the diagnostics assessment report (DAR – see section 6.3) for comment (these comments are considered by the Committee when it formulates its draft recommendations on a topic) • have the same input to the development of diagnostics guidance as members of the public (see below). <p>Identifying potential registered stakeholders is an important part of the process. Registration is open to anyone with an interest in the topic who is (or belongs to) one of the following:</p> <ul style="list-style-type: none"> • a manufacturer, developer, distributor or agent of a relevant technology (see below) • a trade association representing manufacturers, developers, distributors or agents of diagnostic technologies • a national organisation representing healthcare professionals • a national group representing patients and/or carers • a provider of NHS services in England • a commissioner of NHS services in England • a statutory organisation such as the Department of Health • a research organisation. <p>Stakeholders register via NICE’s website. Potential stakeholders may register at any point in the evaluation process. More information about registering is available from: Diagnostic technologies stakeholder registration</p>
<p>Product sponsors</p>	<p>Manufacturers, developers, distributors or agents of:</p> <ul style="list-style-type: none"> • a technology selected for assessment via DAP by the Medical Technologies Advisory Committee (MTAC) (see section 3) or • a technology identified during the scoping period as a

	<p>possible related alternative technology (see section 5.2.1) to the technology selected by MTAC</p> <p>are referred to in this process as product sponsors, and are invited to join the evaluation.</p> <p>Product sponsors:</p> <ul style="list-style-type: none"> • are invited to register as stakeholders • are asked to provide data to support the evaluation of the technology or technologies, as outlined in section 4.2 • are invited to attend the Committee meetings to comment on matters of factual accuracy, and to respond to questions from the Committee about information submitted to inform the evaluation, including confidential information (see section 7 about Committee meetings). <p>When a topic is selected for the Diagnostics Assessment Programme, NICE informs the product sponsor of its intention to evaluate that technology.</p>
<p>Manufacturers, developers, distributors or agents of comparator technologies</p>	<p>Manufacturers, developers, distributors or agents of any technology identified during the scoping period as a comparator (see section 5.2.1) are able to register as a stakeholder (see above). Comparator technologies are those that are most commonly used or are recommended in current NICE guidance for the indications and uses that feature in the evaluation.</p>
<p>External Assessment Group (EAG)</p>	<p>The External Assessment Group (EAG) is an independent academic group that prepares a review of the clinical effectiveness and cost effectiveness of the technology or technologies under consideration. The DAR prepared by the EAG is based on a systematic review of the clinical and health economic literature (including data supplied by the product sponsor or sponsors when appropriate) and appropriate models. The EAG is commissioned to carry out this assessment on NICE's behalf by the National Institute for Health Research (NIHR) - Evaluation, Trials and Studies Coordinating Centre (NETSCC).</p> <p>The EAG is invited to the scoping workshop (see section 5.4), the assessment subgroup meeting(s) (see section 5.5) and Committee meetings. The EAG may also work with the NICE Diagnostics Assessment Programme team during the</p>

	early stages of scope development.
Members of the public	<p>Members of the public may:</p> <ul style="list-style-type: none"> • comment on the diagnostics consultation document (DCD) (see section 7 for more information) • apply to attend Committee meetings (see section 7 for more information) • apply to become a lay specialist Committee member (see section 4.2 for more information).
NICE staff	
Diagnostics Assessment Programme team	<p>The Diagnostics Assessment Programme (DAP) is part of NICE’s Centre for Health Technology Evaluation (CHTE). The DAP team consists of the Associate Director and technical, project and administrative staff who support the DAC in developing diagnostics guidance. Members of the DAP team:</p> <ul style="list-style-type: none"> • develop a detailed draft scope, including carrying out research on the care pathway • liaise with EAGs about evidence assessments • prepare evidence overviews for the Committee • arrange public consultation on the Committee's draft recommendations • prepare guidance for publication • ensure agreed timelines and quality assurance standards are followed • promote awareness of the Programme.
Patient and Public Involvement Programme (PPIP)	<p>The Patient and Public Involvement Programme (PPIP) recruits and supports lay members of the Committee (both standing members and specialist members), identifies appropriate patient and carer organisations to be invited to register as stakeholders, encourages members of the public and patient organisations to respond to consultations, and establishes links with patient organisations with an interest in diagnostics guidance. NICE uses the terms ‘patient organisation’ and ‘patient group’ when referring to patients, carers, and community and other lay organisations and charities, including those representing people from groups protected by equalities legislation.</p>
Information Services	<p>The Information Services team searches for information and evidence from conventional sources and ‘grey’ literature.</p>

	This information is primarily used by the DAP team to prepare the scopes.
Editorial	The editors review the DCD and the diagnostics guidance document (DGD). NICE editors prepare the final guidance for publication on the NICE website and also develop a lay explanation of the recommendations when appropriate.
Implementation	<p>NICE provides advice and tools to support the local implementation of its guidance. In general NICE's implementation team:</p> <ul style="list-style-type: none"> • ensures intelligent dissemination to the appropriate target audiences • actively engages with the NHS, local government and the wider community • works nationally to encourage a supportive environment • provides tools to support putting NICE guidance into practice • demonstrates significant cost impacts – either costs or savings at local and national levels • evaluates uptake of NICE guidance • shares learning • develops educational material to raise awareness of NICE guidance and encourages people to input into its development. <p>There is an implementation support plan for each piece of guidance. The implementation team produces implementation support tools (such as costing tools and audit tools) to help the NHS implement NICE guidance. These tools are developed with advice from Committee members and reference groups as appropriate.</p>

1.6 Information disclosure

To ensure that the diagnostics evaluation process is as transparent as possible, NICE ensures that wherever possible evidence pivotal to the DAC's decisions is publicly available. This section covers

- what information NICE discloses during an evaluation (section 1.6.1)
- what information NICE treats as confidential during an evaluation (section 1.6.2)

1 - Introduction

- registered stakeholder responsibilities in relation to confidential information (section 1.6.3)
- models including costs and clinical outcomes (section 1.6.4).

NICE adheres to the principles and requirements of the Data Protection Act (1998) and the Freedom of Information Act (2000) in dealing with information it receives during a diagnostics evaluation.

1.6.1 Information NICE discloses during an evaluation

Table 2 shows the documents that are made publicly available during the evaluation process. NICE posts these documents on its website. These documents are not considered confidential once they are posted on the website.

Table 2 Documents made publicly available during an evaluation

Documents NICE makes publicly available during the evaluation	For further information see section
Final scope for the evaluation	5
List of registered stakeholders	1.5
Assessment protocol	5.4, 5.5
Diagnostics assessment report (DAR) ^a	6.3
Comments from registered stakeholders on the DAR ^a	6.3
Evidence overview prepared by Programme team ^a	7.1
Diagnostics consultation document (DCD) ^b	7.2
Comments received on the DCD and responses from NICE ^a	7.2, 7.3
Diagnostics guidance document	7.4
^a These documents are made available to stakeholders earlier in the process than their publication on the website.	
^b 5 working days before publishing on its website, NICE releases this document to registered stakeholders.	

Any confidential information provided to NICE in the course of the assessment is made available for review by the EAG and the DAC.

Reference is made in the diagnostics assessment report (DAR) to the existence of documents that have been designated as confidential by the originator. Specific confidential information is redacted from the version of the

1 - Introduction

DAR that is released to registered stakeholders and published on the NICE website. Unless the status of confidential information changes during the period of the diagnostics evaluation, it will not appear in other documents, such as the draft guidance or the final guidance.

NICE reserves the right to use in the DAR, draft guidance and final guidance any material that is provided to it during the course of an evaluation that is not designated by the person providing it as being 'confidential', or ceases to be so during the evaluation.

NICE considers that evidence designated as 'academic in confidence' (but not 'commercial in confidence') can be presented at DAC meetings with members of the public and press present.

1.6.2 Information NICE treats as confidential during an evaluation

Unpublished evidence is accepted under agreement of confidentiality. However, NICE expects the product sponsor or sponsors to keep confidential material made available to the evaluation to an absolute minimum. Types of information that can be classed as confidential include:

- data that are 'commercial in confidence' (CIC)
- data that are intellectual property awaiting publication ('academic in confidence').

Information designated by a [sponsor](#) as either commercial or academic in confidence should be consistent with the following principles:

- Information and data that have been made publicly available anywhere in the world are not considered confidential.
- When it has been decided that study results will be published in a journal after the first public release by NICE of documentation quoting data from the study, as a minimum a structured abstract should be made available for public disclosure. The structured abstract should be a synopsis following a recognised format for a full trial report, such as that provided by the CONSORT (www.consort-statement.org) or STARD statements.

NICE asks data owners to reconsider restrictions on the release of data when either there appears to be no obvious reason for the restrictions, or such restrictions would make it difficult or impossible for NICE to show the evidence base for its guidance.

Guidance on how to identify confidential information when providing documents to the evaluation is available from the Programme team. In the case of technologies that require CE marking, NICE will not make public any draft guidance for public consultation before the technology is CE marked.

1 - Introduction

In the event of an unauthorised disclosure from a confidential document relating to a specific topic, NICE may comment on the guidance publicly before it is published. The decision to do so will be taken by the Chair or Vice Chair of NICE on the recommendation of two Executive Directors. Registered stakeholders will be informed of this decision as soon as possible after it has been taken.

1.6.3 Responsibilities of registered stakeholders in relation to confidential information

Organisations and individuals (including product sponsors and manufacturers) are required to sign a confidentiality agreement before they are recognised as registered stakeholders and evaluation documentation can be released to them. This includes the draft scope prior to the scoping workshop. The confidentiality form is available at

<http://www.nice.org.uk/guidance/dt/diagnostictechnologiesstakeholderregistration.jsp>.

Registered stakeholders must not disclose confidential documents relating to an evaluation until NICE makes the documents public. It is the responsibility of the registered stakeholders, and any other party that has signed a confidentiality agreement for the evaluation, to keep such documents, which are not otherwise publicly available, confidential and secure at all times. NICE considers individuals within a registered stakeholder organisation who see evaluation documentation to be bound by the terms of the confidentiality agreement signed by the organisation.

Any organisation or individual not in the direct employment of the registered stakeholder organisation is a third party. Registered stakeholders may release the evaluation documentation to third parties when this is clearly necessary to enable the registered stakeholder to formulate their contribution to the evaluation, and:

- the third party has seen and agreed to be bound by the terms of the confidentiality agreement, or
- the registered stakeholder is satisfied that the third party has signed and returned their confidentiality agreement to NICE.

1.6.4 Models including costs and clinical outcomes

For a diagnostics evaluation, models may be used for a number of purposes, including modelling of costs, of outcomes as derived through the care pathway and of cost effectiveness.

Models are produced by the EAG as part of its assessment.

1 - Introduction

The EAG normally produces one or more models as part of its assessment. Because the model outputs are used by the Committee to assist their decision-making, when possible NICE distributes an executable version to registered stakeholders. NICE does not make the model publicly available. Before the DAR is distributed prior to the first Committee meeting, registered stakeholders are asked if they wish to receive an electronic copy of the model, subject to agreeing to specified conditions for its use and disclosure. Registered stakeholders are clearly informed that NICE will distribute an executable copy, that the model is protected by intellectual property rights, and that the executable copy can be used only for the purposes of commenting on the model's reliability and informing a response to the DAR and/or draft guidance. The model is offered for consultation at the same time as the DAR.

NICE offers the EAG's model for consultation if it does not contain model inputs designated confidential by the data owner or provider. NICE does not request separate permission to do so.

If the EAG's model contains confidential model inputs that can be redacted without producing severe limitations on the functionality of the model, NICE offers the model for consultation in a redacted form.

Results derived from calculations incorporating confidential data are not considered confidential unless releasing those results would enable back-calculation to the original confidential data.

The only situation in which NICE would not offer the EAG's model for consultation is if it contains confidential model inputs that cannot be redacted without severely limiting the model's function.

1.7 Equality considerations

The DAP operates in accordance with the NICE equality scheme (available from www.nice.org.uk/aboutnice/howwework/niceequalityscheme.jsp). Equality considerations are taken into account at each stage of the development of diagnostics guidance, including scoping, and the development of the Committee's draft and final recommendations. The equality issues raised at each development stage for a topic are recorded in the equality impact assessment (in accordance with the documented equality impact assessment procedure). The equality impact assessment is approved by the Programme's Associate Director and published with the scope, and approved by the Programme or Centre Director and published with the guidance. Any equality issues that directly affect the recommendations in the guidance, or the Committee's consideration of the evidence, are included in the final guidance.

1 - Introduction

2. Characteristics of diagnostic tests

This section includes:

- Introduction (section 2.1)
- Types of diagnostic tests (section 2.2)
- Uses of diagnostic tests (section 2.3)
- Outcome measures (section 2.4).

2.1 Introduction

Diagnostics involves a variety of tests and measurements that can be used to determine what conditions, diseases or syndromes¹ a person may currently have or is likely to develop. These tests can be used in a variety of ways, including screening, suggesting diagnoses, ruling out or confirming suspected diagnoses, monitoring chronic conditions, monitoring a patient's condition following treatment, and predicting future events.

Some diagnostic technologies are used with concomitant treatment. For example, endoscopy can be used to not only detect lesions, but also to remove the lesion, either for further testing or because it forms part of the treatment. In these circumstances, diagnostic interventions can be curative and can avoid outcomes that would otherwise occur later in the care pathway.

Diagnostics are sometimes an integral part of treatment. For example, imaging may be used during radiation or surgical therapy for some conditions. Diagnostics may be specific adjuncts for certain treatments. In these cases, the diagnostic test may be evaluated independently by the Diagnostics Assessment Programme or may be evaluated in conjunction with the treatment by other NICE Programmes.

Regardless of the tests or the way in which they are used, the evaluation of diagnostics is both similar to and different from the evaluation of treatments. It is similar because both are interventions aimed at improving the quantity and quality of life of the patient. As with treatments there are often alternative interventions or series of interventions to compare with the intervention being evaluated. In both cases, overall costs are considered in the evaluation.

The evaluation of diagnostics differs from the evaluation of treatments in several ways. The most important difference is that diagnostic tests have few direct outcomes, that is, outcomes affecting the patient that come directly from the test itself. Most outcomes of interest follow from treatments that are either

¹ These terms are used interchangeably and inclusively in this document to refer to any disease, condition or syndrome being diagnosed.

2 – Characteristics of diagnostic tests

initiated or not initiated based on the results of the tests. The second important difference is that tests are frequently done in conjunction with other tests or measurements, and, where this is the case, it is the composite of the series of tests that is used in clinical decision-making.

These two important differences make the evaluation of diagnostics complex. Only very rarely do studies of diagnostic tests follow patients through treatment to final outcomes. Also, evaluation of diagnostics usually requires that the clinical management process is described and that the effects of that process are known or assumed. If the effects of treatment are not known, analyses can be performed, but the validity of the results will be less certain in ways that may not be completely specifiable. This increases the uncertainty with which decisions can be made on use of diagnostic technologies.

In statistics, '[test accuracy](#)' means the proportion of test results that are correct. This is not a useful definition for the purposes of this document, because a test may be incorrect in more than one way and for more than one reason. This form of accuracy is also dependent on the prevalence of the condition in the population tested. Therefore, when this document refers to test accuracy, it means any measure relating to the correctness of the test, not just the proportion of results which are correct. It includes [sensitivity](#), [specificity](#), predictive values, etc.

Most diagnostic tests are not perfectly accurate from a clinical perspective. Although a test may accurately measure the level of a particular chemical or produce an accurate image of a part of the body, normal levels vary from person to person according to factors such as age, gender, ethnicity or weight. This means that normal and abnormal ranges can overlap.

2.2 Types of diagnostic tests

NICE considers developing guidance on specific types of measurements and tests that are used to evaluate a patient's condition.

2.2.1 Physiological measurements

Physiological measurements include tests such as measurement of temperature and blood pressure, weight and height, eye examinations, and tympanometry. In the NHS the term is commonly used in relation to tests that assess the function of major organ systems.

2.2.2 Laboratory tests and pathology

These tests involve taking samples of body fluids or tissues and subjecting them to some form of analysis. The analysis may be performed mechanically, chemically, or by observation (for example through a microscope). Usually the

2 – Characteristics of diagnostic tests

analysis is performed in a pathology laboratory, but sometimes it can be performed with the patient present (point of care testing). Side effects from these tests are usually limited to the side effects resulting from obtaining the sample. However, in some cases, dietary or other changes are required of the patient before the test.

2.2.3 Imaging tests

Imaging tests include X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, radio-isotope imaging and positron emission tomography (PET), and other tests that produce images of part or all of the body in various forms. These tests usually require a person to interpret the image and considerable reader variability can arise. Moreover, individual patient characteristics, such as weight, muscle mass or tissue density can significantly affect the [accuracy](#) of these tests. Many of these tests involve types of radiation that are potentially direct sources of long-term adverse effects, particularly when the test is likely to be used repeatedly. Contrast media used in these tests may also be a direct cause of adverse events.

2.2.4 Endoscopy

Endoscopic examinations encompass tests such as colonoscopy, OGD (oesophagogastroduodenoscopy) and arthroscopy. These either produce images or are viewed directly through lenses. Endoscopic tests can be used for treatment as well as diagnosis; sometimes the diagnostic process can be curative such as when a biopsy completely removes a lesion. Endoscopic tests can result in direct adverse effects, including discomfort, anaesthesia risk, infection and perforation.

2.2.5 Decision rules and algorithms

Decision rules and algorithms are ways of combining the results of tests and observations to provide diagnostic information. In some cases, these rules are published and freely available. In other cases, the algorithm or rules are proprietary, with clinicians providing either laboratory samples or test results to a company which runs the algorithm and returns the results.

2.2.6 Diagnostic challenges

A diagnostic challenge is a test in which the patient is given a treatment or a chemical to assess its effect or their ability to tolerate the treatment. The patient's reaction is measured using one of the methods above or by asking the patient about changes in symptoms. There can be discomfort and adverse effects associated with the challenge.

2 – Characteristics of diagnostic tests

2.2.7 Questionnaires, structured interviews and surveys

Validated questionnaires can be used for the diagnosis and monitoring of disease. These types of tests do not fall within the EU directives which come within the remit of the Diagnostics Assessment Programme, and questionnaires are therefore beyond the scope of the Programme.

2.3 *Uses of diagnostic tests*

The four main uses of diagnostic tests are for:

- Diagnosis (section 2.3.1)
- Monitoring (section 2.3.2)
- Screening (section 2.3.3)
- Prognosis (section 2.3.4).

2.3.1 **Diagnosis**

Diagnosis is the process of identifying whether the patient has a specific disease, condition or syndrome at the time of testing. It is performed for patients with specific complaints or in whom signs or symptoms have been noted that may indicate a disease. Tests can have several different roles in the process of diagnosis.

Ruling in or out a specific disease

Some tests are used to either make (rule in) or exclude (rule out) a specific diagnosis. Using a test to make a diagnosis is the simplest to analyse because the test is being used for a specific diagnosis in a well-defined patient group. For tests used to exclude a diagnosis, the situation is more complex because the number of alternative diagnoses may be large. The sensitivity and specificity statistics of a test are only applicable for a specific disease being tested for. The value of a test in these circumstances consists not only of the value of the outcomes from the treatment given, but also of the avoidance of unnecessary tests or treatments when a diagnosis is excluded..

General examination looking for clues to the cause of the symptoms

Tests that are not specific for a particular disease pose difficulties. They may have sensitivities and specificities that differ between possible diseases, and the [prior probabilities](#) also vary for the different possible diseases.

Staging or disease severity

In some cases it is known what disease a patient has but not how severe or advanced it is. For many conditions, standard treatments will depend on the stage, grade or other measure of the severity of the disease, and in these

2 – Characteristics of diagnostic tests

cases it is generally advantageous to perform additional testing before treatment.

2.3.2 Monitoring

Monitoring is the process of following a patient over time to observe changes in their condition. Tests are repeated according to a defined schedule or according to changes in the condition. The intention is to detect changes with the aim of allowing timely intervention to prevent further deterioration or appearance of symptoms. Without monitoring, an intervention might be delayed, which could lead to both increased symptoms and reduced [efficacy](#) when the treatment is eventually initiated. Monitoring can be similar to screening (see below) since the purpose is to detect a change in a timely manner. The main difference is that the patient is already known to have, or is suspected of having, the condition being monitored and that monitoring is often looking for changes over time, which is not always the case with screening.

There are two main circumstances under which patients are monitored. The first is checking patients with chronic conditions to ensure treatment adequacy and/or monitor for progression in disease severity. The second is checking patients after treatment for the development of side effects, improvement in the condition, or recurrence of disease.

2.3.3 Screening

Screening tests look for conditions in patients without signs or symptoms of the specific condition. They can be general and given to the entire population, or they can be limited to patients in known risk categories. General population screening is a public health service in which members of a defined population, who may not be aware that they are at risk of, or already affected by, a disease, are offered a test. This test seeks to identify those people for whom further tests or treatments to reduce the risk of the disease, or its complications, are likely to be beneficial. This type of screening is typically initiated by the NHS. Tests being considered for this purpose need to be assessed in the context of screening programme criteria, and are normally assessed by the UK National Screening Committee. If there is ambiguity over which organisation should most appropriately evaluate a screening test, this is decided collaboratively between NICE and the UK National Screening Committee on a case-by-case basis. Screening can also be carried out for patients in known risk categories based on patient characteristics such as age, gender, genetic traits or comorbid conditions. Tests for these purposes may be assessed by the NICE Diagnostics Assessment Programme.

2 – Characteristics of diagnostic tests

Screening tests typically look for a single condition, but it is not uncommon for the tests to be capable of other incidental findings. Some of these findings can be beneficial to the patient and allow early treatment of conditions that would later cause problems. Other incidental findings can cause anxiety or additional costs and adverse effects in the same way as [false positive](#) results for the main diagnosis.

Screening tests can provide multiple types of information, listed below.

Early detection for treatment

The primary motivation for most screening tests is to detect a condition at a stage when treatment is more effective than waiting for the appearance of signs and symptoms. If treatment is equally effective once the symptoms appear, then this benefit is limited to the patient not having to experience the symptoms, which might be offset by the anxiety of knowing earlier or for a longer period that they have the condition.

Risk stratification

Some screening tests detect risk factors rather than a disease itself. This is becoming increasingly common with genetic tests (for example, tests for familial hypercholesterolaemia detect a high probability of very high cholesterol levels which are linked to an increased risk of heart attack and stroke). These tests, including non-genetic tests such as bone density scanning, do not detect the disease but help to determine the probability of it developing. The results of these tests can lead to efforts to modify risk, monitor the patient's condition or introduce treatment at an early stage.

Treatment adjuncts

Tests can be used during treatment to monitor progress or to direct treatment (for example, imaging during surgery or radiation therapy). Such uses of tests may be considered for evaluation by NICE as diagnostics or they may be included in NICE's assessments of the treatment.

2.3.4 Prognosis

Prognostic information allows the prediction of future events and outcomes. It may arise from a test that is used primarily for diagnostic, monitoring or screening purposes. Some tests are developed for the sole purpose of providing prognostic information.

Beyond identifying patients for treatment, the information from test results may be directly of use or value for those tested. In some cases, the value of the test result can be negative (for example, some patients might not want to

2 – Characteristics of diagnostic tests

know if they have a particular condition), but if this is the case, the patient would normally decline the test.

Tests carried out with prognostic intent can be helpful to patients for personal planning purposes. If the patient has information about their condition and its future course they can plan for the future more effectively. In the absence of this type of information, patients must wait until signs and symptoms appear, reducing the opportunity for this forward planning.

Prognostic information can also be of value to others besides the patient and physician. The information can be important to family members when planning for care-giving or other needs the patient may have in the future. It can also be informative to other family members about risks that they or their descendants may have.

The accuracy of most diagnostic tests is assessed by comparing the test with a reference standard at a particular point in time. However, for tests that generate predictions of future events (prognostic information), studies of test accuracy must continue for a longer time period to determine if the predicted events actually occur.

2.4 Outcome measures

Diagnostic tests affect outcomes in several ways. The principal output from diagnostic tests is usually information. A test may also have direct effects itself, such as test side effects, or direct benefits when the diagnostic test provides treatment (for example, a colonoscopy may result in the removal of a polyp or a cancer). Diagnostic tests can provide information that may affect treatment and the outcomes that the patient experiences as a result of that treatment. This section describes some of the relevant outcomes in more detail.

2.4.1 Intermediate measures

Diagnostic test accuracy statistics are intermediate measures and, when incorporated into models, can be used as predictors of future health outcomes experienced by patients. Other intermediate measures include the radiation exposure from an imaging test or the [pathogenicity](#) of specific genetic mutations identified by a genetic test. Diagnostic test accuracy may vary based on laboratory differences, the skill and experience of those administering or reading the test, batch and other variations in the materials, and the [cut-off point](#) on the [receiver operating characteristic \(ROC\) curve](#) used.

2 – Characteristics of diagnostic tests

2.4.2 Side effects from tests

The diagnostic test itself can induce side effects or other effects (both positive and negative). Examples of these effects include injury from invasive tests, reactions to contrast media or other ingested test chemicals, time and travel to get the test, discomfort from the test preparation or the test itself, immediate effects from a radiation overdose (for example, burns or nausea), and anxiety (or reassurance) as a result of the test results. The knowledge of a test result can also result in changes to a patient's expectations, behaviour, and actual health. Moreover, the diagnostic status identified by the test can sometimes affect employment, insurance, and other social and financial aspects of the patient's life.

A test result can lead to follow-up tests, either because it is equivocal, or because the standard protocol requires a confirmatory test (either for positive or negative results). Each of these tests has the potential to be associated with side effects or other negative or positive effects, as stated above.

2.4.3 Outcomes from the disease or disease modification

The most obvious benefits are those that arise from treating the identified disease(s). These treatments modify the identified disease and may also have adverse effects. Patients with negative test results may be spared the adverse effects of unnecessary treatment that might have been given if the diagnostic test were not available. Also, diagnostic errors mean that some patients (**false negatives**) will not receive the treatment or have treatment delayed until further symptoms appear. Second, some patients will receive no benefits from the treatment because they do not have the condition (false positives) but may experience the side effects or complications of the treatment. Patients with false-positive results may have another cause for their symptoms. Discovery of that cause may be delayed by the false-positive result with a reduction or delay in the benefits of treatment for that cause..

PART II: PROGRAMME PROCESSES

3. Selection of diagnostic technologies

NICE's Medical Technologies Advisory Committee (MTAC) selects medical technologies for evaluation, including diagnostic technologies. MTAC has defined criteria for selecting topics and routing them for evaluation by NICE. The following text is extracted from MTAC's methods guide (see <http://www.nice.org.uk/media/4E1/09/MedicalTechnologiesEvaluationProgrammeMethodsGuide.pdf>).

Considerations for routing technologies to the Diagnostics Programme

The Diagnostics Programme evaluates diagnostic technologies that have the potential to improve health outcomes, but the introduction of the technology is likely to result in an overall increase in resource costs to the NHS.

This Programme is likely to be suitable for evaluating diagnostic tests and technologies for which recommendations could only be made on the basis of clinical utility and cost–utility analysis. There should normally be a 'gold standard' or established comparator to enable an assessment of potential benefit of the technology. This Programme can evaluate classes of technologies or individual technologies.

Diagnostic technologies that appear likely to achieve a similar clinical benefit at less cost or more benefit at the same cost as current practice in the NHS may be more suitable for evaluation by the Medical Technologies Evaluation Programme.

One of the aims of the Diagnostic Assessment Programme is to promote the rapid adoption of innovative clinically and cost-effective diagnostic technologies. Some potentially important technologies will require evaluation at an early stage in the product lifecycle at a point at which there is relatively little evidence on which to base an evaluation. Balancing the need to support innovation with the availability of robust evidence is a key consideration at the topic selection stage. In some cases, the potential importance of a new technology may be such that it is selected for assessment by the Programme at an early stage. In other cases, technologies may not be selected for immediate evaluation because more comprehensive data are expected at a later date.

The main source of topic notifications to MTAC is product sponsors (technology manufacturers, developers, distributors and agents) via the NICE website (www.nice.org.uk/mt). Topics may also be suggested by other sources, such as National Clinical Directors, medical Royal Colleges,

3 – Selection of diagnostic technologies

professional bodies, national expert bodies, or national screening programmes. For each topic, a briefing note is prepared by NICE technical staff, based on information provided by the notifier and other sources. When a topic is selected that was not notified to NICE by the product sponsor, NICE contacts the product sponsor to invite them to take part in the evaluation. Product sponsors may choose not to provide data for the evaluation, but the evaluation will proceed without this input.

If a diagnostic technology requires CE marking, the Programme can only carry out an evaluation of that technology if the CE mark is received by the time any documents are issued for public consultation.

The Diagnostics Advisory Committee does not make recommendations regarding the use of a technology outside of its approved CE mark indications for use or licensed indications if applicable.

4. Initiation of the evaluation

4.1 Initiation date of an evaluation

The evaluation formally starts on the initiation date.

When a topic is selected by the Medical Technologies Advisory Committee (MTAC) (see section 3), it is normally scheduled into the next available slot for a diagnostics evaluation. The start of the timeline for that slot is the initiation date. If there is a gap of more than a few weeks between the MTAC referral and the initiation date, the sponsor of the [notified technology](#) is informed of the initiation date in advance in confidence and is sent introductory information about the Programme. The [topic lead](#) at NICE is appointed before the initiation date and is available for informal discussions with the sponsor before the initiation date. The External Assessment Group (EAG) is usually identified before the initiation date.

In exceptional circumstances a topic may not be scheduled for the first available slot but be allocated a different initiation date. For instance a topic of particular urgency to the NHS could be prioritised for evaluation before other technologies already identified. This decision is taken by the Centre Director taking into account the views of the product sponsor.

Should the initiation date need to be postponed for any reason, the sponsor and the EAG are informed.

4.2 Activities undertaken when an evaluation is initiated

On the initiation date four separate strands of activity are started:

- scoping begins (see section 5)
- contact is made with product sponsors and comparator manufacturers (see section 4.2.1)
- recruitment of specialist Committee members commences (see section 4.2.2)
- registered stakeholders are identified (see section 4.2.3).

4.2.1 Contact with sponsors and comparator manufacturers during the evaluation

When a topic is selected by MTAC, a Diagnostics Assessment Programme analyst is assigned to it as the topic lead. On initiation of the evaluation the topic lead contacts the sponsor of the notified technology with the following:

- general information about the Programme
- an invitation to participate in the evaluation

4 – Initiation of the evaluation

- details of how information (including confidential information) will be handled during the course of the evaluation
- the expected timetable of the evaluation
- details of the stakeholder registration process and an invitation to register

and a request to:

- formally agree to participate in the evaluation
- complete a confidentiality form and provide contact details
- provide all relevant data of which the sponsor is aware (including confidential and unpublished data) to enable scoping to start
- declare that all relevant data have been provided.

If [alternative technologies](#) are identified as described in section 5.2.1, the sponsors of these technologies are contacted by the topic lead to inform them that their product is being considered for inclusion in the evaluation and they are invited to register as stakeholders. If their technology is included in the final scope they receive the information and requests outlined in the previous paragraph.

Sponsors of all the technologies listed in the final scope may be contacted during the assessment period by the topic lead on behalf of the EAG with a request for additional information to assist with the assessment.

Manufacturers of non-generic comparator technologies (see section 5.2.1) identified during scoping are contacted and invited to register as stakeholders. They are not asked to provide data.

4.2.2 Recruitment of specialist Committee members

Specialist Committee members (described in section 1) are recruited at the beginning of the evaluation process and are appointed for the duration of a single evaluation (normally about 10 months). Both professional and lay specialist Committee members are appointed. These posts are advertised on NICE's website

(www.nice.org.uk/getinvolved/joinnwc/join_a_nice_committee_or_working_group.jsp) for at least 5 weeks. A panel consisting of the Chair of the Committee, the Programme Director and the Associate Director selects the specialist Committee members and their appointment is reviewed and ratified by the Centre Director.

It is possible that the full range of specialist knowledge and expertise required by the Committee will only become apparent when the final scope has been agreed. Additional specialist Committee members may therefore be appointed if necessary once the scope has been finalised.

4 – Initiation of the evaluation

Specialist Committee members are full decision-making members of the Committee, and are also members of the assessment subgroup (see section 5.5). In addition they may support the EAG on behalf of the Committee during the assessment phase. However they are expected to maintain sufficient independence from the assessment in order to be able to contribute to the Committee's discussions on the quality of the assessment and the development of guidance recommendations from that assessment.

Specialist Committee members must meet the requirements of NICE's code of practice for declaring and dealing with conflicts of interest (www.nice.org.uk/aboutnice/whoweare/policiesandprocedures/policiesandprocedures.jsp). They complete a declaration of interests form with their application and must declare any interests at the beginning of each meeting. Any applicant with an interest that would not permit them to take part in the Committee's decision-making is unlikely to be appointed as a specialist Committee member.

4.2.3 Identification of registered stakeholders

At the beginning of the evaluation the topic lead and the project manager search various sources to identify potential registered stakeholders (described in section 1). Sources may include former and current clinical guideline development groups, patient and carer organisations known to the Patient and Public Involvement Programme (PPIP) at NICE, product sponsors and other manufacturers, Royal Colleges and other professional organisations, and suggestions from the Diagnostics Advisory Committee and the Medical Technologies Advisory Committee. The individuals and organisations identified are contacted and invited to register. Additional potential stakeholders may be identified during the scoping period; if so they too are contacted and invited to register.

Stakeholders may register at any time during the development of guidance. There is more information here: [Diagnostic technologies stakeholder registration](#).

5. Developing the scope

The purpose of the scoping process is to ensure that the topic for evaluation is well defined and relevant, and that the evaluation is achievable within the time and with the resources available. The scope sets out what the evaluation will cover and the questions that need to be addressed in the evidence assessment.

The scoping process outlined in this manual is tailored to the specific requirements of the evaluation of diagnostic technologies.

This section includes:

- Understanding the care pathway (section 5.1)
- Contents of the scope (section 5.2)
- Contributors to the development of the draft scope (section 5.3)
- The scoping workshop (section 5.4)
- The assessment subgroup (section 5.5)
- Stopping evaluations at the scoping stage (section 5.6)
- Scoping steps and timings (section 5.7).

5.1 *Understanding the care pathway*

NICE aims to collect information about the outcomes (benefits and harms to the patient) through the entire care pathway, including the stages following diagnosis ([‘post diagnostic care pathway’](#)). Most of the relevant health outcomes will be the result of treatments given after diagnosis, and the treatment pathway or range of treatment pathways must be understood for the value of the diagnostic technology to be assessed.

Many diagnostic technologies are designed for multiple uses which, together with the need to estimate outcomes through the post diagnostic care pathways, result in the potential for assessments to become highly resource intensive. This assessment process can be complex and often requires significant input from clinicians with expertise in the topic under consideration. It is important to develop the scope for a topic so that the final guidance is as useful as possible to the NHS. The considerations include: the uses of the technology most likely to maximise benefit to the NHS and the population of England; whether there are sufficient data to carry out the evaluation; and the degree of complexity of the assessment.

Existing and emerging NICE clinical guidelines and other sources are used during the development of diagnostics guidance to ensure recommendations in diagnostics guidance are consistent with generally agreed post diagnostic care pathways.

5.2 Contents of the scope

The starting point for the development of the scope is the briefing note on the technology notified by the Medical Technologies Advisory Committee (MTAC) (see section 3). The briefing note includes the claim made by the sponsor for the advantages of the technology over current practice.

In most cases, the scope defines an assessment of the technology in terms of its use in specific clinical situations. In those cases, the scope defines the following aspects of each situation:

- patient population
- intervention (technology or test) to be evaluated and comparators
- care pathway
- outcomes and costs.

There is more detail on the above in part III of this manual.

NICE evaluates the scientific or engineering validity of a technology in the context of its impact on costs and patient outcomes.

5.2.1 How NICE decides which diagnostic technologies to include in an assessment

MTAC normally selects single products for assessment by the Diagnostics Assessment Programme (DAP) (see section 3). It may sometimes be appropriate for NICE to assess a diagnostic technology alongside alternative technologies. These are normally diagnostic tests that are similar in action or intent to the notified technology and, like the notified technology, are not in common use. This is generally done when, for example, the tests might be used in very similar settings or circumstances and there is likely to be some benefit to the NHS in developing guidance on more than one product or technology.

These additional technologies are identified during the scoping phase as a result of searches by the Information Services and Programme teams. Alternative technologies must meet MTAC's eligibility criteria (see www.nice.org.uk/mt), including any necessary CE marking or licensing before guidance is issued. The sponsors of alternative technologies are informed about the evaluation and invited to register as stakeholders.

The decision on which technology(ies) to include in the evaluation is taken by NICE after the assessment subgroup meeting (see section 5.5.1). Factors considered include both the resources available for the assessment and technical considerations that could affect the assessment, such as, but not confined to, the degree of [heterogeneity](#) of the tests and their purpose. The

5 – Developing the scope

final decision is made by the Centre Director and Diagnostics Advisory Committee Chair. The technology or technologies to be included in the evaluation are listed in the final scope. The notified technology is always included in the evaluation.

The sponsors of the notified technology and the sponsors of additional alternative technologies identified during scoping are involved in the evaluation process in the same way (see section 5.3).

During scoping, NICE takes advice from expert advisers (see section 5.3.2) to establish what tests or sequences of tests constitute current practice and these become the comparators in the assessment. Comparators are normally only considered within the requirements of their CE marking (if required) or licensed indication, unless they are used outside of these in routine clinical practice in the NHS.

If a comparator is a widely used generic technology (for example, conventional X-ray, cholesterol testing), manufacturers are not notified. If a comparator is non-generic (that is, a specific proprietary technology), the manufacturer is informed and invited to register as a stakeholder.

An in-house test may be considered for inclusion in an evaluation as an alternative technology or a comparator, providing it is used in compliance with regulatory requirements and is expected to be or is generally available.

NICE's diagnostics guidance contains recommendations on the use of the notified technology and the alternative technologies (if any are included). If by the time of public consultation on the draft guidance any of these technologies has not been CE marked or licensed (and this is required), they are not included in the draft guidance or in the final guidance. NICE's diagnostics guidance does not make recommendations on the comparator(s) technologies.

Product sponsors may choose not to provide data for the evaluation. In this case the evaluation proceeds without this input. They may nevertheless register as stakeholders and comment on the evaluation of their product.

5.3 Contributors to the development of the draft scope

The development of the scope involves a literature search undertaken by Information Services or the DAP technical team. If necessary this includes finding evidence from non-standard sources such as grey literature, manufacturers' data and other unpublished data. The DAP technical team also obtains information from other sources as outlined below.

5 – Developing the scope

5.3.1 Product sponsors and manufacturers of comparator technologies

The sponsor of the notified technology is contacted during the scoping process for information about the product and relevant available evidence. Manufacturers of alternative technologies that are being considered for inclusion in the evaluation are also contacted during this period. Manufacturers of proprietary comparator technologies may also be contacted.

5.3.2 Expert advisers

NICE identifies advisers with expertise in the technology and the care pathway (expert advisers) to contribute to the development of the scope. These experts are identified through literature searches, by asking Committee members for suggestions, by consulting existing clinical advisory support within NICE (for example, expert advisers who contributed to the briefing note or members of guideline development groups), and by contacting specialist Committee member applicants and registered stakeholders.

5.3.3 External Assessment Group

The External Assessment Group (EAG) may be consulted during the scoping process to ensure that the group is familiar with the direction the scope is taking, to take advantage of the group's expertise, and to assess the group's workload.

5.4 *The scoping workshop*

Once the DAP technical team has produced the draft scope, NICE holds a scoping workshop. The aims of the workshop are to:

- ensure the scope is appropriately defined, including verification and/or modification of the care pathway
- identify important evidence and any other issues relevant to the evaluation.

Discussions at the scoping workshop also help the EAG develop an understanding of the key issues which will feed into the evidence assessment.

NICE invites all registered stakeholders to attend the scoping workshop. Attendees, including representatives of relevant patient and carer organisations, are expected to have specific knowledge or experience of the condition, the technology, how the condition affects the patient, or the care pathway. Each person attends from their own perspective and does not represent the views of their stakeholder organisation. A maximum of two people from each registered stakeholder organisation may attend.

Stakeholders registered up to 1 week before the date of the scoping workshop are invited to attend.

5 – Developing the scope

Specialist Committee member applicants are invited to attend the scoping workshop.

5.4.1 After the scoping workshop

The DAP technical team revises the scope, taking account of discussions at the scoping workshop. Additional evidence identified at the scoping workshop is also investigated for its relevance to the scope.

The EAG sends its draft assessment protocol to NICE.

NICE agrees the revised scope and sends it with the draft assessment protocol to the assessment subgroup.

5.5 The assessment subgroup

An assessment subgroup is set up for each topic being evaluated. It normally comprises:

- the specialist Committee members for that topic
- the Chair of the Diagnostics Advisory Committee
- two standing members of the Committee
- NICE technical staff.

The purpose of the assessment subgroup is to ensure that the scope of the topic being evaluated, the assessment protocol (the work plan of the EAG) and the assessment itself are appropriately informed by the specialist knowledge and expertise of the Committee members.

The assessment subgroup and the EAG meet approximately 2 weeks after the scoping workshop to:

- review the revised scope and suggest amendments if necessary, and
- discuss the assessment protocol.

5.5.1 After the assessment subgroup meeting

The scope for an assessment may grow through the scoping process and become too large for the available assessment resources. If this becomes apparent, the scope may be revised by the DAP technical team in collaboration with the assessment subgroup and the EAG. The final scope is developed by the DAP technical team, signed off by the Programme Director or Centre Director, and published on NICE's website. Registered stakeholders are informed of this.

Should there be remaining scoping issues following the assessment subgroup meeting or should the EAG request an additional meeting to deal with issues

5 – Developing the scope

that arise, then a second meeting of the assessment subgroup may be scheduled at the request of the Chair of the Diagnostics Advisory Committee.

It may also be necessary to reduce the scope at an even later stage, if for instance it becomes clear during the assessment that a given technology or a particular indication cannot be assessed. Any amendments to the scope must be agreed by the Programme or Centre Director.

5.6 Stopping evaluations at scoping stage

It may become clear during the detailed scoping phase that a topic is not suitable for evaluation by the Diagnostics Assessment Programme and NICE may decide to terminate the evaluation at the scoping stage. This is expected to be uncommon. Registered stakeholders (including product sponsors), specialist Committee members and specialist Committee member applicants are advised if this occurs. The decision is made by the Centre Director.

5.7 Scoping steps and timings

These are approximate timings and may vary in response to individual evaluation requirements.

Table 3 Scoping phase timelines

Stage	Weeks (average) since phase began
Initiation of evaluation: <ul style="list-style-type: none">• NICE contacts sponsor of notified topic• NICE identifies potential registered stakeholders• NICE initiates specialist Committee member recruitment• External Assessment Group is identified (this may occur earlier)	0–2
NICE undertakes care pathway research and develops a detailed draft scope, making use of available clinical expertise	0–9
If potential alternative technologies are identified, NICE invites their sponsors to join the process	0–11
NICE holds the scoping workshop and following this develops a revised scope	9
NICE selects the specialist Committee members and appoints the assessment subgroup	10
The assessment subgroup meets with the External Assessment Group to review the revised scope and discuss the assessment protocol	11
NICE agrees the final scope for publication. This is normally published within 2 weeks (that is, by week 14)	12

5 – Developing the scope

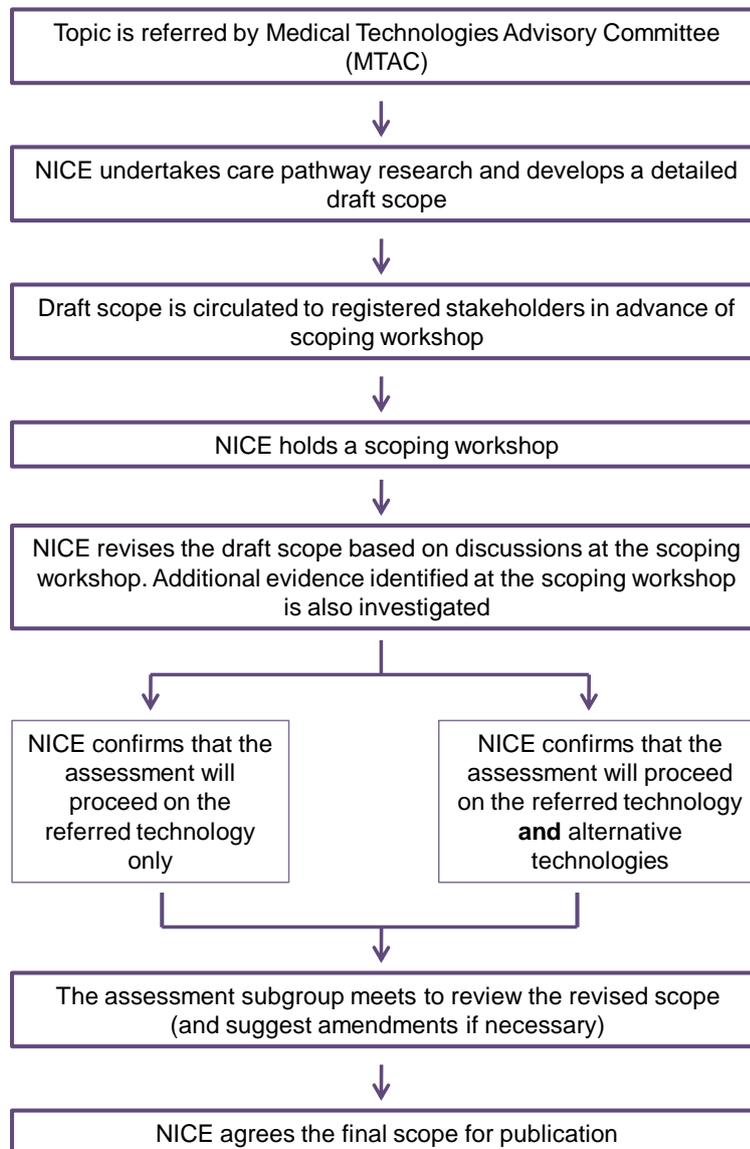


Figure 1 Steps in developing the scope

6. Assessing clinical outcomes and the cost effectiveness of diagnostic technologies

The assessment of evidence is carried out by the External Assessment Group (EAG) and presented in a diagnostics assessment report (DAR). This section considers:

- Impact of the scope on the assessment structure (section 6.1)
- Assessment methods used by the External Assessment Group (section 6.2)
- Diagnostics assessment report (section 6.3)
- Assessment steps and timings (section 6.4).

6.1 *Impact of the scope on the assessment structure*

The EAG is responsible for the structure of the assessment, but suggestions may be given by NICE based on the information obtained during scoping. The searches that are performed during the preparation of the scope may provide information about the quantity and nature of the available research evidence. The results of these searches are made available to the EAG.

It may become clear during scoping that it is possible to simplify the analysis. This is most likely if there are long-term studies or an already agreed [dominant](#) test. Similarly, if there are long-term follow-up trials, it may be possible to determine long-term outcomes without having to model this through intermediate outcomes. However, modelling is likely to be necessary to assess cost effectiveness.

The scope may provide information about the availability of data for key parameters that are required to model the care pathway. The scope may also include a suggested model structure based on the availability of evidence uncovered.

6.2 *Assessment methods used by the External Assessment Group*

The EAG undertakes an assessment of the diagnostic test accuracy, clinical outcomes and cost effectiveness of the technologies. The assessment is based on systematic reviews of the literature and data provided by the sponsors and information from the specialist Committee members, as well as modelling of patient outcomes, costs and cost effectiveness. The EAG's assessment highlights the uncertainties in the evidence and may include an analysis of the value of reducing those uncertainties.

6 – Assessing clinical outcomes and the cost–effectiveness of diagnostic technologies

The EAG may engage their own expert advisers and, if they wish, use information from the specialist Committee members or the expert advisers identified by NICE at the topic selection stage.

Part III of this manual contains discussions of methodological issues and options that relate to the assessment phase. Section 13 discusses issues relating to systematic reviews and evidence evaluation for diagnostics. Section 14 discusses aspects of modelling related to diagnostics, particularly aspects relating to assessing clinical outcomes. Section 15 contains a discussion of the reference case modelling of cost effectiveness.

6.3 *Diagnostics assessment report*

The EAG develops an assessment protocol, derived from the final scope of the evaluation. The assessment protocol outlines what the EAG will do during the assessment and the information it will provide in the DAR. The protocol is signed off by NICE and published on its website.

The EAG prepares the DAR in accordance with quality criteria agreed with National Institute for Health Research (NIHR) – Evaluation, Trials and Studies Coordinating Centre (NETSCC) and standard report templates. The EAG is responsible for the content and quality of the report. The Programme team and the Committee Chair liaise with the EAG and NETSCC to ensure that a satisfactory assessment report is produced.

The DAR does not contain recommendations on the use of a technology. The report forms part of the evidence base for the evaluation. The EAG further develops the report for subsequent publication as a topic in the Health Technology Assessment Programme (see www.hta.ac.uk/project/htapubs.asp).

NICE sends the DAR, with any confidential material removed, to registered stakeholders for comment. Stakeholders have 10 working days to return comments. Models supporting the DAR are made available to registered stakeholders on request during this period as outlined in section 1.6.

NICE presents anonymised registered stakeholder comments on the DAR, along with any responses from NICE or the EAG, to the Committee and later publishes these comments on its website. Comments should therefore not contain any confidential information.

After comments are received and considered, the EAG may need to perform additional analysis before the Committee meets to develop its draft recommendations. Any additional analysis forms part of the evidence base for the development of guidance, and is distributed to the Committee in advance of this meeting.

6 – Assessing clinical outcomes and the cost–effectiveness of diagnostic technologies

If possible, additional analysis is completed in time for the scheduled Committee meeting. If this is not possible, NICE may extend the timelines for the evaluation. NICE advises registered stakeholders as soon as possible of any change to the timelines and the reasons for that extension. The decision to undertake additional analysis or to extend the evaluation timelines is not taken lightly; it is done to ensure that NICE is able to provide robust guidance to the NHS.

6.4 Assessment steps and timings

These are approximate timings and may vary according to the requirements of individual evaluations.

Table 4 Assessment timelines

Stage	Weeks (average) since phase began
Final scope, final assessment protocol, list of specialist Committee members and list of registered stakeholders published on NICE website	2
NICE asks for relevant data from product sponsors on behalf of the External Assessment Group (EAG) (deadline for receipt of data is set by the EAG). EAG draws up diagnostics assessment report (DAR). Specialist Committee members contribute expertise	0–23
The EAG submits the DAR	24–25
DAR distributed to registered stakeholders for comment ^a	26
Deadline for receipt of registered stakeholder comments on DAR	28
DAR, registered stakeholder comments and EAG written response if any, and evidence overview sent to Diagnostics Advisory Committee (DAC)	28–29
^a The DAR is distributed to the product sponsors 48 hours before distribution to other registered stakeholders	

6 – Assessing clinical outcomes and the cost–effectiveness of diagnostic technologies

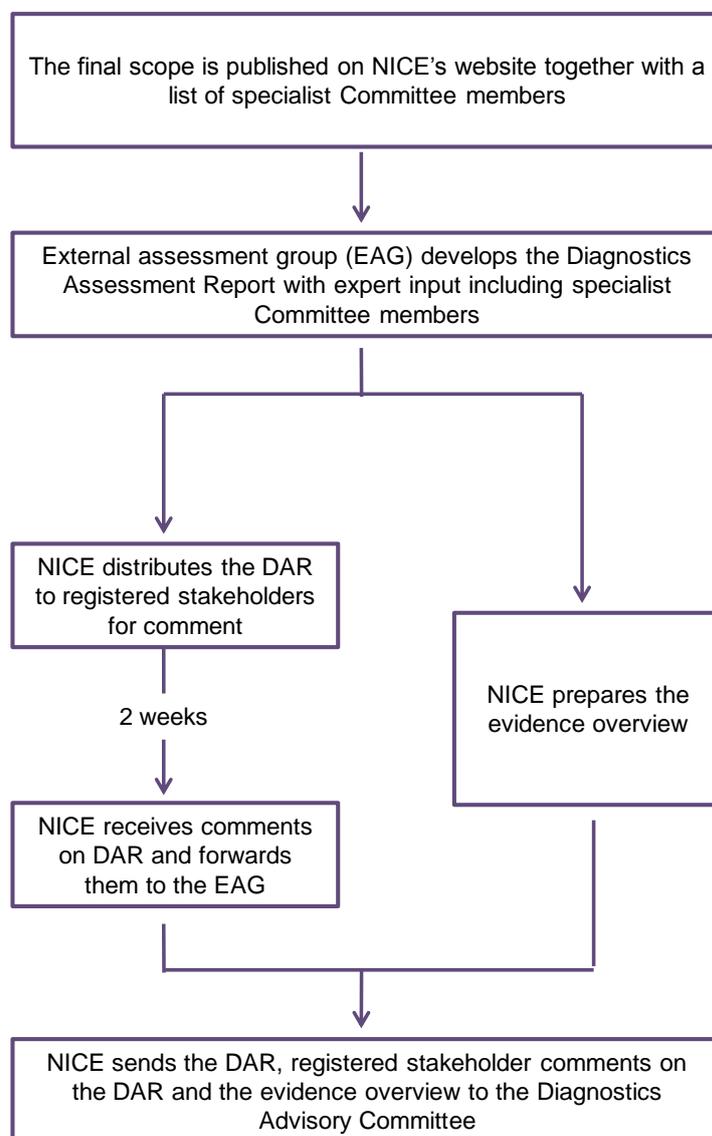


Figure 2 Steps in the assessment phase

7. Evaluation by the Diagnostics Advisory Committee

The purpose of this section is to explain the process followed by the Diagnostics Advisory Committee (DAC) in evaluating the evidence and formulating its recommendations. The methods that the Committee uses to do this are outlined in part III of this manual.

The evaluation phase has four stages which are explained in this section:

- Consideration of evidence and development of draft recommendations (section 7.1)
- Development of and consultation on the diagnostics consultation document (section 7.2)
- Review of the diagnostics consultation document in light of consultation comments (section 7.3)
- Preparation of the diagnostics guidance document (section 7.4).

The role of the Committee is:

- to consider evidence
- to make draft recommendations
- to consider public consultation comments
- to make final recommendations for publication in NICE guidance.

7.1 Consideration of evidence and development of draft recommendations

7.1.1 How Diagnostics Advisory Committee meetings are organised

The Committee normally meets 11 times a year in public. Agendas and minutes of Committee meetings are published on the NICE website.

Committee members are required to submit a declaration of interests on appointment, in every year of their tenure, and at each Committee meeting, in line with NICE's code of practice for declaring and dealing with conflicts of interest².

To promote public attendance at Committee meetings, NICE publishes a notice of each meeting and a draft agenda on its website at least 20 working days before the meeting. At this point, members of the public or an organisation who wish to attend the meeting can register their interest on NICE's website. Up to 20 places are available, depending on the size of the venue. In the event that attendance at any meeting is oversubscribed, NICE selects attendees according to its allocation procedure (see

² www.nice.org.uk/aboutnice/whoweare/policiesandprocedures/policiesandprocedures.jsp

7 – Evaluation by the Diagnostics Advisory Committee

www.nice.org.uk/newsroom/publicmeetings/advisorycommitteemeetings/advisorycommittees.jsp).

To allow wide public access, NICE reserves the right to limit attendees to one representative per organisation. The closing date for receipt of completed application forms is 10 working days before the meeting. NICE publishes the final agenda on its website 5 working days before the meeting. When the registration period has closed, NICE contacts successful applicants to invite them to the meeting. Along with the invitation, applicants receive a code of conduct for public attendees and a list of frequently asked questions. If a meeting is cancelled or a topic is re-scheduled, NICE gives attendees as much notice as possible.

Public access to meetings is granted in accordance with NICE policies (see www.nice.org.uk/newsroom/publicmeetings/advisorycommitteemeetings/advisorycommittees.jsp) and is subject to the standing orders of the Committee.

NICE publishes unconfirmed minutes of DAC meetings on its website within 15 working days. When the Committee has approved the minutes, NICE publishes confirmed minutes on its website, normally within 6 weeks of the meeting. The minutes of a Committee meeting provide a record of the proceedings and a list of the issues discussed. They do not record the Committee's decisions in relation to the topics under consideration.

7.1.2 The Committee meeting to develop draft recommendations

The following information is circulated to Committee members, usually 1 week before the Committee meeting:

- the final scope
- the list of registered stakeholders
- the diagnostics assessment report (DAR), including confidential material
- comments from registered stakeholders on the report
- the External Assessment Group's (EAG) written response, if any, to the comments on the DAR and any supplementary analysis the EAG has undertaken as a result of the comments
- an evidence overview written by NICE's topic lead for the evaluation.

The Committee meeting is in two parts: the public part (part 1) and the closed part (part 2). The Committee discusses the evidence and advice it has received in part 1 of the meeting. In part 2 the Committee considers any confidential information and formulates its draft recommendations. On occasion a meeting may be entirely public or entirely private – public if there is no confidential information and the Committee is not making any decisions,

7 – Evaluation by the Diagnostics Advisory Committee

and private if all the content of the meeting is confidential. This decision is made by the Committee Chair and the Programme Director.

In part 1 of the meeting the evidence is presented by the lead team which consists of two Committee members, one standing and one specialist member. This presentation highlights the most important issues relating to the evidence and outlines the issues and options facing the Committee. The presentation does not pre-empt the Committee's debate or the formulation of the Committee's recommendations.

EAG representatives attend both parts of the meeting to answer questions from the Committee and provide clarification on the diagnostics assessment report.

NICE staff members attend both parts of the meeting and may present evidence, provide advice on NICE policies and procedures, and respond to questions from the Committee.

Two representatives of each sponsor are invited to attend part 1 of the Committee meeting. The Chair may ask these representatives to respond to questions from the Committee. The Chair may also ask the representatives to comment on any matters of factual accuracy before concluding part 1 of the meeting. The Chair may ask the representatives to remain for part of the closed session (part 2) of the Committee meeting, specifically to respond to questions from the Committee about confidential information. The sponsor representatives are not present when the Committee develops its recommendations in closed session.

Each sponsor representative:

- should be an employee of the manufacturer or sponsor or have been contracted by the manufacturer or sponsor for the purposes of the evaluation
- should have relevant detailed knowledge of the technology under evaluation to engage effectively with the Committee
- should be able to comment on the diagnostic test accuracy, clinical effectiveness and cost effectiveness of the technology
- must agree to be bound by the terms and conditions of NICE's confidentiality agreement
- must be willing and able to discuss the condition and the technology with members of a large committee at a meeting where there may be members of the public and press observing
- should be familiar with the purpose and processes of NICE.

7 – Evaluation by the Diagnostics Advisory Committee

The diagnostics consultation document (DCD), diagnostics guidance document (DGD) and the minutes of Committee meetings report industry representation at the meetings but do not name the individual representatives.

Registered stakeholders, including manufacturers of comparator technologies, may apply to attend Committee meetings as members of the public. Because the Committee does not make recommendations on the comparator technologies, these manufacturers are not questioned by the Committee and their company names are not listed in the minutes of the meeting.

7.2 Development of and consultation on the diagnostics consultation document

After the Committee meeting, the Programme team drafts the DCD based on the discussions at the meeting, including the draft recommendations agreed by the Committee.

The DCD normally contains the following:

- the Committee's draft recommendations to the NHS on the use of the technology or technologies
- a description of clinical need and practice in the relevant clinical area
- a description of the technology or technologies
- a summary of the evidence on clinical effectiveness and cost effectiveness
- the issues the Committee considered important in reaching its recommendations ('Committee considerations')
- proposed recommendations for further research, if appropriate
- implementation considerations if appropriate
- a list of related NICE guidance.

The DCD is issued for a 4-week consultation period, normally within 15 working days of the Committee meeting. Registered stakeholders are sent the DCD and have 20 working days to comment. Five working days after the DCD is sent to registered stakeholders, the DCD, the DAR (with confidential information removed) and the evidence overview are put on the NICE website and members of the public have 15 working days to comment. Comments may be submitted via the website, by email, fax or post. Comments are collated by the Diagnostics Assessment Programme team for the Committee's consideration. The anonymised consultation comments are made available on NICE's website when the final guidance is posted, and should therefore not contain confidential information.

Comments of more than 20 pages are not normally permitted; this may be waived in exceptional circumstances at NICE's discretion.

7 – Evaluation by the Diagnostics Advisory Committee

The purpose of the consultation is to seek views on the Committee's draft recommendations and to determine whether they are an appropriate interpretation of the evidence considered. NICE invites comments on whether:

- all the evidence available to the Committee has been appropriately taken into account
- any significant evidence was missed or incorrectly recorded
- the summary of the evidence available on clinical and cost effectiveness is appropriate
- the draft recommendations are sound and constitute a suitable basis for guidance to the NHS
- there are any equality issues that need special consideration that are not covered in the DCD.

All comments are important and potentially influential in developing the guidance, including those that support the draft recommendations.

7.2.1 DAR addendum

Diagnostic models are complex and the Committee may ask the EAG to undertake additional work at the same time as the DCD is developed and consulted on. This additional work is documented in the [DAR addendum](#).

7.3 Review of the diagnostics consultation document in light of consultation comments

NICE normally sends the Committee members the full text of the comments received. NICE may, at its discretion, summarise comments that are overly lengthy.

The Committee meets to consider the DCD in the light of the consultation comments received and the DAR addendum. This meeting is held in public and is split into part 1 (open) and part 2 (closed). The External Assessment Group is invited to attend the meeting. The Committee discusses the responses to the consultation on the DCD in part 1 of the meeting and moves to a closed session to consider any confidential information and to agree the final recommendations.

The Committee considers the impact of the consultation comments on all sections of the DCD, but in particular on:

- the draft recommendations on the use of the technology
- the recommendations for further research.

The Committee considers all comments and, when appropriate, amends its recommendations, exercising its own judgement on the nature and

7 – Evaluation by the Diagnostics Advisory Committee

importance of the comments from consultation. The content of the 'considerations' section in the guidance document is modified as necessary to include the issues identified via the consultation which the Committee has taken into consideration and its view of these issues.

If the DAR addendum, comments and/or new evidence submitted during consultation lead to a substantial revision of the proposed guidance (such as a major change in the recommendations, considerations and/or evidence base), the Centre or Programme Director decides whether it is necessary to repeat the consultation process. If the decision is to repeat the consultation process, a revised DCD is prepared, submitted to the Committee and issued for consultation. NICE distributes the same documents with the second DCD as with the first, as well as the DAR addendum and the new evidence (if applicable) and the consultation comments on the first DCD. The process continues as with the initial round of consultation followed by a third Committee meeting to consider any additional comments received. The timelines for the evaluation are extended accordingly.

7.4 Preparation of the diagnostics guidance document

After the final Committee meeting the Programme team drafts the DGD based on the discussions at the meeting, and the final recommendations agreed by the Committee.

The DGD has the same structure and content as the DCD except for any revisions proposed by the Committee at its meeting. The Programme team submits a report to NICE's [Guidance Executive](#) (made up of NICE's Executive Directors and Centre Directors). The Guidance Executive decides whether it is satisfied that the guidance has been developed in accordance with NICE's published processes. If so, the Guidance Executive approves the DGD for publication on behalf of the NICE Board (subject to resolution – see section 8).

In exceptional circumstances, for example, if relevant information is published after the consultation period closes but before the DGD is published, NICE may undertake further analysis before circulating the DGD. If warranted, a new consultation is undertaken as outlined in the last paragraph of section 7.3. The Centre or Programme Director takes this decision in discussion with the Chair of the Committee and the Programme team. The decision is not taken lightly and is made to ensure NICE is able to provide robust guidance to the NHS.

8. Resolution and publication of guidance

This section covers the following:

- Resolution grounds (section 8.1)
- Eligibility to make a resolution request (section 8.2)
- Resolution requests (section 8.3)
- Initial scrutiny of resolution requests (section 8.4)
- The resolution panel (section 8.5)
- Publication of diagnostics guidance (section 8.6)
- Steps and timings for evaluation, resolution and publication (section 8.7).

The resolution process takes place after NICE's Guidance Executive has approved the guidance for publication and before it is published. The resolution process is a final quality-assurance step to ensure that NICE acts fairly, follows its own processes and produces clear, accurate guidance. It prevents the inadvertent publication of guidance that contains factual errors or is developed other than in accordance with this manual.

If NICE receives a resolution request, it suspends publication of the final guidance while it investigates the request. If NICE does not receive a request, the final guidance is published as soon as possible after the resolution period ends.

The resolution process applies only to the guidance document. It does not apply to the diagnostics assessment report or other documents produced in the course of developing the guidance.

8.1 Resolution grounds

The resolution panel (see section 8.5) only considers resolution requests that clearly meet one or both of the following grounds:

- Ground 1: Breach of NICE's published process for the development of diagnostics guidance.

An example would be when a step is missed in the process.

- Ground 2: Factual errors in the guidance.

A factual error is an objective error of material fact in the final guidance. Conflicting scientific or clinical interpretations or judgements are not considered to be factual errors. For example, if a resolution request states that a statistic quoted in the guidance is incorrect, NICE establishes whether the final guidance misquoted the statistic, or if one statistic was preferred out

of several because the Committee considered it to be more reliable. The former is a factual error; the latter is a difference of scientific or clinical judgement.

8.2 Eligibility to make a resolution request

After the Guidance Executive approves the guidance, NICE sends an email with the DGD, the public consultation comments on the DCD and NICE's response to those comments to all those who responded to the draft guidance consultation. It is important that any organisation or person who may wish to make a resolution request submits a consultation response at the appropriate time, even if this is a simple comment or a comment supportive of the draft guidance. They should bear in mind that the final guidance may change significantly from the consultation document because of comments received during consultation and considered by the Committee when formulating its final guidance.

8.3 Resolution requests

A resolution request on one or both of the grounds given above must be made within 15 working days of receiving the email. NICE accepts requests by email, fax or letter addressed to the Associate Director of the Diagnostics Assessment Programme. People making requests should specify the resolution they seek. NICE can then fully understand the nature of their concern and take appropriate action.

8.4 Initial scrutiny of resolution requests

All eligible resolution requests are subject to an initial scrutiny process. The Associate Director investigates the matters raised and reports the findings to the Centre Director, who decides whether the request falls within the scope of the resolution process. Initial scrutiny continues for 15 working days after the resolution request period ends. If multiple resolution requests are made, either from the same or different sources, each request is treated as outlined below and in table 5.

Ground 1: breach of process

If the Centre Director considers that the resolution request does not meet ground 1 (breach of process), or does not have a reasonable prospect of success, the Associate Director informs the person or organisation that made the request and NICE publishes the guidance.

If the Centre Director considers that ground 1 appears to have been met, the Associate Director convenes the resolution panel (see section 8.5).

Ground 2: factual errors

If the Centre Director considers that the resolution request does not meet ground 2 (factual errors), or does not have a reasonable prospect of success, the Associate Director informs the person or organisation that made the request and NICE publishes the guidance.

If the Centre Director considers that the guidance contains a minor factual error or a point that requires clarification but does not affect the Committee’s recommendation(s), the guidance is amended and signed off by the Committee Chair without being referred to the resolution panel. NICE then publishes the final guidance in the usual way.

If the Centre Director considers that there may be a major factual error that cannot be remedied by minor amendment, they instruct the Associate Director to convene the resolution panel.

In the event of multiple resolution requests, in the view of those conducting the initial scrutiny, not all requests may qualify to be referred to the resolution panel. In order to avoid pre-empting the outcome of resolution, NICE informs everyone who has submitted a resolution request that the panel is to be convened, and that NICE will tell them the outcome of their request after the panel's decision is made.

Table 5 Initial scrutiny of resolution requests

Outcome of initial scrutiny	NICE action
Ground 1 not met	Guidance is published
Ground 1 met	Resolution panel is convened
Ground 2 not met	Guidance is published
Ground 2 met, minor factual error	Guidance is amended and published
Ground 2 met, major factual error	Resolution panel is convened

8.5 The resolution panel

The panel consists of two NICE Board members: one Non-Executive Director and one Executive Director not previously involved in developing guidance on the technology. The aim of the panel is to decide whether there has been a breach of process or factual error and, if so, what action is appropriate. The outcome of the panel meeting is outlined below and in table 6.

8.5.1 Meeting

The Associate Director organises the resolution panel meeting, which takes place no more than 20 working days after the initial scrutiny process has ended. Panel members may attend the meeting by video conference or telephone.

8 – Resolution and publication of guidance

The Programme team prepares a briefing, which the panel uses when considering the resolution request. For ground 1, this means establishing what process was followed when developing the guidance and what events or omissions were alleged in the resolution request. In the case of ground 2, this involves setting out what evidence lies behind the alleged errors.

The Associate Director and, if needed, the Committee Chair attend the meeting to provide clarification. They are not members of the panel and do not contribute to the outcome of the resolution. Members of the Programme team may also attend the meeting to answer questions.

8.5.2 The outcome

Ground 1: Breach of process

If the resolution panel decides that there has been no breach of process, NICE can publish the final guidance. If the panel decides that there has been a breach of process, it decides what action is appropriate. This may involve repeating part of the assessment process and, if necessary, referring the technology back to the Committee and/or carrying out another consultation.

Ground 2: Factual errors

If the resolution panel decides that there are no factual errors, NICE can publish the final guidance. If the panel decides that there are factual errors or elements to be clarified, NICE produces an amended version of the guidance. The panel must decide whether the error can be corrected and the amended version of the guidance approved by the Guidance Executive before publication, or whether the Committee should review the wording of the amended guidance in light of the error identified.

NICE considers whether to publish the amended guidance or whether there is a need for further consultation. This need normally arises if:

- NICE makes a substantive change to the wording of the recommendation(s)
- changes to the guidance not involving the recommendations are significant or likely to be of interest to the people who made the resolution request.

The Associate Director implements the panel's decision and informs everyone who made resolution requests of the outcome of resolution. This normally occurs 2 days before NICE publishes the final guidance, although this timescale does not apply if the Committee needs to reconsider the recommendation(s).

8 – Resolution and publication of guidance

The resolution panel's decision is final and there are no further opportunities for redress within NICE.

Table 6 Outcome of resolution panel meeting

Outcome of resolution panel meeting	NICE action
Ground 1 not met	Guidance is published
Ground 1 met	Appropriate action as decided by resolution panel
Ground 2 not met	Guidance is published
Ground 2 met	Appropriate action as decided by resolution panel

8.6 Publication of diagnostics guidance

After resolution is complete, NICE publishes the diagnostics guidance. Guidance is published on NICE's website only.

NICE publishes the following on its website with the diagnostics guidance, after the resolution period has closed:

- the diagnostics assessment report (DAR), with confidential data removed, registered stakeholders' comments on the DAR and any response from the External Assessment Group (EAG)
- evidence overview
- anonymised consultation comments on the diagnostics consultation document, including any new non-confidential evidence
- NICE's response to the consultation comments
- further analysis or correction, if any, undertaken by NICE or the EAG subsequent to the DAR (the DAR addendum)
- implementation support tools
- lay explanation of the recommendations when appropriate.

Implementation support tools, published along with the guidance, help the NHS to implement the diagnostics guidance and may include audit support, costing tools, slide sets to explain how the guidance can be put into practice, or other specific products when needed.

Within 2 weeks of publication of the final guidance, a debriefing meeting may be arranged with product sponsors if such a meeting is considered of value by one or more of the product sponsors or by NICE.

If NICE is advised of any potential errors in the guidance or the supporting documents after publication, these are dealt with according to NICE's standard procedures.

8.7 Steps and timings for evaluation, resolution and publication

These are approximate timings and may vary according to the requirements of individual evaluations.

Table 7 Approximate timings for the evaluation phase

Stage	Weeks (average) since phase began
NICE sends the diagnostics assessment report (DAR), registered stakeholders' comments on the DAR and the evidence overview to the Diagnostics Advisory Committee (DAC)	0–1
DAC meeting to develop draft recommendations	2
NICE agrees the diagnostics consultation document (DCD) for consultation	5
Consultation on the DCD (4 weeks for registered stakeholders, 3 weeks for general public)	5–8
NICE collates consultation comments	8
Final DAC meeting to consider the consultation comments and develop final recommendations	10
Diagnostics guidance document (DGD) is produced	12
NICE Guidance Executive approves guidance for publication, subject to resolution	14
Resolution period	15–18
NICE publishes diagnostics guidance	23

A timeline of the complete guidance development process is in appendix C.

Steps in the evaluation, resolution and publication phase

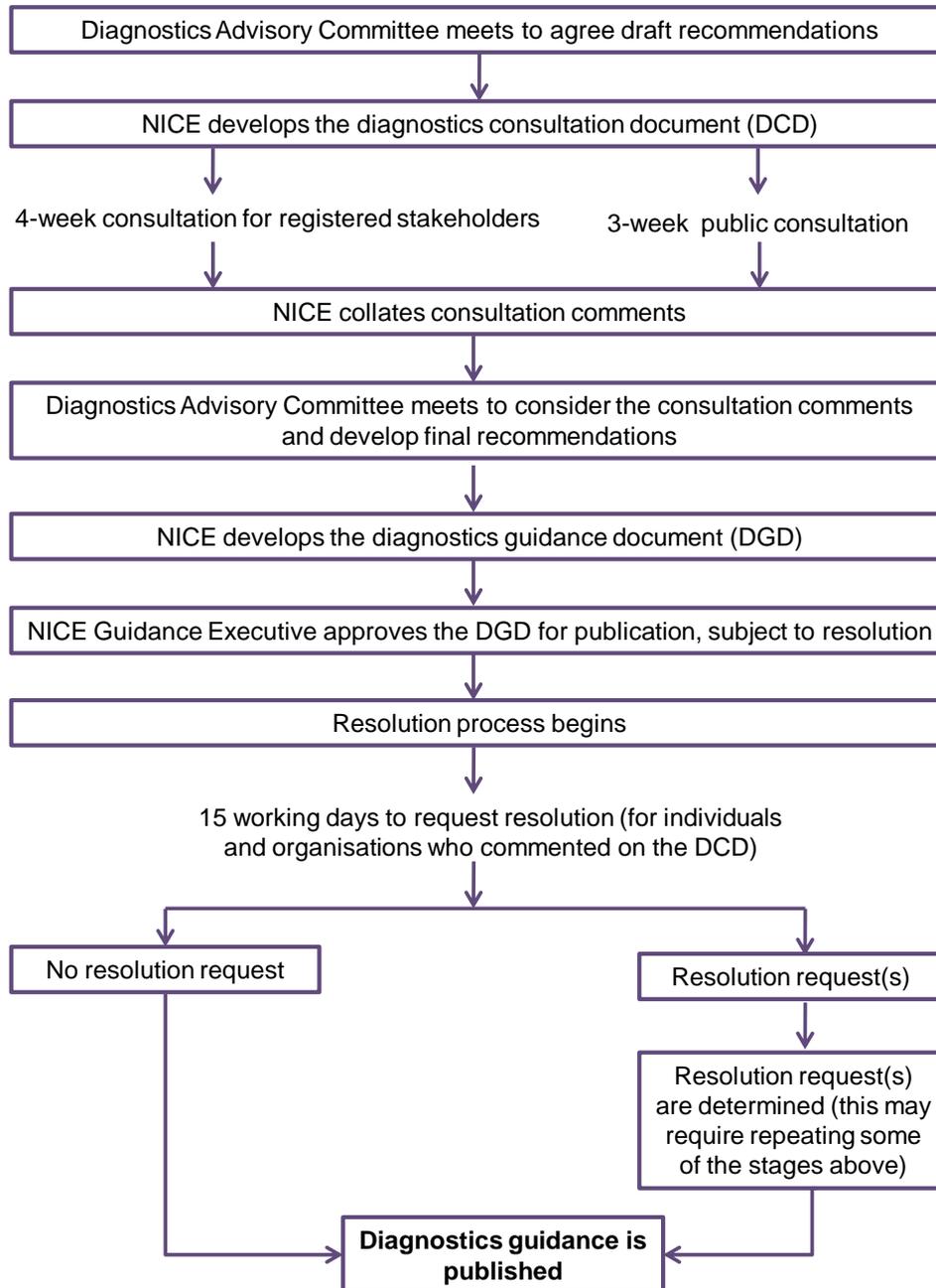


Figure 3 Steps in the evaluation, resolution and publication phase

9. Guidance reviews

After the guidance is published the Programme team updates the literature search at least every 3 years to ensure that relevant new evidence is identified. At the same time, NICE contacts product sponsors and other stakeholders about issues potentially affecting the value of the diagnostic technologies, including significant changes to the price of the product or the comparator. In addition to this, NICE may review and update diagnostics guidance at any time if significant new evidence becomes available. Stakeholders, including product sponsors, researchers and clinicians, can inform NICE of developments in the evidence base. When NICE reviews diagnostics guidance it may decide not to update it; or to reassess the topic with a view to issuing updated guidance; or, if appropriate, to withdraw the guidance.

The process of reviewing guidance and submitting review proposals to the Guidance Executive forms part of the normal workload of the Programme. NICE includes guidance updated as a result of the review process in the Programme's annual target for guidance development.

10. Updating this manual

This manual is subject to the approval of the NICE Board and a review will normally be initiated 3 years after its publication. It may be necessary to make minor changes to the procedures for developing diagnostics guidance before that time. Procedural changes will be made in accordance with NICE's policy. Minor changes that may be made without consultation are those that:

- do not add or remove a fundamental stage in the process
- do not add or remove a fundamental methods technique or step
- do not disadvantage one or more stakeholders
- improve the efficiency, clarity or fairness of the process or methodology.

Changes meeting these criteria will be published on the NICE website 4 weeks before their implementation.

Any other changes will only be made after a public consultation period of 3 months.

PART III: METHODS USED FOR DECISION-MAKING

11. Introduction to Programme methods

This part of the manual describes the methods used for decision-making in the Diagnostics Assessment Programme. The choice of method for any one technology will vary depending on the nature of the technology and the amount of available evidence. This manual therefore describes a range of methodological options to be applied within the Programme. Selection of the specific methods for a particular topic is based on the technology being evaluated and the available evidence.

The approach to the evaluation of diagnostics involves estimating the outcomes that the patient will experience as a result of using the diagnostic technology, estimating the costs to the healthcare system, and determining the cost effectiveness of using the technology. The outcomes and costs typically include those arising from treatments following the use of the technology and cover the entire relevant portion of the care pathway.

In principle, the approach to assessing the cost effectiveness of diagnostic technologies is similar to treatment assessments. Added complexity arises because controlled trials reporting all relevant outcomes are rare for diagnostics. More extensive modelling is involved, including the initial testing, follow-up testing, treatment and monitoring. The specific content of the assessment will vary depending on the nature of the diagnostic technology or technologies being evaluated and the availability of evidence about the care pathway. Modelling for diagnostics is often required, both to estimate clinical effectiveness (patient outcomes) and to assess cost effectiveness. Often the same model is used for both purposes. The model provides the framework for the use of information to guide Committee decision-making.

The scope (described in section 12) defines the overall nature of the patient groups, interventions and outcomes to be included in the evaluation. This provides an outline for the basic structure for evidence acquisition and modelling, although the details are developed during the assessment based on the nature of the available evidence.

12. Developing the scope

The scope is developed as outlined in section 5. The detailed information gathered in the scope is outlined in this section as follows:

- Patient population (section 12.1)
- Intervention (technology or test) to be evaluated and comparators (section 12.2)
- Care pathway (section 12.3)
- Outcomes and costs (section 12.4)
- Other considerations (section 12.5).

12.1 Patient population

This covers the patient characteristics, the conditions to be diagnosed, and the [aetiologies](#) of the conditions if relevant to the assessment.

A diagnostic test can affect outcomes in a number of ways. Tests can help determine whether or not treatment is undertaken, which treatment is undertaken and the intensity of the treatment (for example, dose of a drug, extent of surgery, duration or frequency of treatment). The selection of treatments directly affects the final outcomes that matter to the patient, so an understanding of the outcomes of treatment based on patient characteristics is essential to evaluating the benefits of the tests. Subgroups included in the analysis of the diagnostic test may therefore include those that have different outcomes from treatment even when the test accuracy is the same.

The group or groups of patients to be studied is carefully defined. Outcomes can vary significantly depending on the patient population evaluated: there may be differences in the [prior probabilities](#) for various conditions identified by the tests; there may be differences in test accuracy in different patient populations; there may also be differences in the impact of treatment, and side effect or complication rates. For many tests, there may be multiple patient populations. In order to keep the evaluation to a reasonable size, some patients who are potential users of the diagnostic test may not be included in the scope. Importantly, the exclusion of patients from the scope should not be taken to mean that the test is inappropriate for these patients. Because resources for the assessment are limited, patient groups may need to be selected carefully to maximise the benefit of the assessment. Examples of relevant factors in identifying patient populations for the scope include:

- probability of disease
 - genetic factors
 - prior testing results
 - presenting symptoms or situation

12 – Developing the scope

- relevant physiology or body type
- ethnicity
- previous exposure to risk factors
- **aetiology** of the disease
- disease stage, grade, or severity
- factors that may affect the test or test accuracy
 - body habitus (size, shape and other conditions affecting test accuracy)
 - current medications
 - ability to complete the test
 - disability that affects application of the test
- factors that affect the benefits or risks of treatment
 - comorbidities
 - age, gender, ethnicity or other genetic factors.

Definition of the patient population also includes where, why and how the test is used in the care pathway. These are described for each defined population.

12.2 Intervention (technology or test) to be evaluated and comparators

As outlined in section 5, the technology is usually defined during the topic selection process and details of the technology are provided in a briefing note. During the scoping process details about the technology and the available evidence are sought from the relevant manufacturers, literature searches and experts in the field.

The notified technology may have multiple uses. An imaging device, for example, may be used in many ways to diagnose or monitor different conditions. The scoping stage aims to identify the most important potential uses for the technology (within their specified indications for use or licensed indications when appropriate) and the subsequent evaluation then focusses on these specific uses.

Alternative tests or technologies either not in common use and newly available or soon to be available may be included alongside the notified technology. The scoping process for these alternatives is similar to that for the notified technology. If the alternative is available, or is likely to become available during the evaluation, it may be considered for inclusion in the assessment.

If the diagnostic test or tests under consideration are used in sequence or in sequence with other tests, the technologies that comprise the potential sequence are also included.

12 – Developing the scope

The setting for use of the tests (for example, hospital, specialist centre or general practice) is specified.

The comparator or comparators are the technologies or tests that are most commonly used or are recommended in current NICE guidance for the functions in the evaluation. There may be multiple tests or variants or test sequences in common use and all are included as comparators.

In the scope the comparators are described in the same amount of detail as the technologies being evaluated. The remainder of this section therefore applies both to the interventions being studied and to the comparators.

The description of the tests to be studied needs to be precise because there may be many variants of a single technology that could be used, and these variants may need to be evaluated separately. All potential, relevant alternatives are outlined in the scope.

Below is a list of some of the issues that may be considered when developing the scope. NICE seeks expert advice (see section 5.3.2) when considering their relevance.

12.2.1 Alternative cut-off points for test interpretation

Many tests have a range of cut-off values that can be used to determine when a test is considered positive. These can be different values for a laboratory measurement, different sizes or densities on an X-ray image, or even different means of doing the test (for example, position of a blood pressure cuff or different laboratory instruments). Sometimes the values are derived by clinical consensus (for example, the glucose level for identifying diabetes). Some differences are a result of equipment differences and some differences in human accuracy in reading the tests or pictures. This can result in variation in the reported test sensitivity and specificity between studies, and may also reduce [applicability](#). It may be appropriate at the scoping stage to specify one or more cut-off points to be included in the analysis.

12.2.2 Alternative sets of follow-up/confirmatory tests

Tests are frequently done in conjunction with other tests. Sometimes they are done concurrently; at other times, subsequent tests may be required based on the results of earlier testing. Each of these combinations or planned sequences can be viewed as a separate intervention. If the combination or sequence of tests is viewed as a unit, an overall sensitivity and specificity can either be found from direct data about the test combination/sequence or be computed (subject to some assumptions about correlation and independence). Because there may be many reasonable combinations and

12 – Developing the scope

sequences, it is usually appropriate to focus on the most likely or most efficient sequence(s).

12.2.3 Variations of the test

Tests with the same general name can vary in practice (for example, with/without contrast, T1 versus T2 weighting for MRI, and alternative laboratory procedures). Even a routine conventional X-ray can vary according to the exposure characteristics used. Other imaging tests can vary according to the type and amount of contrast, the method of computation used and the probe in use. Laboratory tests can vary according to the machine and reagents used. When specifying the scope of the evaluation it is important to specify any variations that are relevant.

12.2.4 Timing between tests or for initial test

One of the more complex issues can be the timing of tests. As a disease progresses, the accuracy of a test can change. For example, if the disease is more advanced, tests are typically more likely to uncover it. The timing of the discovery of a disease can affect the efficacy of treatment and quality of life. If these are potential issues, the scope indicates the patient population by disease state or stage and the likely timing of the tests or test sequences.

In the case of screening tests, the timing of the initial and subsequent screening is critical. Because of the problems of [length bias](#) and [lead-time bias](#) (see section 13), different timings for the screening can greatly affect the benefits of the screening. In such cases, different timings need to be investigated in the assessment because of the impact on both costs and benefits.

The timing is similarly relevant to monitoring tests. The timing of the changes being monitored can affect outcomes in a manner similar to diseases being screened. Different timings for monitoring are considered as part of the scope if feasible.

12.3 Care pathway

The care pathway is an important part of the process of assessing diagnostic effectiveness and costs. The care pathway includes the entire sequence of tests and treatments relevant to the topic. It may also include tests or treatments that are performed to deal with the adverse effects of the tests and treatments in the pathway. The care pathway can vary depending on the patient's conditions, characteristics or comorbidities.

The scope includes a description of the care pathway, including any variations according to test results or the tests used. This care pathway defines the time

frame for the treatments covered, key steps leading to final outcomes, and the outcomes relevant to treatments that will be included in the assessment. It covers the diagnostic sequences, treatments, monitoring, retreatment, treatment for adverse effects and complications that may be experienced by the patient. In some cases, the care pathway includes tests and interventions that are not carried out because of the results of the test under study. For example, if a test diagnoses a condition that would not have been diagnosed by the comparator, then the benefits of not undergoing other treatments or tests would be relevant. Even if a test diagnoses an untreatable condition, the costs and harms of treatment that can now be avoided are relevant.

The scope may include a flow chart or other diagram to illustrate the pathway.

12.4 Outcomes and costs

The scope defines the key health outcomes for the assessment. It is unlikely to describe all parameters needed for the development of the model. The External Assessment Group may need to specify these parameters as the model is being constructed. The scope defines the relevant cost areas for the assessment but it does not detail all the specific costs and other resource details to be incorporated in the assessment.

Relevant outcomes include any health outcomes resulting directly or indirectly from the use of the test. They may also include informational outcomes of value to the patient for the relief (or imposition) of anxiety or for personal planning. All health benefits (or harms) resulting directly or indirectly from the use of the diagnostic tests (including both the true and false results) should be included. These include longer-term outcomes in most cases. Similarly all costs stemming from the use of the test should be included. This analysis should be done for the test(s) being evaluated and the comparator(s).

Outcomes that follow directly from the diagnostic tests are always included in the scope, and outcomes from tests and treatments undertaken based on the results of the diagnostic tests are usually included. Downstream outcomes may be omitted if the test under study has an accuracy that is essentially identical to that of the comparator tests. In that case, the scope can be shortened to include only the outcomes from the test and the test costs, because the downstream outcomes are expected to be the same. Generally, only those outcomes and costs are included that vary depending on the diagnostic test selected.

Several questions related to outcomes are addressed during scoping. These questions inform the development of the protocol for the assessment that is carried out by the External Assessment Group.

12.4.1 What is the time horizon for the analysis?

The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect all important differences in costs or outcomes between the technologies being compared. Many technologies affect costs and outcomes over a patient's lifetime. This is particularly the case with treatments for chronic diseases. In such instances, a lifetime time horizon for clinical and cost effectiveness is appropriate. A lifetime time horizon is also needed for any mortality component in order to quantify the implications of any differential survival effect between alternative technologies. For a lifetime time horizon, extrapolation modelling is often necessary. If the impact of treatment beyond the results of the clinical trials is uncertain, analyses should be presented that compare several alternative scenarios, reflecting different assumptions about future treatment effects (see section 14 on modelling). Such assumptions should include the limiting assumption of no further benefit as well as more optimistic assumptions. Analyses with a time horizon shorter than the expected impact of treatment are not usually considered to provide the best estimates of costs and benefits.

A time horizon shorter than lifetime could be justified if there is no differential mortality effect between options, and the differences in costs and relevant benefits relate to a relatively short period (for example, in the case of an acute infection). Consideration of the time horizon and the uncertainty around extrapolating data beyond the duration of the clinical trials is a critical component of the assessment.

12.4.2 Should the impacts on the treatment of other conditions be included?

Costs and outcomes related to the condition of interest and occurring during additional years of life gained as a result of treatment should be included in the reference-case analysis (see section 15). Costs that are considered to be unrelated to the condition or technology of interest occurring during periods of additional longevity should generally be excluded.

12.4.3 What are the relevant costs?

It is important that the costs model parallels the benefits included, so that it includes all costs necessary to obtain the benefits (or harms) stemming from the testing. These include the costs of the test itself (including any retests), and of follow-up testing, treatment, treatment of adverse effects from the test or treatment, and any monitoring needed before or after the treatment. For the reference case (see section 15), the perspective on outcomes should be all health effects, whether for patients or, when relevant, other people (principally carers). The perspective adopted on costs should be that of the NHS and personal social services (PSS). Technologies for which a substantial

proportion of the costs (or cost savings) are expected to be incurred outside the NHS and PSS, or which are associated with significant non-resource effects other than health, are identified during the scoping stage of an evaluation.

The reference-case perspective on outcomes has the objective of maximising health gain from available healthcare resources. Some features of healthcare delivery that are often referred to as ‘process characteristics’ may ultimately have health consequences; for example, the mode of treatment delivery may have health consequences by affecting treatment adherence. If significant characteristics of healthcare technologies have a value that is independent of any direct effect on health, these should be noted. These characteristics include the convenience with which healthcare is provided and the level of information available for patients.

NICE works in a specific context; in particular, it does not set the NHS budget. The Diagnostics Assessment Programme offers guidance on the efficient use of available NHS and PSS resources. Therefore the reference-case perspective on costs is that of the NHS and PSS.

Some health technologies may have a substantial impact on non-health outcomes or costs to other government bodies. These impacts are usually identified during scoping. Diagnostic evaluations that consider costs incurred outside the NHS and PSS will always be agreed with the Department of Health (and other relevant government bodies as appropriate) and detailed in the final scope. For these non-reference-case analyses the benefits and costs (or savings) to other government bodies are presented separately from the reference-case analysis. Productivity costs and costs borne by patients and carers that are not reimbursed by the NHS or PSS are not included in either the reference-case or non-reference-case analyses. Section 15 gives more information about the reference case.

12.5 Other considerations

In some cases, the scope identifies issues relating to the technology that are not particular to specific clinical situations. For example, a new imaging machine may have cost or radiation exposure aspects that cover a broad range of clinical conditions. In this case the scope defines a wide category of patients, but the subsequent care pathway for those patients may not be included in the assessment. Specialised scope layouts may be used in such situations.

The scope may discuss special considerations. These include anything likely to affect the potential assessment, including equality and diversity issues, or special implementation issues. These are topic specific.

13. Evidence for assessment and evaluation

This section sets out in detail the kind and sources of evidence to be used by the External Assessment Group (EAG) when assessing the technology and by the Diagnostics Advisory Committee when evaluating it. Much of the evidence is the same as for the assessment and evaluation of other technologies. Although evidence for treatment [effectiveness](#) may be included in a diagnostics assessment as part of the care pathway, this section focuses on the aspects of assessment that are specific to diagnostics.

This section covers:

- Introduction (section 13.1)
- Types of evidence (section 13.2)
- Identifying and synthesising evidence on diagnostic test accuracy (section 13.3)
- Identifying and synthesising evidence for health outcomes, including test side effects (section 13.4)
- Identifying evidence for cost effectiveness (section 13.5).

13.1 Introduction

A comprehensive evidence base is needed when assessing diagnostic technologies. However, the amount and quality of the evidence directly relating to diagnostic tests is generally much lower than for other technologies such as drugs.

A first step in the evidence gathering process is to search for studies that follow patients from testing, through treatment, to final outcomes (these are sometimes termed ‘end-to-end studies’). These end-to-end studies may be of varying quality and design and could include randomised controlled trials (RCTs), cohort studies and observational studies. If these studies exist, a systematic review of this evidence may remove the need for more extensive searches to identify evidence for model parameters. The next step is to search for data on test accuracy, other direct outcomes and costs stemming from the test. Another early step is to explore existing models of the management or treatment of the condition after diagnosis. If these models exist then they can be used to inform the overall model.

In each case, the methods of identification, selection and analysis of data need to be presented in a transparent way.

13.2 Types of evidence

If, as is likely, there are no end-to-end studies available for a diagnostic technology, then different types of evidence are collected and a linked evidence approach taken. If no data can be found for a particular parameter relating to the care pathway or treatment effect, or if a wide range of values are identified, then expert opinion or [expert elicitation](#) may be used to provide a parameter value for the model(s), or the model(s) can be redesigned to use other parameters.

13.2.1 Test accuracy

Diagnostic test accuracy studies compare test results of people with a disease or condition to those of people without it. Designs are generally prospective cohort or cross-sectional studies, or retrospective case-control studies. Most compare a single index test of interest with a reference standard in order to calculate the accuracy. Paired design studies compare two index tests with each other, and often also with a reference standard. These studies are less prone to [bias](#) resulting from confounding but are rarely available. Identifying and synthesising test accuracy evidence is explored later in this section.

13.2.2 Test side effects

Data may be identified in RCTs and other comparative studies. However, cross-sectional studies, case studies and patient registries may be of most benefit for assessing side effects, particularly adverse effects from tests. If appropriate high-quality systematic reviews of test side effects are available, these can be used to provide estimates for the assessment. There is more information later in this section about identifying and synthesising health outcomes, including test side effects.

13.2.3 Existing models

If appropriate high-quality models of the management and treatment following a diagnosis exist, these are used to provide information for the assessment.

13.2.4 Treatment effectiveness

If relevant high-quality systematic reviews of treatment effectiveness are available, these can be used to provide estimates for the assessment. If a systematic review is not available, either an individual RCT or a [meta-analysis](#) of multiple RCTs is the optimal source of evidence on treatment effectiveness. Other comparative designs, such as controlled studies, cohort studies and case-control studies may provide useful evidence, but are at a higher risk of bias (most notably patient selection bias, but other biases are also more likely).

13.2.5 Care management

Clinical guidelines from NICE and other organisations can provide a good background to care management and the care pathway. Diagnostic before-and-after studies also provide useful information on any change in management following the introduction of an index test to clinical practice. These studies are rarely available, especially when assessing a new test that is not in routine clinical use. As such, expert clinical input on the usual care pathway is likely to be important.

13.2.6 Impact of misdiagnosis

The direct impacts of a false-negative or false-positive result on outcomes are very rarely reported in the literature and often need to be estimated using expert clinical judgement.

13.2.7 Impact of test usage differences

A test being evaluated may affect outcomes because it is used differently or has different characteristics than the alternatives or comparators. The test may produce results more quickly, thus reducing the need for the patient to attend extra appointments or reducing the lag time to treatment. The test may produce fewer direct adverse effects and this, in addition to the benefits of the patient not experiencing the adverse effect, may increase patient adherence to the treatment. These outcomes can be included in the assessment, but may require expert opinion or expert elicitation because evidence may be lacking in most cases.

13.2.8 Health-related utility, costs and resource use

Direct data may be available about the outcomes of interest, including costs resulting from treatments after use of a diagnostic technology. These data may be used if appropriate. Data might also arise from existing models, which can be used as an alternative to de novo modelling if the existing models are adequate and appropriate.

13.3 Identifying and synthesising evidence on diagnostic test accuracy

The objective of analysing diagnostic test accuracy data is to produce unbiased estimates of accuracy data for all interventions and comparators included in the scope. Data on diagnostic test accuracy should be identified by a systematic review process.

13.3.1 Systematic review

Diagnostic test accuracy studies should be systematically reviewed using a pre-defined protocol. This protocol should permit the inclusion of evidence from all sources likely to inform the decision about the use of the assessed technologies by the NHS within the scope of the assessment. These sources can include published and unpublished data, data from non-UK sources, and data from registries and other observational sources. Selection of the data for inclusion should be based on incorporating as much data as possible while minimising the biases to both internal and external validity. Thus higher quality data should be used where available but lower quality data may be considered in the absence of such data.

Search strategies

Search strategies for reviews of diagnostic test accuracy tend to be longer and more complex than search strategies to identify treatment effects. Filters should not be used to narrow the search to diagnostic studies because indexing of these types of studies is often poor. A search strategy should be developed iteratively, with new terms added based on the results from the previous search. In addition to searches of key literature databases it is critical to review other sources of evidence.

- Manufacturers and experts in the field should be contacted and asked if they know of unpublished data or studies not identified by searching the literature databases.
- Systematic reviews should be identified by searching specialist databases, for example MEDION, ARIF and DARE. The bibliographies of any relevant systematic reviews should be searched for studies missed.
- Bibliographies of identified studies should be reviewed to identify any studies that have been missed, and the bibliographies of any new studies identified should also be reviewed. Hand searching conference abstracts may yield additional relevant studies.

Increased searching expenditure gives increased search recall. However, the optimal strategy is not necessarily the one that would retrieve the maximum amount of data.

Study selection and data extraction

As with systematic reviews of treatment effects, a list should be compiled of potential studies identified through the searches. Each study should be assessed to determine whether it meets the inclusion criteria for the review. These criteria are usually in the protocol developed by the EAG. A list of excluded studies should be maintained, and include the reason for exclusion. The validity of the process is increased if two reviewers assess each study

and disagreements are resolved by discussion or a third reviewer. All methods should be detailed in the diagnostics assessment report.

Biases

Variability in the results of different diagnostic test accuracy studies is to be expected by chance, and imprecision may arise for other reasons, for example from small sample sizes. The risk of bias depends on the quality of the study method design, and execution, and other circumstances of the study. A related issue is the applicability of the results to routine clinical practice. Types of bias that particularly apply to diagnostic tests include but are not limited to:

- **Selection bias:** arises from an error in selecting people to take part in a study, resulting in variation between the arms of the study; this is less likely to be an issue for diagnostic cohort studies in which all patients receive both tests, but can be highly significant in diagnostic case-control studies in which the cases are not carefully matched to the controls.
- **Information bias:** occurs when the results of the index test are interpreted with knowledge of the results of the reference test or vice versa.
- **Imperfect reference standard:** arises if the reference standard does not correctly classify patients with the target condition, yielding over- or underestimates of test accuracy.
- **Disease progression bias:** occurs due to a delay between administering the index test and the reference test, during which time a target condition could change.
- **Partial verification bias:** arises when a non-random subset of patients does not undergo verification with the reference standard.
- **Differential verification bias:** occurs when some patients are verified by one reference standard and others by another reference standard, particularly if the choice of reference test depends on the result of the index test. The difference in reference standard is not necessarily a different test, but could include differences in implementation (for example, different laboratory setups, manufacturers of reagents) or differences in interpretation (for example, different readers of X-ray images.)
- **Incorporation bias:** occurs if the reference test and index test are not independent; for example, the result of the index test is used in establishing the presence of the target condition and therefore becomes a part of the reference test, or parts of the reference test are included in the index test.
- **Excluded data:** occurs when uninterpretable results or withdrawals are not described or included in the analysis.
- **Spectrum bias:** arises if included patients do not represent the patient in whom the test is intended for use in clinical practice; examples include

differences in disease severity or in how early or late in the referral process the test is performed.

- **Lead time bias:** occurs in screening and monitoring studies, in which early detection can appear to result in a longer survival time; in reality, morbidity and mortality remain the same but patients are aware of their condition for longer.
- **Length bias:** occurs in screening and monitoring studies, in which aggressive disease is more likely to present between monitoring points of a given interval than less aggressive disease.
- **Hawthorne effect:** occurs if clinicians or patients adjust an aspect of their behaviour as a result of being included in a study.

Poor reporting quality in diagnostic test accuracy studies can hinder the assessment of bias. The STARD (STAndards for the Reporting of Diagnostic accuracy studies) initiative aims to improve the accuracy and completeness of diagnostic accuracy study reporting. The STARD statement consists of a 25-item checklist and its use is encouraged.

Assessment of quality and risk of bias

Each study included in the systematic review should be critically appraised to assess the validity of its results. This ensures potential sources of bias are identified and can be taken into account when drawing conclusions from the review. The QUADAS (Quality Assessment of Diagnostic Accuracy Studies) is a checklist developed for critically appraising diagnostic accuracy studies. It is recommended that this, or a modified version, is used to assess the quality of the diagnostic accuracy studies included in the review.

NICE is aware that a revision of the QUADAS checklist is currently in progress (QUADAS-2). Once published, the updated version of the checklist may be used to critically appraise included studies.

It should be noted that although the QUADAS checklist gives a good indication of the quality of a study, it does not cover all possible aspects and does not help in assessing whether the study will be useful for modelling the decision problem. Care should be taken to assess studies specifically to determine whether their use is informative in the context of the scope.

Study applicability

A key issue is the applicability of the study to the situation for which the recommendations are being made. Ideally, the study population, patient conditions (such as comorbidities or genetic makeup), the setting and nature of the intervention should correspond exactly to those for which the guidance is being written. Studies from other countries, other aetiologies of the disease

or other diseases, other patient groups, or non-typical healthcare settings may not accurately reflect the usefulness of the test for the situation under consideration. These differences should be documented, with reference to the decisions taken about which studies to include in the assessment.

Evidence combination

Meta-analysis of test characteristics may be used if multiple, comparable studies are available. Results from meta-analysis of test accuracy data may have more power to detect differences between tests than results from single studies alone. However, meta-analysis of test accuracy data is more complicated than meta-analysis of clinical effectiveness data because of the correlation between sensitivity and specificity. In addition, there are likely to be many sources of heterogeneity across test results, arising from differences in setting, patient population, reference standard, equipment, procedures and skill levels of test operators. The cut-off point at which test accuracy data are reported is also likely to differ between studies.

Several methods for meta-analysis of test accuracy data are given in the literature. These vary in complexity and in the assumptions that need to be made. What method to use depends on the data available. The choice of method should be justified by the EAG in the diagnostics assessment report.

13.3.2 Graphical presentation of test accuracy results

Meta-analysis produces useful graphical displays that can be used to investigate heterogeneity across results from different studies. Coupled forest plots and summary ROC curves can both be used to help interpret meta-analysis results. Which form to use depends on the nature of the available data.

Paired forest plots

Paired forest plots present sensitivity and specificity on separate but adjacent plots. Accuracy data from each study are presented on the same row, together with confidence intervals. These plots can be useful to show the heterogeneity between sensitivity estimates and the heterogeneity between the specificity estimates, although they do not display the correlation between the two accuracy measures.

Summary ROC curves

Summary ROC curves, derived from meta-analysis, depict how sensitivity and specificity vary as cut-off points vary. Several methods can be used to generate a summary ROC curve, depending on the situation: the Moses-

Littenberg model, the hierarchical summary ROC model or the bivariate model (see section 13.3.3).

Data from multiple studies of paired tests can also be presented. Points are plotted for a normal summary ROC curve, but with the two estimates (one from each test) from each study joined by a dotted line. This shows how accuracy differs between tests within a study as well as the variability between studies. Summary ROC curves and point estimates can also be computed and added to each plot.

13.3.3 Types of meta-analysis

Types of meta-analysis include among others:

- separate meta-analysis of sensitivity and specificity
- meta-analysis of [likelihood ratios](#) and predictive values
- meta-analysis of diagnostic [odds ratios](#)
- Moses-Littenberg summary ROC curves
- hierarchical models.

The choice of method for meta-analysis is a trade-off. Data are not always available for the more complex and technically accurate methods. Also the benefits from more complex methods may not be worth the increased analytical costs. The choice of method should be customised to the specific situation.

Separate meta-analysis of sensitivity and specificity

Separate meta-analysis of sensitivity and specificity assumes there is no correlation between sensitivity and specificity. It should be used only if data are reported at similar cut-off points. Separate analyses for different cut-off points can be conducted if enough data are available. Each study can contribute accuracy data from one cut-off point to each analysis. Random-effects meta-analysis methods are recommended to deal with other sources of heterogeneity between the studies.

The assumption about consistency of cut-off point and lack of heterogeneity does not hold true in most cases. Therefore separate pooling of sensitivity and specificity will usually not be suitable.

Meta-analysis of likelihood ratios and predictive values

Separate analysis of positive and negative likelihood ratios does not take into account the correlation between the two measures. Pooling of [positive predictive values](#) and [negative predictive values](#) are affected by disease

prevalence, which is likely to vary between studies. Neither method is recommended for combining diagnostic test accuracy data.

Meta-analysis of diagnostic odds ratios

The advantages of analysing diagnostic odds ratios are that sensitivity and specificity are analysed as a pair, and that standard analytical techniques can be used to combine the data. A confidence interval can be calculated that allows comparison of performance between different tests. Random-effects meta-analysis methods are recommended because of the likelihood of heterogeneity between studies. The disadvantage is that there are challenges in interpreting the clinical implications of the results.

Moses-Littenberg summary ROC curves

The simplest way to calculate a summary ROC curve is using the Moses-Littenberg method (Littenberg 1993). It should be used only if there is little between-study heterogeneity and may be preferred over separate analysis of sensitivity and specificity if data are reported at a variety of different cut-off points. Each study can contribute accuracy data, using one cut-off point, to the analysis of the summary ROC curve.

Disadvantages are that it does not provide an estimate of the heterogeneity between studies, and it should not be used to calculate a point estimate value or standard errors. Although useful to investigate how test accuracy may depend on covariates, it does not provide a clinically informative estimate of sensitivity and specificity because the cut-off point corresponding to a chosen accuracy cannot be identified.

Hierarchical models

Hierarchical models require fewer assumptions than are needed for analysing sensitivity and specificity separately and for creating a summary ROC curve using the Moses-Littenberg method. Two possible approaches are the hierarchical summary ROC model (Rutter 2001) and the bivariate model (Reitsma 2005). The output from these analyses include the summary ROC curve, a point estimate, a 95% credible or confidence region, and a 95% predictive region. The two methods have been shown to be mathematically equivalent when no covariates are fitted (Harbord 2007). One of these methods should be used in diagnostic assessments if data are available but do not meet the assumptions needed for a separate meta-analysis of sensitivity and specificity.

Results from paired design tests can be analysed using hierarchical models. However, this method of analysis assumes that the data have come from a randomised design – that is, patients were randomised to receive either index

test A or index test B, plus the reference standard. Methods for analysing data from paired studies in which each patient received both index tests plus the reference standard are not well established and are not currently recommended.

Other methods

Analytical methods are currently being developed that will be able to incorporate data using multiple cut-off points (Dukic 2003, Hamza 2009). These should not be used in NICE diagnostic assessments unless the EAG can justify their use in a specific assessment.

13.3.4 Alternatives to meta-analysis

If studies are heterogeneous, as demonstrated by graphical presentation of the results, point estimates calculated from meta-analysis can provide misleading results and it may be preferable to use individual study results in the modelling, combined with ranges (derived from the range and uncertainty of the results of the available studies) for [sensitivity analysis](#). Expert clinical input can help guide the decision as to which particular study or data point is preferred. The existence and magnitude of the heterogeneity should be captured and reported clearly.

13.3.5 Imperfect reference standards

Reference standards may not perfectly reflect whether the patient has the condition. The actual condition of the patient may not become known until the condition develops further, at surgery, or in some cases until autopsy. With this in mind, the imperfect reference standard may still be the best practical assessment to use in studies of the technology of interest. If the reference standard is not universally used or can vary from study to study or site to site, this can generate additional heterogeneity and uncertainty. Even if the reference standard is universally used and reasonably consistent between sites, if it is imperfect then the resultant sensitivity and specificity estimates will also be imperfect.

Depending on whether and how the new test and the reference standard are correlated, the error introduced in the estimate for the new test could be in either direction. It may be possible to correct the error if the nature and magnitude of the correlation is known. For example, if the new test and the reference standard are closely correlated, the errors from the reference test probably apply to the new test. Lack of correlation means that differences between the reference test and the new test could be errors from the reference test as opposed to the new test. If the degree or nature of the correlation is not known, the analysis should explore the possible impact of alternative assumptions about the strength and direction of the correlation.

13.3.6 Expert elicitation

In some situations published studies do not exist for key parameters of the assessment. In these cases, expert elicitation may be used, preferably with a formal approach that provides some indication of the expert's uncertainty. Such formal approaches typically involve assessing probability distributions, usually after training the responders about the various types of common cognitive biases. Less formal approaches to solicit expert opinion may be considered if necessary.

13.4 Identifying and synthesising evidence for health outcomes including test side effects

Existing systematic reviews of studies of test side effects and treatment effectiveness should be reviewed. If high-quality systematic reviews exist, a de novo review is not necessary. If they do not exist then data on test side effects and health outcomes from management and treatment after diagnosis should be reviewed using the principles described below. These principles are in alignment with standard methods used for health technology assessment.

13.4.1 Systematic review of health outcomes

Study selection and data extraction

A systematic review should be conducted according to a previously prepared protocol. Once a search strategy has been designed and literature searches undertaken, study selection and data extraction should be performed as described in 13.3.1.

Critical appraisal

The validity of results of an individual study depends on the robustness of its overall design and execution. Each study included in the systematic review should be critically appraised using a method appropriate to the study type.

Validity of results

Many factors can affect the overall estimate of relative treatment effects or test side effects obtained from a systematic review. Issues affecting internal and external validity include the characteristics of the patients, the care setting, any additional care provided to patients and when the study was conducted (because clinical techniques progress over time). These issues need to be identified before data are analysed and conclusions are drawn.

Meta-analysis

It is appropriate to synthesise outcome data through meta-analysis, provided there are sufficient relevant and valid studies available that use comparable outcome measures.

If there are multiple end-to-end studies comparing the same diagnostic techniques in the same treatment algorithms in comparable patient populations, it is reasonable to directly meta-analyse the studies. This situation is unusual.

Forest plots are useful for illustrating individual study results, as an addition to tabulating the characteristics and possible limitations of the data.

Pooling of data using statistical techniques should be accompanied by an assessment of heterogeneity. Statistical heterogeneity in results can, to some extent, be taken into account using a random (as opposed to fixed) effects model. However, the degree of, and the reasons for, heterogeneity should be explored as fully as possible. Known clinical heterogeneity may be managed by careful use of meta-regression.

If there is doubt about the relevance of a particular study, a sensitivity analysis should be undertaken that examines the effect of excluding such trials. If the risk of an event differs substantially between the control groups of the studies in a meta-analysis, an assessment should be carried out of whether the measure of relative treatment effect is constant over different baseline risks.

13.5 Identifying evidence for cost effectiveness

13.5.1 Existing models

The search process for existing models of cost effectiveness does not need to be as extensive as the search process for data because the objective is not to identify all available models, but to identify appropriate, high-quality models. Once identified, models need to undergo a critical appraisal using a suitable tool and an assessment of external validity in relation to the decision problem. If suitable models are found, they can be used or modified as appropriate. If no suitable models are found then a de novo model can be constructed during the assessment process.

13.5.2 Costs and resource use

The base case of the model should use actual costs of the technology as it is or will be used in the NHS. If data are available for costs in the NHS these should be used. If not, prices submitted by the manufacturer should be used.

Ideally these prices should reflect actual national prices paid, not just a list price.

Unit costs and prices associated with testing procedures, treatments and resource use should ideally be taken from current official listing published by the Department of Health. In addition, national databases on healthcare resource groups, such as reference costs or the Payment by Results tariff, are a valuable source of information and should be considered for use if they are appropriate and available. Data based on healthcare resource groups may not be appropriate in all circumstances (for example, if the definition of the group is broad or the mean cost does not reflect the actual resource use). In such cases other sources of evidence, such as microcosting studies, may be more appropriate. In all cases all relevant costs should be included, such as the costs of the test, follow-up, treatment, monitoring, staffing, facilities, training and any other required modifications.

13.5.3 Duration and health-related quality of life

The measure preferred by NICE for duration and quality of life is [quality-adjusted life years](#) (QALYs). This measure is described further in section 15. QALYs include not only longevity or mortality effects, but also the health-related quality of life (HRQL). The impact on HRQL is based on health changes through the entire care pathway. The quality of life is based on the health states experienced through a lifetime of health states, although only those states that are changed as a result of the changes in the diagnostic tests should be included in the analysis.

In some cases, there may be direct data expressed in QALYs or a similar measure. These data can be identified by reviewing the literature for studies reporting HRQL. Determining HRQL usually comprises two elements: the description of changes in HRQL itself and a valuation of that description of HRQL. Information on changes in HRQL as a result of treatment should be reported directly by patients. If it is not possible to obtain the information directly from patients, then data should be obtained from their carer (not from healthcare professionals).

The valuation of changes in HRQL reported by patients should be based on public preferences elicited using a choice-based method in a representative sample of the UK population. The [EQ-5D](#), a standardised and validated generic instrument, is the preferred measure for valuing changes in HRQL, although other measures may be used if necessary as long as consistent measures can be obtained. In some cases, other measures are needed because the EQ-5D is insufficiently sensitive.

13 – Evidence for assessment and evaluation

In some cases, HRQL may be based on existing studies. If no such studies are available, then patient or expert opinion or elicitation may be required.

1

14. Modelling clinical outcomes and the cost effectiveness of diagnostic technologies

This section describes modelling issues that are likely to arise during an evaluation of a diagnostic technology. Section 15 describes the reference case that applies to NICE cost-effectiveness analyses of diagnostic technologies. Although the details of models are not finalised during the scoping phase, the structure of the model likely to be needed is explored during scoping to ensure that the assessment phase can proceed effectively.

This section covers:

- Simplified analyses (section 14.1)
- Structuring the assessment (section 14.2)
- Considerations for assessment and modelling (section 14.3)
- Identifying future research needs from the evidence (section 14.4).

14.1 Simplified analyses

The best approach to establishing the relative effectiveness of diagnostic technologies is with studies that randomise patients and follow them from the initial diagnostic tests through treatment to final outcomes. If high-quality end-to-end outcome studies are available, the analysis can be greatly simplified and complex modelling processes may not be needed to estimate final outcomes. However, studies of this type are expensive and rarely done, and those that are done are sometimes not of high quality.

Another situation in which a simple model structure may be developed is if the new diagnostic test is superior to its comparator(s) in both sensitivity and specificity (at the relevant thresholds) and is no worse in direct test side effects. In this case, and assuming the use of a cost-effective treatment once diagnosed, a test could be cost saving if the test costs less than the comparator. Complex cost-effectiveness analysis would be unnecessary in this situation. However, important assumptions need to be made about equivalence of the tests, therefore care should be taken to ensure that the estimates for test accuracy and side effects demonstrate that a more detailed and complex evidence assessment is not necessary.

A simple model structure may also be sufficient if robust estimates of quality-adjusted life years (QALY) gains and costs exist for both true and false negative and positive tests. For example, this can occur if the treatment pathway has already been studied or modelled extensively. If the studies match the situation being evaluated, only test accuracy needs to be explored further and cost effectiveness may be computed more easily.

14.2 Structuring the assessment

The scientific literature for diagnostics largely consists of studies of analytical and clinical validity. Data on the impact of diagnostic technologies on final patient outcomes are limited.

If data on the final patient outcomes of a diagnostic technology are not available, it may be necessary to combine the evidence from different parts of the care pathway. In this case the linkages between diagnosis, treatment and final outcomes need to be specified, and relevant data about those linkages needs to be obtained and reviewed.

Data about test accuracy and the nature of the care pathway and its outcomes can be used to create an assessment comparing the effect of different testing approaches.

14.2.1 Diagnostic technology performance

The performance of a diagnostic technology is assessed from a review of the test accuracy statistics and from studies of the direct effect of the test or the testing process on patient outcomes. Such direct effects can include adverse events or benefits from the test (for example, some biopsies may be curative), inconvenience to the patient, psychological outcomes (for example, relief or anxiety caused by the test results), and any psychological effects from being labelled as having a disease after a positive test result, even if it was a false positive ('labelling' effects).

Diagnostic test accuracy statistics measure the probability of correct and incorrect results from the test. The accuracy of test results determines whether correct diagnoses are made and therefore indirectly affects the final patient outcomes, which are the results of the care the patients receive based on the diagnoses.

In order to assess the diagnostic accuracy, the assumption is made that the patient undergoing the test being studied (index test) either has or does not have the condition of interest. This assumption is normally based on the results of a reference standard, but it is often not known for certain because of an imperfect reference standard.

Test accuracy data can be presented in a 2 x 2 table showing the results of the index test against the results of the reference standard. The data provided by a diagnostic test can be dichotomous, ordinal or continuous. Therefore, it may be necessary to select a cut-off point to reclassify ordinal and continuous data into dichotomous data to calculate test accuracy statistics.

14 – Modelling clinical outcomes and the cost–effectiveness of diagnostic technologies

Diagnostic test accuracy statistics represent the agreement between the index test and the reference test. The most commonly used are sensitivity and specificity; however, there are many different statistics that are commonly reported including positive and negative predictive values, likelihood ratios and odds ratios. Where there is the potential for variable cut-off points, data may be plotted on a ROC curve. Although the area under the ROC curve is often reported as a test statistic in studies, it is of little use for computing either test accuracy in practice or in test cost effectiveness. If tests are frequently indeterminate (either the test failed technically or the results can't be interpreted as either positive or negative) care must be taken in interpreting any reported test statistics to ensure the indeterminacy has been appropriately handled. Side effects associated with the test can be positive or negative. These are generally temporary and short in duration. However, the impact of these events may last longer than the events themselves, for example 'labelling' effects.

14.2.2 Impact of test results on the care pathway – the diagnostic process

The benefits from diagnostic testing generally arise from the results of treatment or prevention efforts that take place based on the testing. There may be some direct benefits from the knowledge gained and some direct harm from the testing, but most of the outcomes are indirect and come downstream. In order to assess these outcomes, consideration should be given not only to the diagnostic process itself, but also to treatment and monitoring.

A new diagnostic technology can affect the care pathway in two major ways. The first is how the test is used in the diagnostic process. The second is the impact of changed diagnostic information on subsequent disease management. A new technology can be a like-for-like replacement for an existing test or test sequence or it can be an addition to an existing test or test sequence. New diagnostics can be integrated together with parts of the existing diagnostic process to create a new sequence.

Tests can be given in sequence so that subsequent tests decrease the number of either false negative or false positive diagnoses, but one test sequence cannot simultaneously do both.

Often initial tests in a sequence have high sensitivity to triage those patients likely to have a condition, with subsequent tests with higher specificity used to limit the false positives. This is often done when the first test is either cheaper or less invasive than the subsequent test. A highly specific initial test may be given if a risky intervention is contemplated.

14 – Modelling clinical outcomes and the cost–effectiveness of diagnostic technologies

Diagnostic assessments should include all reasonable test sequences that that are likely to be considered by clinicians.

14.2.3 Impact of the intervention on the care pathway – disease management and health outcomes

Once the diagnostic process options are defined, the health outcomes from identified interventions or changes in intervention based on test results should be assessed. Often the intervention may be some form of treatment. The diagnostic technology may result in treatment being started, modified or stopped. Care should be taken to ensure the populations assessed in the studies of diagnostic test accuracy are comparable with those in the evaluation of the intervention.

14.3 Considerations for assessment and modelling

14.3.1 Direct outcomes of the diagnostic technology

There are several complexities to be considered when modelling direct outcomes of diagnostic technologies.

The number of technologies being evaluated

The assessment scope may include multiple technologies to be evaluated, as well as one or more comparators (see section 5). When modelling diagnostic test accuracy, all tests and test sequences identified in the final scope (technologies to be evaluated and comparators) need to be included.

Test side effects

Significant side effects and test preparation effects (for example, dietary modifications) may need to be considered for some tests. The temporary nature of some side effects means the timing of when they are measured may affect their magnitude. Sometimes these changes can be substantial and need to be considered and, if appropriate, incorporated into quality of life calculations.

Prognostic information

Prognostic information is information a test provides about future health events the patient can expect. Prognostic information does not drive treatment decisions and does not improve health outcomes directly other than through psychological effects. It can be difficult to quantify these benefits and they are not normally included in the base-case analysis.

14.3.2 Test accuracy

Issues that affect test accuracy should be considered when designing and carrying out the assessment. These issues can be quite complex and an attempt should be made to determine whether they are likely to affect the overall decision. If they are likely to affect the decision, then they should be fully investigated, at least with sensitivity analysis and further if possible.

Timing of tests

The timing of the test can influence outcomes. First, it can affect the accuracy of the test. Second, it can influence the effectiveness of treatment (for example, by delaying it or by providing opportunity for earlier intervention). Third, it can affect the patient directly by meaning they need additional visits to the GP's surgery or hospital. If differing test timings are likely, the different options should be modelled so that recommendations for test timing can be made if necessary.

Test sequences

More than one test is often used in making a diagnosis. A test may be repeated or a second test performed for several reasons:

- a second test in people identified as negative in the first test, to decrease the number of false negatives
- a second test in people identified as positive in the first test, to decrease the number of false positives
- a repeat because of uninterpretable results; this can occur if a test is unreadable or technical problems interfere with interpreting the test, such as laboratory errors or technical difficulties
- a repeat as part of a monitoring or screening programme; these would usually be planned repeat tests at intervals to detect changes over time.

Different sequences and timings of test sequences may need to be modelled as alternative test strategies.

Test correlation and conditional dependence

If tests are used in sequence, the correlation between the tests (particularly the correlation of test errors or conditional dependence) needs to be explored. As the correlation of the errors increases, the utility of the second test decreases. If there is perfect correlation then any misclassified patients will still be misclassified after the second test. In this case performing the second test will not add any value.

14 – Modelling clinical outcomes and the cost–effectiveness of diagnostic technologies

Test correlation data are rarely available so various assumptions on test correlation may need to be modelled.

Imperfect reference standard

The reference standard for a condition is often not perfectly accurate. Depending on the correlation between the technologies being evaluated and the reference standard, the resultant values of sensitivity and specificity can be either unduly high or low. If the correlations are known and the error rate for the reference standard is known, the estimates of sensitivity and specificity can be adjusted to correct for this systematic error. If not, sensitivity analysis needs to be undertaken to explore the ranges of the uncertainty.

Cut-off points

The diagnostic accuracy of a test depends on the threshold value used to distinguish positive results from negative results. If the threshold value is set to maximise sensitivity (that is, to maximise the number of patients with the disease who have positive test results), then the number of false positive results also usually increases (that is, specificity decreases). If the threshold is set so that specificity is increased (that is, to reduce the number of false positive tests), then sensitivity usually decreases. The accuracy at different cut-off points can be presented on a ROC curve, and each point on the ROC curve could be considered a different version of the test and evaluated separately.

The variability in sensitivity and specificity can stem from a human interpretation factor, which means that test performance can vary depending on the skill of the person administering or interpreting the test. Variability can also arise from mechanical differences between machines or from differences in laboratory techniques, which may result in a different effective cut-off point or operation below the ROC curve.

The optimal cut-off point is usually one of the decision problems explored in the assessment, but in some cases it may already be known based on routine clinical practice.

14.4 Identifying future research needs from the evidence

Future research needs may be identified by the External Assessment Group and included in the diagnostics assessment report. Candidate topics for future research can be identified from evidence gaps found during the systematic review and cost-effectiveness analysis. These may be best prioritised by considering the value of additional information in reducing the degree of decision uncertainty.

14 – Modelling clinical outcomes and the cost–effectiveness of diagnostic technologies

Part of the analysis of uncertainty is to identify the parameter and structural uncertainties to which the decision is most sensitive. This information can then be fed into decisions about future research priorities. Formal [value-of-information](#) methods may be useful in this context. These use probabilistic sensitivity analysis to establish the value for money that will be obtained from additional research to reduce parameter uncertainty, and how that research should be focused.

15. Cost-effectiveness evaluation – the reference case

This section describes the NICE approach to cost-effectiveness analysis. It covers:

- The concept and structure of the reference case (section 15.1)
- Modelling methods (section 15.2)
- Characterisation of potential bias and uncertainty (section 15.3)
- Presenting data and results (section 15.4)
- Analysis of data for patient subgroups (section 15.5)
- Reflecting equity considerations in cost-effectiveness analysis (section 15.6).

15.1 The concept and structure of the reference case

There is considerable debate about the most appropriate methods to use for some aspects of health technology assessment and these issues apply to the assessment of diagnostic technologies. This uncertainty relates to choices that are essentially value judgements; for example, whose preferences to use for valuation of health outcomes. It also includes methodological choices that relate to more technical aspects of an analysis; for example, the most appropriate approach to measuring health-related quality of life (HRQL). NICE has to make decisions across different technologies and disease areas. It is therefore crucial that analyses of clinical and cost effectiveness used in the evaluation adopt a consistent approach. To allow this, NICE has defined a 'reference case'.

The reference case specifies the methods NICE considers the most appropriate for the Diagnostics Advisory Committee's purpose, and consistent with an NHS objective of maximising health gain from limited resources. It does not preclude the Committee from considering non-reference-case analyses if appropriate.

There may be important barriers to applying reference-case methods. In these cases, the reasons for a failure to meet the reference case should be clearly specified and justified and the likely implications should, as far as possible, be quantified. The Committee should make a judgement on the weight it attaches to the results of a non-reference-case analysis.

The reference case includes a problem definition or scoping step. The outcomes of interest are health effects for patients or, when relevant, other people (principally carers). The reference-case perspective on outcomes is to maximise health gain from available healthcare resources.

15 – Cost–effectiveness evaluation – the reference case

Some features of diagnostic technologies affect ‘process characteristics’ and these may have health consequences; for example, the diagnostic technology may have health consequences by affecting the speed of correct diagnosis. If a diagnostic technology has significant characteristics that are independent of a direct effect on health, these should be noted. These characteristics include the convenience with which healthcare is provided and the level of information available for patients. The objective of NICE’s Diagnostics Assessment Programme is to offer guidance that represents an efficient use of available NHS and personal social services (PSS) resources. For this reason, the reference-case perspective on costs is that of the NHS and PSS (see section 12.4).

15.1.1 Type of economic evaluation

For the reference case, cost-effectiveness (specifically cost–utility) analysis is the preferred form of economic evaluation. This seeks to establish whether differences in costs between options can be justified in terms of health effects related to quality of life. Health-related quality of life changes should be expressed in terms of quality-adjusted life years (QALYs).

The focus on cost-effectiveness analysis is justified by the more extensive use and publication of these methods compared with cost-benefit analysis and the focus of NICE on maximising health gains from a fixed NHS/PSS budget. Given its widespread use, the QALY is considered to be the most appropriate generic measure of health benefit that reflects both mortality and HRQL effects. If the assumptions underlying QALYs (for example, constant proportional trade-off and additive independence between health states) are considered inappropriate in a particular case, evidence to this effect should be produced and analyses using alternative measures may be presented as an additional non-reference-case analysis.

15.1.2 Time horizon

The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect all important differences in costs or outcomes between the technologies being compared. In particular, it needs to be long enough to uncover the differences that result from outcomes related to treatments ordered because of the tests.

Some diagnostic technologies have effects on costs and outcomes over a patient’s life. In such instances, a lifetime time horizon for clinical and cost effectiveness is appropriate. A lifetime perspective is not used simply because the condition may last a lifetime; it is needed to incorporate a mortality component and quantify the implications of differential survival effects between alternative strategies. Analyses that limit the time horizon to periods

15 – Cost–effectiveness evaluation – the reference case

shorter than the expected full impact of the diagnostic test do not provide the best estimates of costs and benefits.

Modelling effects into the future usually needs extrapolation from shorter-term evidence. Therefore, sensitivity analyses should be performed on both the structural and data assumptions underlying that extrapolation.

Consideration of the time horizon and the uncertainty around extrapolating data beyond the duration of the clinical trials is a critical component of the evaluation.

15.1.3 Measuring and valuing health effects

For diagnostics assessments modelling is usually needed to measure and value health effects, because ‘end-to-end’ controlled trials with follow-up through the care pathway are uncommon. As discussed below the aim of the process is to evaluate outcomes in terms of QALYs. In some cases, HRQL and mortality data for patients with certain conditions may be directly available. However, in most cases the clinical outcomes captured in trials need to be converting by mapping them into QALYs.

The analysis should include all relevant patient outcomes that change in the care pathway as a result of the diagnostic test or sequence of tests. The nature, severity, time and frequency of occurrence, and the duration of the outcome may all be important in determining the impact on quality of life and should be considered as part of the modelling process.

For cost-effectiveness analysis, the value of health effects should be expressed in terms of QALYs for the appropriate time horizon. For the reference case, the measurement of changes in HRQL should be reported directly from patients and the value of changes in HRQL (that is, utilities) should be based on public preferences using a choice-based method – EQ-5D is a preferred measure in adults.

The EQ-5D is a health state instrument and a widely used measure of HRQL and has been validated in many different patient populations. A set of preference values elicited from a large UK population study using a choice-based method of valuation (the time trade-off method) is available for the EQ-5D classification system. This set of values can be applied to people’s self-reported descriptions of their HRQL to generate health-related utility values. The methods to elicit EQ-5D utility values should be fully described.

Data using the EQ-5D instrument may not always be available. If EQ-5D data are not available, or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D.

15 – Cost–effectiveness evaluation – the reference case

When EQ-5D data are not available, EQ-5D utility data can be estimated by mapping EQ-5D utility data from other HRQL measures included in the relevant study(ies) if an appropriate, validated mapping function is available. Mapping should use studies based on actual preferences of patients or potential patients, ideally rating using both instruments, and the statistical properties of the mapping function should be clearly described.

Another possibility, if EQ-5D utility data are not available, is to submit direct valuation of descriptions of health states based on standardised and validated HRQL measures included in the relevant clinical trial(s). In these cases, the valuation of descriptions should use the time trade-off method in a representative sample of the UK population, with ‘full health’ as the upper anchor, to retain methodological consistency with the methods used to value the EQ-5D.

Data that have been collected directly in relevant clinical trials using condition-specific, preference-based measures should be presented in a separate economic analysis.

The EQ-5D may not be an appropriate measure of health-related utility in all circumstances. For diagnostics, a new technology may be as accurate as the comparator, but may be less invasive, less painful, or quicker. The psychological effects of testing including anxiety, relief, or ‘labelling’ may also be difficult to quantify with the anxiety scale of the EQ-5D because it has a limited number of options. If the EQ-5D is considered inappropriate, empirical evidence (if available) should be provided on why the EQ-5D properties are not suitable for the particular patient population. If an alternative measure is preferred, the analysis should provide justification, supported by empirical data if possible, on the properties of the instrument used. It should also indicate any evidence that will help the Committee understand to what extent the choice of instrument affects the valuation of the QALYs gained.

The current version of the EQ-5D has not been designed for use in children. When necessary, consideration should be given to alternative standardised and validated preference-based measures of HRQL, such as the Health Utility Index 2 (HUI 2), that have been designed specifically for use in children.

The justification for choosing a particular data set should be clearly explained. Health-related utility data that do not meet the criteria for the reference case should be accompanied by a carefully detailed account of the methods used to generate the data and a consideration of how these methods may affect the values. If more than one plausible set of health-related utility data are available, a sensitivity analysis should be undertaken.

15.1.4 Evidence on resource use and costs

NHS and PSS costs

For the reference case, costs should relate to resources that are under the control of the NHS and PSS if it is possible to compare differential effects on costs between the technologies. These resources should be valued using the prices relevant to the NHS and PSS.

If the acquisition price paid for a resource varies significantly (for example, the diagnostic technology or consumables may be sold at reduced prices to NHS institutions), either the public list price or the lower price generally available to the NHS should be used in the reference-case analysis. Sensitivity analysis should assess the implications of variations from this price. Analyses based on price reductions for the NHS are considered only if the reduced prices are transparent and can be consistently available across the NHS, and if the period for which the specified price is available is guaranteed. In these circumstances, advice is taken from institutions such as the executive agency of the Office of Government Commerce (OGC) or Welsh Health Supplies.

Given the perspective in the reference case, it is appropriate for the financial costs relevant to the NHS/PSS to be used as the basis of costing, although these may not always reflect the full social opportunity cost of a given resource. As far as possible, estimates of unit costs and prices for particular resources should be used consistently across evaluations.

Diagnostic tests should generally be priced at average cost. The average cost should be based on the expected total use of the technology in the settings in which it would be installed. In some cases, if a device is already recommended for use for another purpose and sufficient spare capacity exists to allow the use for the condition envisioned in the current assessment, an analysis using [marginal costs](#) may be supplied in addition to the analysis based on average costs.

For devices with multiple uses, where only some uses are being evaluated, the average cost should initially be identified based on the expected usage or throughput of the device for only the uses being evaluated. Additional sensitivity analyses may be carried out using average costs computed through assigning some of the fixed costs to other uses of the device, if there is evidence that the other uses also provide good value for money.

If several alternative sources are available, a justification for the costs chosen should be provided and discrepancies between the sources explained. If appropriate, sensitivity analysis should be used to assess the implications for the results of using alternative data sources.

15 – Cost–effectiveness evaluation – the reference case

Costs related to the condition of interest and incurred in additional years of life gained as a result of treatment should be included in the reference-case analysis. Costs that are considered to be unrelated to the condition or technology of interest should be excluded. For diagnostic technologies, if the prognostic information generated allows cost savings in unrelated conditions, these offsets may be included in a non-reference-case analysis but must be explained and justified.

If introduction of the technology requires additional infrastructure to be put in place, these costs should be incorporated into the analysis, usually by inclusion in the average cost.

If a group of related technologies are being evaluated as part of a ‘class’ of treatments, an analysis should normally be presented in the reference case using the individual unit costs specific to each technology. Exceptionally, if the technologies can be justified as being represented as a class and there is a very wide range of technologies and costs to be considered, then analyses using the highest and lowest cost estimates can be presented.

Value added tax (VAT) should be excluded from all economic evaluations but included in budget impact calculations at the appropriate rate (currently 20%) if the resources in question are liable for this tax.

Non-NHS and non-PSS costs

Some technologies may have a substantial impact on the costs (or cost savings) to other government bodies. In these exceptional circumstances, costs to other government bodies may be included. If non-reference-case analyses include these broader costs, explicit methods of valuation are required. In all cases, these costs should be reported separately from NHS/PSS costs. These costs should not be combined into an [incremental cost-effectiveness ratio](#) (ICER; where the QALY is the outcome measure of interest).

Costs borne by patients may be included if they are reimbursed by the NHS or PSS. If the rate of reimbursement varies between patients or geographical regions, such costs should be averaged across all patients. Productivity costs and costs borne by patients that are not reimbursed by the NHS and PSS should be excluded from the reference-case analysis. If such costs may be a critical component of the value of the intervention, they should be included as additional information for the Committee to consider, but not as part of the reference-case analysis.

15.1.5 Discounting

Cost-effectiveness results should reflect the present value of the stream of costs and benefits accruing over the time horizon of the analysis. For the reference case, an annual discount rate of 3.5% should be used for both costs and benefits. The annual rate of 3.5%, based on the recommendations of the UK Treasury for the discounting of costs, should be applied to both costs and health effects. If results are potentially sensitive to the discount rate used, sensitivity analyses should be presented that use differential rates for costs and outcomes and/or that vary the rate between 0% and 6%.

15.2 Modelling methods

The models used to generate estimates of clinical and cost effectiveness for NICE's needs should follow accepted guidelines. Full documentation and justification of structural assumptions and data inputs should be provided. If there are alternative plausible assumptions and inputs, sensitivity analyses should be undertaken of the effects of the key assumptions on model outputs.

Modelling provides an important framework for synthesising available evidence and generating estimates of clinical and cost effectiveness in a format relevant to the Committee's decision-making process. Models are required for most evaluations. Situations when modelling is likely to be required include those in which:

- full end-to-end studies of diagnostics are not available, in which case modelling is used to estimate final outcome
- all the relevant evidence is not contained in a single trial
- patients in studies do not match the typical patients likely to use the technology in the NHS
- intermediate outcomes measures are used rather than effect on HRQL and survival
- relevant comparators have not been used, or studies do not include evidence on relevant subgroups
- the long-term costs and benefits of the technologies extend beyond trial follow-up.

It is not possible to provide an all-embracing definition of what constitutes a high-quality model, but some guidelines are available.

- In general, all structural assumptions should be fully justified, and data inputs should be clearly documented and justified in the context of a valid review of the alternatives. This is particularly important to avoid outlying values being selected that create a bias analogous to the selection bias

produced when using one or two clinical trials from a selection of several relevant trials.

- Estimates of treatment effect should be based on the results of the systematic review and modelling where appropriate. Modelling is often needed to extrapolate costs and health benefits over an extended time horizon.
- Assumptions used to extrapolate treatment effects should have clinical validity, be reported transparently and be clearly justified.
- Alternative scenarios should be considered to compare the implications of different assumptions around extrapolation for the results. For example, for the duration of treatment effects scenarios might include the treatment benefit in the extrapolated phase: being nil; being the same as during the treatment phase and continuing at the same level; or diminishing in the long term.

Study data may not be sufficient to quantify baseline risk of some health outcomes or events for the population of interest. Quantifying the baseline risk of health outcomes and how the disease would naturally progress with the comparator intervention can be a useful step when estimating absolute health outcomes in the economic analysis. Relative treatment effects observed in randomised trials may then be applied to data on the baseline risk of health outcomes for the populations or subgroups of interest. The methods used to identify and critically appraise sources of data for these estimates should be stated and justified.

The methods of quality assurance used in the development of the model should be detailed and the methods and results of model validation should be provided. In addition, the results from the analysis should be presented in a disaggregated format. This should include presenting information on estimates of life years gained, mortality rates (at separate time points if appropriate) and the frequency of selected clinical events predicted by the model.

15.3 Characterisation of potential bias and uncertainty

It is important to identify potential bias in the selection of inputs to the model and for the model to quantify the decision uncertainty associated with a technology (that is, the probability that a different decision would be reached if the true cost effectiveness of each technology could be ascertained before making the decision).

It is necessary to make assumptions when constructing a model. The potential bias and consequent uncertainty of these assumptions is sometimes referred to as ‘structural uncertainty’. Examples of structural uncertainty may include the categorisation of different states of health and the representation

of different pathways of care. These structural assumptions should be clearly documented and the evidence and rationale to support them provided. The impact of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of plausible scenarios.

A second type of potential bias arises from the selective use of data sources to provide values for the key parameters, such as different costs and utilities, or estimates of relative effectiveness and their longevity. The implications of different estimates of key parameters must be reflected in sensitivity analyses (for example, by including alternative scenarios). Inputs must be fully justified and uncertainty explored by sensitivity analysis using alternative input values.

A third source of uncertainty arises from parameter precision, once the most appropriate sources of information have been identified (that is, the uncertainty around the mean health and cost inputs in the model). Distributions should be assigned to characterise the uncertainty associated with the precision of mean parameter values. This uncertainty arises both from the basic statistical uncertainty of the parameter estimates and from any biases that may exist in the studies used to estimate the parameters. (See section 13.3 for a discussion of diagnostic study biases.) Probabilistic sensitivity analysis is the preferred method of investigating these uncertainties. This enables the uncertainty associated with parameters to be simultaneously reflected in the results of the model. In non-linear decision models, probabilistic methods provide the best estimates of mean costs and outcomes.

The mean value, distribution around the mean, and the source and rationale for the supporting evidence should be clearly described for each parameter included in the model. The distributions chosen for probabilistic sensitivity analysis should not be arbitrarily chosen – they should represent the available evidence on the parameter of interest, and their use should be justified. Formal elicitation methods are available if there is a lack of data to inform the mean value and associated distribution of a parameter. If alternative plausible distributions that could be used to represent uncertainty in parameter values are available, this should be explored by separate probabilistic analyses of these scenarios.

Accuracy parameters (usually sensitivity and specificity) present a special case. Because sensitivity and specificity are usually correlated and may vary based on how the test is used or interpreted, point estimates with distributions as described above are not usually appropriate. As discussed in section 13, ROC curves provide an appropriate way of presenting the relationship between sensitivity and specificity. Some methods of meta-analysis that provide summary ROC curves also provide confidence intervals around those

curves. Since part of the cost-effectiveness assessment includes determining the optimal point on the ROC curve, that point can then be the starting point for sensitivity analysis based on the confidence bounds. For probabilistic sensitivity analysis, a joint distribution over sensitivity and specificity would be required which may be computable from the meta-analysis outputs in some cases. Sometimes it may be appropriate to approximate this distribution in order to perform a probabilistic sensitivity analysis. The nature and the basis of the approximation should be documented.

Evidence about the extent of correlation between individual parameters should be carefully considered and reflected in the probabilistic analysis. Assumptions made about the correlations should be clearly presented.

The computational methods used to create an appropriate model structure may occasionally make it difficult to conduct probabilistic sensitivity analysis. The use of model structures that limit the feasibility of probabilistic sensitivity analysis should be clearly specified and justified. Models should always be fit for purpose, and should enable a thorough consideration of the decision uncertainty associated with the model structure and input parameters. The choice of a ‘preferred’ model structure or programming platform should not result in the failure to express uncertainty.

The level of effort in exploring uncertainty should usually be based on the level of decision uncertainty. Decisions may be robust in some cases within the range of likely variation of the parameters, and the amount of effort can be reduced. However, even if extensive modelling of uncertainty is not needed for the Committee to make a decision, in some cases additional modelling of uncertainty about test accuracy may help clinicians using the test.

15.4 Presenting data and results

15.4.1 Presenting data

All parameters used to estimate clinical and cost effectiveness should be presented clearly in tables and include details of data sources. For continuous variables, mean values should be presented and used in the analyses. For all variables, measures of precision should be detailed. For probabilistic analyses, the distributions used to characterise the uncertainty in input parameters should be documented and justified. As much detail as possible should be provided on the data used in the analysis.

15.4.2 Presenting expected cost-effectiveness results

The expected value of each component of cost and expected total costs should be presented. The main contributing components of expected QALYs

for each option compared in the analysis should also be detailed. ICERs should be calculated as appropriate.

The main individual components, comprising both costs and QALYs for the intervention and control treatment pathways, should be tabulated. For QALYs this includes presenting the life-year component separately. Consideration should also be given to presenting separately the costs and QALYs associated with different stages of the disease. Standard decision rules should be followed when combining costs and QALYs. These should reflect any situation in which [dominance](#) or extended dominance exists. ICERs reported must be the ratio of expected additional total cost to expected additional QALYs compared with alternative treatment(s). In addition to ICERs, expected net monetary or health benefits may be presented, using values of £20,000 and £30,000 for a QALY gained. If models consist of non-linear combinations of parameters, probabilistic sensitivity analysis should be used to generate mean costs and QALYs. In such models, setting parameters to their mean values will not provide the correct estimates of mean costs and QALYs.

15.4.3 Dealing with uncertainty around structural assumptions in cost-effectiveness analysis

Sensitivity analysis should be used to explore uncertainty around the key structural assumptions used in the analysis. Analysis of a representative range of plausible scenarios should be presented and each alternative analysis should present separate results.

An important element of uncertainty around cost-effectiveness results arises from the uncertainty in the structure of the decision model. The analysis of the uncertainty in all parameters for decision uncertainty assumes that factors such as a model's structure and data inputs are considered to be appropriate. However, these characteristics of the model are also subject to uncertainty, which should be identified and formally examined using sensitivity analysis.

Common examples of when this type of sensitivity analysis should be conducted are if there:

- is uncertainty about the most appropriate assumption to use for extrapolation of costs and outcomes beyond trial follow-up
- is uncertainty about how the pathway of care is most appropriately represented in the analysis
- may be economies of scale in the use of diagnostic technologies.

Uncertainty about the appropriateness of the methods used in the reference case can also be dealt with using sensitivity analysis, but these analyses must be presented separately.

15.4.4 Dealing with uncertainty around the selection of data sources in cost-effectiveness analysis

The uncertainty around the appropriate selection of data sources should be dealt with through sensitivity analysis. This includes uncertainty about the choice of sources for parameter values. Such sources of uncertainty should be explored through sensitivity analyses, preferably using probabilistic methods of analysis.

The choice of sources of key data to include in an analysis may not be clear cut. In such cases, the analysis should be re-run, using an alternative source of data or excluding the study over which there is doubt, and the results reported separately. Examples of when this type of scenario analysis should be conducted are if:

- alternative sets of plausible data are available on the health-related utility associated with the disease or intervention
- the cost of a particular resource or service differs between hospitals
- there are doubts about the quality or relevance of a particular study in a meta-analysis or mixed treatment comparison.

15.4.5 Additional factors

The report should include descriptions and analysis about additional factors that are not part of the reference case and that may be relevant for decision-making. These may include discussions of issues such as end-of-life costing, incremental improvements, system and process improvements and patient convenience and cost improvements.

15.5 Analysis of data for patient subgroups

For many technologies, the capacity to benefit from treatment differs for patients with differing characteristics. This should be explored as part of the reference-case analysis by providing estimates of clinical and cost effectiveness separately for each relevant subgroup of patients. The characteristics of patients in the subgroup should be clearly defined and preferably identified on the basis of an a priori expectation of differential clinical or cost effectiveness resulting from known, biologically plausible mechanisms, social characteristics or other clearly justified factors. If possible, potentially relevant subgroups are identified at the scoping stage, with consideration being given to the rationale for expecting a subgroup effect.

However, this does not preclude the identification of subgroups later in the process; in particular, during the deliberations of the Committee.

Given NICE's focus on maximising health gain from limited resources, it is important to consider how clinical and cost effectiveness may differ because of differing characteristics of patient populations. Typically, the capacity to benefit from treatment differs between patients, but this may also affect the subsequent cost of care. There should be a clear justification and, if appropriate, biological plausibility for the definition of the patient subgroup and the expectation of a differential effect. Post hoc data 'dredging' in search of subgroup effects is to be avoided and is viewed sceptically.

The estimate of the overall net treatment effect of an intervention is determined by the baseline risk of a particular condition or event and/or the relative effects stemming from the use of the technology compared with the alternatives and comparators. The overall net effect may also be determined by other features of the people comprising the population of interest. It is therefore likely that relevant subgroups may be identified in terms of differences in one or more contributors to absolute treatment effects.

For subgroups based on differences in their baseline risk of specific health outcomes, data to quantify this needs to be systematically identified. It is important that the methods for identifying appropriate baseline data for the subgroup analysis are provided in sufficient detail to enable replication and critical appraisal.

Care should be taken to specify how subgroup analyses are undertaken, including the choice of scale on which any effect modification is defined. The statistical precision of all subgroup estimates should be reflected in the analysis of parameter uncertainty. The characteristics of the patients associated with the subgroups presented should be clearly specified to allow the Committee to judge the appropriateness of the analysis with regard to the decision problem.

The standard subgroup analyses performed in randomised controlled trials (RCTs) or systematic reviews are often based on differences in relative effects (through the analysis of interactions between the effectiveness of the technology and patient characteristics). The possibility of differences emerging by chance, particularly when numerous subgroups are reported, should be explored.

In considering subgroup analyses, the Committee takes specific note of the biological or clinical plausibility of a subgroup effect in addition to the strength of the evidence in favour of such an effect. The evidence supporting biological

15 – Cost–effectiveness evaluation – the reference case

or clinical plausibility for a subgroup effect should be fully documented, including details of statistical analysis.

Individual patient data are preferred, if available, for estimating subgroup-specific parameters. However, as for all evidence, the appropriateness of such data is always assessed by considering factors such as the quality of the analysis, the representativeness of the available evidence and relevance to the decision problem.

In some circumstances it may be appropriate to consider subgroups based on differential cost; for example, if the cost of managing a particular complication of treatment is known to be different in a specific subgroup.

The Committee pays particular attention to its obligations with respect to legislation on human rights, discrimination and equality when considering subgroups.

Types of subgroups that are not considered relevant are those based solely on:

- individual utilities for health states and patient preference
- differential treatment costs for individuals according to their social characteristics
- the costs of providing treatment in different parts of the UK (for example, if the costs of facilities available for providing the technology vary according to location).

15.6 Reflecting equity considerations in cost-effectiveness analysis

In the reference case, an additional QALY should receive the same weight regardless of any other characteristics of the people receiving the health benefit.

The estimation of QALYs, as defined in the reference case, implies a particular position regarding the comparison of health gained between individual patients. Therefore, an additional QALY is of equal value regardless of other characteristics of the patients, such as their sociodemographic details, or their pre- or post-treatment level of health. There are several unresolved methodological issues concerning how and in what circumstances to apply additional weights to QALY calculations. Until such issues are resolved, the use of differential QALY weights is not recommended as part of the reference case.

16. Development of recommendations by the Diagnostics Advisory Committee

The Diagnostics Advisory Committee (DAC) develops recommendations following the process outlined in section 7. The section below details the considerations relevant to their decision-making. This section includes:

- Committee review of the evidence (section 16.1)
- Developing recommendations (section 16.2)
- Types of recommendation (section 16.3)
- Framework for research recommendations (section 16.4).

16.1 Committee review of the evidence

The Committee reviews the evidence contained in the documentation. The Committee has the discretion to take account of the full range of studies that have been carried out and is not expected to restrict itself to consideration of only certain categories of evidence. The Committee is required to consider all of the evidence it deems relevant, from randomised controlled trials (RCTs) to observational studies, and any qualitative evidence related to the experiences of patients, carers and clinical experts who have used the technology being evaluated or are familiar with the relevant conditions and patient groups. If sufficient direct data on outcomes stemming from the diagnostic interventions are not available, the Committee evaluates indirect evidence and models of the care pathway. In evaluating the evidence base, the Committee exercises its scientific and clinical judgement when deciding whether particular forms of evidence are suitable for answering specific questions.

The importance given to the various kinds of evidence depends on both the overall balance and quality of the evidence from different sources, and the suitability of a particular type of evidence to address the issues under consideration. In general, greater importance is given to evidence derived from high-quality studies that are designed to minimise bias.

The Committee considers the evidence on:

- diagnostic test accuracy
- clinical effectiveness
- cost effectiveness.

16.1.1 Evaluating diagnostic test accuracy

Diagnostic test accuracy is not a direct contributor to cost effectiveness, but information on the test accuracy is an important tool for clinicians. Because of their usefulness to clinicians, the summary results on test accuracy may be

included in the draft and final guidance along with estimates of clinical and cost effectiveness. The DAC, at its discretion, may review the analysis of the accuracy statistics provided in the diagnostics assessment report to examine the validity and inclusiveness of the underlying data, meta-analytic techniques used, the selection of cut-off points, and the resultant uncertainties generated.

16.1.2 Evaluating clinical effectiveness

The DAC has the discretion to take account of the full range of clinical studies and modelling that have been carried out and is not expected to restrict itself to consideration of only certain categories of evidence. The recommendations developed by the DAC take account of the level of uncertainty surrounding the underlying evidence base.

The DAC's judgements on clinical effectiveness take account of the following factors:

- The nature and quality of the evidence derived from:
 - the analysis of the External Assessment Group
 - the written comments of the registered stakeholders
 - the experience of the specialist Committee members, particularly of the use of the technology in clinical practice
 - the views of the lay members of the Committee (both standing and specialist members) on patients' experiences during and following the use of the technology
- Uncertainty generated by the evidence and differences between evidence gained in research conditions and that relating to effectiveness in clinical practice
- The possible differential benefits or greater risk of adverse effects in different groups of patients
- The risks (adverse effects) and benefits of the technology as seen from the patient's perspective
- The position of the technology in the overall pathway of care and the available alternative treatments.

The extent to which the above factors are taken into account in making judgements about the evidence of clinical effectiveness is a matter for the Committee's discretion.

16.1.4 Evaluating cost effectiveness

NICE is asked to take account of the overall resources available to the NHS when determining cost effectiveness. Therefore, decisions on the cost effectiveness of a new technology must include judgements on the

implications for healthcare programmes for other patient groups that may be displaced by the adoption of the new technology.

The potential budget impact of adopting a new technology does not determine the DAC's decision. The Committee does take account of how its advice may enable the more efficient use of available healthcare resources. In general, the Committee will want to be increasingly certain of the cost effectiveness of a technology as the impact of the adoption of the technology on NHS resources increases. Therefore, the Committee may require more robust evidence on the effectiveness and cost effectiveness of technologies that are expected to have a large impact on NHS resources.

The DAC takes account of how the incremental cost effectiveness of the technology being evaluated relates to other interventions and technologies currently being applied in the NHS. In addition, as far as possible, the Committee will want to ensure that its judgements regarding the cost-effective use of NHS resources are consistently applied between evaluations.

The Committee has to make judgements on the appropriateness and relevance of comparator technologies because this is crucial to the consideration of the cost-effectiveness evidence.

If the evidence on key parameters used to estimate cost effectiveness (for example, clinical effectiveness and effect on health-related quality of life) has serious limitations and/or a variety of assumptions have been necessary in the cost-effectiveness modelling, the additional uncertainty this generates is a key factor in underpinning the judgements of the Committee. The Committee is aware that the evidence base is often weak for diagnostic technologies. Taking this into account, the DAC is still likely to consider more favourably technologies for which evidence on cost effectiveness is underpinned by the best-quality clinical data than those for which supporting evidence is dependent to a large extent on theoretical modelling alone.

The Committee's judgements on cost effectiveness are influenced by:

- the strength of the supporting evidence on impact on patient health outcomes
- the robustness and appropriateness of the structure of the care pathway and economic models; in particular, it considers carefully whether the model reflects the decision problem at hand and the uncertainties around the assumptions on which the model structure is based
- the plausibility of the inputs into, and the assumptions made, in the economic models
- its own evaluation of the modelling approach, taking into account all of the economic evidence submitted

- the range and plausibility of the incremental cost-effectiveness ratios (ICERs) generated by the models reviewed
- the likelihood of decision error and its consequences.

The DAC considers carefully which patients benefit most from the technology and whether there are subgroups of patients for whom the effectiveness evidence suggests differential cost effectiveness. The DAC may recommend the use of an intervention for subgroups of the population only if there is clear evidence that the characteristics defining the subgroup influence the effectiveness and/or cost effectiveness of the intervention.

The DAC does not use a precise ICER threshold above which a technology would automatically be defined as not cost effective or below which it would. Given the fixed budget of the NHS, the appropriate threshold to be considered is that of the opportunity cost of programmes displaced by new, more costly technologies. NICE does not have complete information about the costs and QALYs from all competing healthcare programmes in order to define a precise threshold. However, NICE considers that it is most appropriate to use a threshold range as described below. Furthermore, the cost effectiveness of a technology is a necessary, but is not the sole, basis for decision-making. Consequently, NICE considers technologies in relation to this threshold range, and the influence of other factors on the decision to recommend a technology is greater if the ICER is closer to the top of the range.

Below a most plausible ICER of £20,000 per QALY gained, the decision to recommend the use of a technology is normally based on the cost-effectiveness estimate and the acceptability of a technology as an effective use of NHS resources. If the estimated ICERs presented are less than £20,000 per QALY gained and the Committee judges that particular interventions should not be provided by the NHS, the recommendations make specific reference to the Committee's view on the plausibility of the inputs to the economic modelling and/or the certainty around the estimated ICER. This might be affected, for example, by sensitivity analysis or limitations to the applicability of findings regarding effectiveness.

Above a most plausible ICER of £20,000 per QALY gained, judgements about the acceptability of the technology as an effective use of NHS resources specifically take account of the following factors:

- the degree of certainty around the ICER; in particular, the Committee is more cautious about recommending a technology if it is less certain about the ICERs presented
- whether there are strong reasons to indicate that the assessment of the change in HRQL has been inadequately captured, and may therefore misrepresent the health utility gained

- the innovative nature of the technology, specifically if the innovation adds demonstrable and distinctive benefits of a substantial nature that may not have been adequately captured in the QALY measure.

As the ICER of an intervention increases in the £20,000 to £30,000 range, the Committee's judgement about the acceptability of the technology as an effective use of NHS resources makes explicit reference to the relevant factors listed above.

Above a most plausible ICER of £30,000 per QALY gained, the Committee needs to identify an increasingly strong case for supporting the technology as an effective use of NHS resources, with regard to the factors listed above.

NICE has a strong preference for expressing health gains in terms of QALYs. In most circumstances, when the health gain is expressed in terms of life-years gained, the range of most plausible 'life-years gained' ICERs that are acceptable is substantially lower than those described above. In these circumstances, the Committee imputes a plausible QALY value from the estimated life-years gained. The exact adjustment that the Committee makes takes account of the differences between QALYs and life-years gained. It is guided by reference to the population norms for HRQL for the affected population. In general, however, patients with a disease or condition have lower QALYs on average than the norms for the overall population.

16.2 Developing recommendations

After reviewing the evidence the Committee agrees draft recommendations on the use of the technology in the NHS in England. When formulating these recommendations, the Committee has discretion to consider those factors it believes are most appropriate to the evaluation. In doing so, the Committee has regard to any relevant provisions of NICE's Directions, set out by the Secretary of State for Health, and legislation on human rights, discrimination and equality. In undertaking evaluations of healthcare technologies, NICE takes into account:

- the broad balance of clinical benefits and costs
- the degree of clinical need of patients under consideration
- any guidance issued to the NHS by the Secretary of State that is specifically drawn to the attention of NICE by the Secretary of State, and any guidance issued by the Secretary of State
- the potential for long-term benefits to the NHS of innovation.

The Committee takes into account advice from NICE on the approach it should take to making scientific and social value judgements. Advice on social value judgements is informed in part by the work of NICE's Citizens Council.

Guidelines that describe the social value judgements that should generally be considered by the Committee are provided in, ‘Social value judgements: principles for the development of NICE guidance’ (www.nice.org.uk/aboutnice/howwework/socialvaluejudgements/socialvaluejudgements.jsp).

The Committee takes into account how its judgements have a bearing on distributive justice or legal requirements in relation to human rights, discrimination and equality. Such characteristics include, but are not confined to: race, gender, disability, religion or belief, sexual orientation, gender reassignment and pregnancy or maternity.

The Committee considers the application of other Board-approved NICE methods policies, such as the supplementary guidance on discounting and the end-of-life criteria, if they are relevant to the evaluation.

Because the Programme often evaluates new technologies that have a thin evidence base, in formulating its recommendations the Committee balances the quality and quantity of evidence with the expected value of the technology to the NHS and the public.

The credibility of the guidance produced by NICE depends on the transparency of the DAC’s decision-making process. It is crucial that the DAC’s decisions are explained clearly, and that the contributions of registered stakeholders and the views of members of the public are considered. The reasoning behind the Committee’s recommendations is explained, with reference to the factors that have been taken into account.

The language and style used in the documents produced by the Committee are governed by the following principles:

- Clarity is essential in explaining how the DAC has come to its conclusions.
- The text of the documents does not need to reiterate all the factual information that can be found in the information published alongside the guidance. This needs careful judgement so that enough information and justification is given in the recommendations to enable the reader to understand what evidence the DAC considered and, if appropriate, who provided that evidence.

The Committee may take into account factors that may provide benefits to the NHS or the population, such as patient convenience. It may also consider costs and other positive or negative impacts on the NHS that may not be captured in the reference-case cost analysis, such as improved processes.

The Committee is not able to make recommendations on the pricing of technologies to the NHS.

16.3 Types of recommendation

Recommendations may have a variety of formats depending on the circumstances.

If there is sufficient evidence to demonstrate the technology's cost effectiveness, the Committee makes a recommendation for use of the technology. Recommendations for use of a diagnostic test, or use of a diagnostic test as an option, may be limited to specific circumstances such as: the patient's characteristics, aetiology of the disease, the training and skills of those providing the test, availability of equipment, and the availability of other portions of the care pathway. In some cases adoption recommendations may be made on the basis that additional research is performed as the technology is adopted.

If there is not sufficient evidence to determine the cost effectiveness of the technology, the Committee may make various types of recommendation, such as a recommendation for use only in research or, in particular circumstances, combined with a recommendation for further research. The Committee's recommendations depend on factors such as the quality of the evidence and the degree of risk, both in terms of cost and patient outcomes associated with the use of the technology. The rationale for their recommendations is outlined in the 'considerations' section of the guidance.

If there is sufficient evidence to demonstrate the technology is not cost effective, the Committee does not recommend it for use. This recommendation may be for either general or specific circumstances. If, on the basis of expert advice or ongoing research, the Committee considers that the technology has the potential to be of benefit to the NHS in the foreseeable future, it may decide to not recommend the technology for use *at the present time* instead of a general recommendation against its use

If there is considerable uncertainty about the cost effectiveness of a technology, the Committee may consider issuing a recommendation that highlights the importance of gaining additional information. The factors it may consider include: the costs and benefits of the additional research; the probability of the research affecting future use; the non-recoverable investment costs of early implementation; the losses in patient benefits from delaying adoption to await research; the probability that the uncertainty will resolve over time; and the impact of adoption recommendations on the feasibility of doing the research

As described in section 12.2, a technology may have multiple uses and not all of these may be explored in the evaluation. In this case the Committee formulates recommendations only for the uses of the technology described in the scope.

The Committee's recommendations may take into account that the technology has already been purchased and its recommendations are made in the context of additional use of existing equipment.

When the Committee makes research recommendations it follows the framework described in section 16.4.

The diagnostics guidance document describes the degree of uncertainty on which the Committee's recommendations are based, and the potential impact of such uncertainties.

16.4 Framework for research recommendations

The Committee develops research recommendations using the principles described in NICE's Research Recommendations Manual, available on the NICE website

(www.nice.org.uk/media/FC2/5E/ResearchRecommendationManual.pdf).

When making a research recommendation, the Committee aims to:

- describe the most important clinical, economic, technical or other evidence gaps relating to use of the technology in the NHS
- state the research questions that future studies need to address.

These recommendations may include recommendations for research about the care pathway after the use of the diagnostic test if uncertainties about the pathway affect the value of testing.

The Committee considers the following factors when deciding whether to recommend future evidence generation and data collection:

- the most important evidence gaps relating to the uncertainty about the technology, and the [value of information](#) that could be derived from generating evidence to address them
- information about ongoing or planned research on the technology
- ethical and/or practical aspects of conducting further research.

These considerations aim to help guide decisions about investment in future research by prioritising the studies that will address research questions and generate new evidence of greatest value to population health.

PART IV: APPENDICES AND REFERENCES

Appendix A: Glossary

Accuracy – see Test accuracy

Aetiology

The origin or cause of the condition or disease under consideration.

Alternative technology

A test or technology that performs similar or overlapping functions to the notified technology but that is not in common or recommended practice and is not a potential comparator technology.

Applicability

How well the results of a study or NICE evidence review can answer a clinical question or be applied to the population being considered. This term is similar to generalisability and external validity, which are also used in the literature.

Assessment – see Evidence assessment

Assessment protocol

The assessment protocol is derived from the scope of the assessment, taking into account comments from organisations attending the scoping workshop. It forms the basis of the diagnostics assessment report.

Audit tool

Criteria and data collection tools to improve patient care and clinical practice by helping clinical services to compare current practice against NICE guidance.

Bias

Systematic (as opposed to random) deviation of the results of a study from the 'true' results.

Care pathway

This usually refers to the sequence of practices, procedures and treatments that should be used with people with a particular condition. The aim is to improve the quality of care.

Clinical effectiveness

The extent to which a specific treatment or intervention, when used under usual or typical conditions, has a beneficial effect on the course or outcome of

disease compared with no treatment or other routine care. Clinical effectiveness is not the same as efficacy.

Companion diagnostic technology

A diagnostic technology that identifies people who are likely to benefit from a specific therapy for their condition. It may also help in stratifying disease status, selecting the proper medication and tailoring dosages to patients' needs. In some cases, the use of companion diagnostic technologies may be necessary to comply with the licensed indications of pharmaceuticals.

Comparator

The technology or technologies that are most commonly used or are recommended in current NICE guidance for the functions in the evaluation.

Cost effectiveness

Value for money. A test or treatment is said to be 'cost effective' if it leads to better health than would otherwise be achieved by using the resources in other ways.

Cost-effectiveness analysis

Cost-effectiveness analysis is one of the tools used to carry out an economic evaluation. The benefits are expressed in non-monetary terms related to health, such as symptom-free days, heart attacks avoided, deaths avoided or life years gained (that is, the number of years by which life is extended as a result of the intervention).

Costing tool

A tool developed by NICE to accompany guidance, which helps healthcare organisations determine the cost of implementing the guidance locally.

Cut-off point

The sensitivity and specificity pair from a receiver operating characteristic (ROC) curve that represents what should be used in practice.

Diagnostic technology

A medical technology used to gain information about a person's condition or future condition. It can be used for a variety of purposes including diagnosis, screening, monitoring or providing prognostic information

Diagnostics assessment report (DAR)

A report prepared by the External Assessment Group, based on a systematic review of the clinical and health economic literature including data supplied by the sponsor or sponsors. The report includes modelling as appropriate to estimate health outcomes and cost effectiveness.

Diagnostics assessment report (DAR) addendum

Corrections, clarifications or additional analysis (if any) undertaken by the External Assessment Group following the Committee meeting at which the diagnostics consultation document is agreed. The Committee considers this additional information when it meets to agree the diagnostics guidance document.

Diagnostics consultation document (DCD)

Draft guidance developed from the Diagnostics Advisory Committee's draft recommendations about using a diagnostic technology (or group of similar technologies) in the NHS.

Diagnostics guidance

NICE guidance about the adoption and use of a diagnostic technology.

Diagnostics guidance document (DGD)

The final guidance document from the evaluation process. This document is published on the NICE website and represents official NICE guidance.

Dominance

A test is dominated if another test has equal or greater sensitivity and specificity and lower or equal costs and adverse events. The dominated test should be worse on at least one criterion. A test that dominates other tests under discussion can be called dominant.

Dominant – see Dominance

Effectiveness

How beneficial a test or treatment is under usual or everyday conditions, compared with doing nothing or opting for another type of care.

Efficacy

How beneficial a test, treatment or public health intervention is under ideal conditions (for example, in a laboratory), compared with doing nothing or opting for another type of care.

EQ-5D

A standardised five-dimensional instrument used to measure health outcomes. It is completed by the responder themselves and is quick to use.

Evaluation

In this document, 'evaluation' is used to mean the process of developing diagnostics guidance on the use of diagnostic technologies within the NHS in England.

Evidence assessment (or assessment)

The process or result of reviewing the evidence about a topic and creating a report covering clinical and cost effectiveness. This process may include systematic reviews, meta-analysis, modelling and other evidence gathering or creation activities.

Evidence overview

A document that summarises the findings from the evidence and modelling reported for a diagnostic technology assessment. It is used to inform the Diagnostics Advisory Committee about the technology so that the Committee can then agree the draft recommendations.

Expert elicitation

Obtaining subjective information from experts about important parameters of interest. These can include probabilities, probability distributions, or magnitudes and distributions of the parameters. Formal approaches are often used.

External Assessment Group

An independent group of researchers commissioned by NETSCC on behalf of NICE to review the evidence on diagnostic technologies. The External Assessment Group includes researchers who assess the quality of studies on diagnostic technologies, and health economists who look at whether the technology is good value for money. The Diagnostics Advisory Committee bases its discussions on the diagnostics assessment report produced by the External Assessment Group.

False negative

Errors in screening that mean that not all patients with a condition are identified as having it.

False positive

Errors in screening that mean that some patients without a condition are incorrectly identified as having it.

Guidance Executive

NICE directors who approve all NICE guidance for publication.

Heterogeneity

Used in meta-analyses and systematic reviews to describe if the results or estimates of effects of a treatment from separate studies seem to be very different (for example, the size of treatment effects may vary across studies, or some studies may indicate beneficial treatment effects whereas others suggest adverse treatment effects). Such differences in results may occur by chance, because of variation in study quality, or because of variation in populations, interventions, or methods of outcome measurement in the included studies.

Incremental cost-effectiveness ratio (ICER)

The incremental cost-effectiveness ratio is a useful way of expressing cost effectiveness that compares change in costs with change in effects.

Indication

A sign, symptom or other condition that leads to the diagnostic process.

Lead-time bias

A bias in screening and monitoring studies, in which the screened population appears to have longer survival simply because the disease is caught earlier in its natural progression even if no actual survival benefit exists.

Length bias

A bias in screening and monitoring studies, in which more aggressive disease processes are less likely to be detected because they are more likely to move to observable signs and symptoms between screening episodes. The result makes the screened group appear to have inappropriately greater benefits than an unscreened group.

Likelihood ratio

There are two likelihood ratios, LR+ and LR-. LR+ is the ratio computed by dividing the true positive rate by the false positive rate or the sensitivity by 1 –

Appendix A – Glossary

specificity. LR– is computed by dividing the false negative rate by the true negative rate or $1 - \text{sensitivity}$ divided by the specificity.

Marginal cost

The additional cost for each use of a technology. It excludes any fixed costs.

Medical technology

A medical technology is any instrument, apparatus, appliance, material or other article, whether used alone or in combination, including the software necessary for its application, intended to:

- diagnose, prevent, monitor, treat or alleviate disease
- diagnose, monitor, treat, alleviate or compensate for an injury or disability
- investigate, replace or modify the anatomy or a physiological process
- control conception.

Meta-analysis

Results from a collection of independent studies (investigating the same treatment) are pooled, using statistical techniques to synthesise their findings into a single estimate of a treatment effect. If studies are not compatible, for example because of differences in the study populations or in the outcomes measured, it may be inappropriate or even misleading to statistically pool results in this way.

Negative predictive value

The proportion of patients who have a negative test result who also do not have the condition of interest.

Notified technology

The technology routed for evaluation to the Diagnostics Assessment Programme by the Medical Technologies Advisory Committee.

Odds ratio

The odds (the number of positives divided by the number of negatives) in one group divided by the odds in another group.

Pathogenicity

The ability to cause disease, usually the probability of a genetic variation resulting in disease.

Patient and carer organisation

Organisations of patients, carers, communities and other lay members, including those that represent people from groups protected by equalities legislation.

Patient outcomes

The health outcomes to patients from following the care pathway. These outcomes usually result from the treatment that follows the use of a diagnostic technology. These outcomes include benefits and harms from either the diagnostic technology or the treatment.

Positive predictive value

The proportion of patients who have a positive test result who also have the condition of interest.

Post diagnostic care pathway

The portion of the care pathway that occurs after the diagnostic test is used.

Prior probability

The prior probability of a model state is the computed probability with which it will occur based on existing data. When new data about the model are collected, the probability is revised and this adjusted figure is called the 'posterior probability'.

Product sponsor

The manufacturer, developer, distributor or agent of the technology or technologies being evaluated. Manufacturers of comparative technologies are not considered to be product sponsors.

Quality-adjusted life years (QALYs)

A measure of health outcome that looks at both length of life and quality of life. QALYs are calculated by estimating the years of life remaining for a patient following a particular care pathway and weighting each year with a quality of life score (on a 0 to 1 scale). One QALY is equal to 1 year of life in perfect health, or 2 years at 50% health, and so on.

Receiver operating characteristic (ROC) curve

A graphical plot of true positive rate (sensitivity) against false positive rate (1 – specificity). ROC analysis may help to discriminate between good and bad tests and is a useful tool for differentiating the diagnostic accuracy of

different tests or test sequences. A cut-off point or cut point on a ROC curve is a point on the curve (that is, a single sensitivity/specificity pair) at which the technology is evaluated, actually used or recommended for use.

Registered stakeholder

An organisation with an interest in a topic on which NICE is developing diagnostics guidance. Stakeholders may be:

- product sponsors (manufacturers, developers, distributors or agents) of diagnostic technologies
- national patient and carer organisations
- NHS organisations
- national organisations representing healthcare professionals.

Scope

Document created at the start of producing a piece of guidance outlining what the guidance will and will not cover. It provides a detailed framework for the evaluation and defines the disease, the patients, the technologies, the outcomes, and the costs that will be covered by the evaluation, as well as the questions the evaluation aims to address. The final version of the scope is used as a starting point for developing the guidance.

Scoping workshop

The scoping workshop is a meeting held to help define the scope of an evaluation. Its attendees include product sponsors, registered stakeholders, the External Assessment Group and NICE staff.

Sensitivity

In diagnostic testing, sensitivity refers to the chance of having a positive test result if you have the disease; 100% sensitivity means that all those with the disease will test positive, but this is not the same the other way around. A patient could have a positive test result but not have the disease – this is called a 'false-positive'. To fully judge the accuracy of a test, its specificity must also be considered.

Sensitivity analysis

A form of modelling that evaluates the impact of alternative values for some of the model parameters. Often used when there is significant uncertainty about the value of the parameter. This has nothing to do with test sensitivity defined above.

Specificity

In diagnostic testing, specificity refers to the chance of having a negative test result if you do not have the disease; 100% specificity means that all those without the disease will test negative, but this is not the same the other way around. A patient could have a negative test result yet still have the disease – this is called a 'false-negative'. To fully judge the accuracy of a test, its sensitivity must also be considered.

Sponsor – see Product sponsor

Test accuracy

Any measure relating to the correctness of a test, such as sensitivity, specificity, predictive values, and the proportion of results that are correct.

Topic lead

The member of the staff at NICE responsible for the topic. The topic lead writes the scope and overview documents and drafts the guidance documents. The topic lead is the primary interface with the External Assessment Group and primary contact for all technical issues.

UK National Screening Committee

The UK National Screening Committee assesses the evidence for screening programmes against a set of internationally recognised criteria covering the condition, the test, the treatment options and the effectiveness and acceptability of the programme.

Value of information

The value of information from additional research based on any net value resulting from improved decision-making.

Appendix B: Members of diagnostics methods working group

Dr Phil Alderson, NICE

Ms Selma Audi, Boston Scientific

Dr Ian Barnes, Department of Health

Dr Hanan Bell, NICE

Dr Meindert Boysen, NICE

Professor Alan Brennan, SchARR

Ms Jennifer Butt, NICE

Professor Martin Buxton, Health Economics Research Group, Brunel University

Dr Kalipso Chalkidou, NICE

Mr Ravi Chana, Roche Diagnostics

Dr Helen Chung, NICE

Ms Carole Cohen, Edwards Lifesciences

Dr Nick Crabb, NICE

Professor Jon Deeks, Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham

Ms Jill Dhell, Department of Health R&D

Ms Eleanor Donegan, NICE

Dr Sarah Garner, NICE

Dr Elisabeth George, NICE

Professor Paul Glasziou, Department of Primary Healthcare, University of Oxford

Mr Franz Hessel, Abbott Diagnostics

Professor Sue Hill, Diagnostics Programme Board, Department of Health

Dr Chris Hyde, Professor of Public Health and Clinical Epidemiology, Peninsula College of Medicine and Dentistry, University of Exeter

Appendix B – Members of diagnostics methods working group

Dr Anthony James, NHS Institute for Innovation and Improvement

Mr Gurleen Jhuti, NICE

Professor Peter Littlejohns, NICE

Dr Myfanwy Lloyd Jones, SchARR

Professor Carole Longson (Chair), NICE

Ms Sandra Lopes, ABHI

Dr Jo Lord, Health Economics Research Group, Brunel University

Dr Susanne Ludgate, MHRA

Dr Georgios Lyratzopoulos, NICE

Dr Anne Mackie, UK National Screening Committee

Ms Mirella Marlow, NICE

Mr Baish Naidoo, NICE

Ms Frances Nixon, NICE

Mrs Laura Norburn, NICE

Mr David Owolabi, Roche Diagnostics

Ms Seren Phillips, NICE

Ms Toni Price, NICE

Dr Craig Ramsey, College of Life Sciences and Medicine, University of Aberdeen

Mr Francis Ruiz, NICE

Mr Mark Samuels, Roche Diagnostics

Professor Mark Sculpher, Team for Economic Evaluation and Health Technology Assessment, University of York

Ms Tarang Sharma, NICE

Professor Andrew Stevens, Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham

Appendix B – Members of diagnostics methods working group

Dr Matt Stevenson, SchARR

Mr Matthew Stork, AXrEM

Dr Rod Taylor, Peninsula Technology Assessment Group, University of Exeter

Ms Victoria Thomas, NICE

Ms Rebecca Trowman, NICE

Professor Tom Walley, NIHR Health Technology Assessment Programme

Appendix C: Process timeline

Weeks			
0	<p>NICE contacts sponsor of referred topic and requests initial information. NICE identifies stakeholders to participate in the diagnostic evaluation and sends invitation to participate. NICE initiates specialist Committee member recruitment</p>	<p>NICE undertakes care pathway research and develops detailed draft scope, making use of available clinical expertise</p>	<p>If potential alternative technologies are identified, NICE invites their sponsors to join the process</p>
1			
2			
3			
4			
5			
6			
7			
8			
9	<p>NICE holds scoping workshop then revises the scope</p>		
10	<p>Specialist Committee members selected & assessment subgroup appointed</p>		
11	<p>Assessment subgroup and External Assessment Group meet to review revised scope and discuss assessment protocol</p>		
12	<p>NICE asks sponsors and specialist Committee members for relevant data on behalf of the External Assessment Group. The External Assessment Group carries out the assessment</p>		
13			
14		<p>Final scope, final assessment protocol, list of specialist Committee members and list of registered stakeholders published on NICE website</p>	
15			
16			
17			
18			

Appendix C – Process timeline

19		
20		
21		
22		
23		
24		
25		
26		
27		(continued)
28		NICE asks sponsors and specialist Committee members for relevant data on behalf of the External Assessment Group. The External Assessment Group carries out the assessment
29		
30		
31		
32		
33		
34		
35		
36	NICE receives the final diagnostics assessment report (DAR). NICE distributes the DAR to registered stakeholders for comment (10 working days)	
37		
38		
39	NICE sends the DAR, registered stakeholders' comments on the DAR, and the evidence overview to the Diagnostics Advisory Committee (DAC)	
40		
41		
42	Diagnostics Advisory Committee meets to agree draft recommendations	

Appendix C – Process timeline

43	
44	
45	Diagnostics consultation document (DCD) is agreed. DCD consultation starts for reg'd stakeholders
46	DCD public consultation starts
47	
48	DCD consultation ends
49	
50	DAC meets to review public consultation comments & agree final recommendations
51	NICE finalises the diagnostics guidance document (DGD)
52	
53	
54	NICE Guidance Executive approves DGD for publication, subject to resolution
55	Resolution period starts
56	
57	Resolution period ends, if there are no resolution requests. (If resolution requests are made, the timeline to final publication is extended until resolution is agreed.)
58	
59	
60	
61	
62	
63	NICE publishes diagnostics guidance

Appendix D: References

Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Medical Test Reviews (in process):

www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=558&pageaction=displayproduct

Claxton K, Palmer S, Longworth L, et al. (2011) Uncertainty, evidence and irrecoverable costs: informing approval, pricing and research decisions for health technologies. University of York CHE research paper 69

Cochrane handbook for DTA reviews (in process):

srdta.cochrane.org/handbook-dta-reviews

Cochrane handbook for systematic reviews of interventions:

www.cochrane.org/training/cochrane-handbook

Deeks J (2001) Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 323: 157–62

Dukic V, Gatsonis C (2003) Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics* 59: 936–46

Hamza TH, Arends LR, van Houwelingen HC, et al. (2009) Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology* 9: 73–87

Harbord RM, Deeks JJ, Egger M, et al. (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8: 239–51

Litlenberg B, Moses LE (1993) Estimating Diagnostic Accuracy from Multiple Conflicting Reports. *Medical Decision Making* 13: 313–21

Medical Services Advisory Committee (2005) Guidelines for the assessment of diagnostic technologies.

[www.health.gov.au/internet/msac/publishing.nsf/Content/D81BE529B98B3DB6CA2575AD0082FD1B/\\$File/Diag%20Guidelines%20Sept%202005%20updated%2021%20may%202007.pdf](http://www.health.gov.au/internet/msac/publishing.nsf/Content/D81BE529B98B3DB6CA2575AD0082FD1B/$File/Diag%20Guidelines%20Sept%202005%20updated%2021%20may%202007.pdf)

NICE (2008) Guide to the methods of technology appraisal.

www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guidetothemethodsoftechnologyappraisal.jsp?domedia=1&mid=B52851A3-19B9-E0B5-D48284D172BD8459

Appendix D – References

NICE (2011) Medical Technologies Evaluation Programme methods guide. www.nice.org.uk/media/4E1/09/MedicalTechnologiesEvaluationProgrammeMethodsGuide.pdf

Reitsma JB, Glas AS, Rutjes AW, et al. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 58: 982–90

Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 20:2865–84

Sica GT (2006) Biases in Research Studies. *Radiology* 238: 780–9

Sutton AJ (2008) Integration of Meta-analysis and Economic Decision Modeling for Evaluating Diagnostic Tests. *Medical Decision Making* 28: 650–67

Whiting PF, Rutjes AWS, Reitsma JB, et al. (2004) Sources of Variation and Bias in Studies of Diagnostic Accuracy. *Annals of Internal Medicine* 140: 189–202

Whiting PF, Sterne JAC, Westwood ME, et al. (2008) Graphical presentation of diagnostic information. *BMC Medical Research Methodology* 8: 20