Guidelines for preparing assessments for the Medical Services Advisory Committee

*Application form for consideration of a health technology by MSAC
*PICO Confirmation for a health technology considered by PASC
*Applicant developed assessment report (ADAR)
*Department contracted assessment report (DCAR)
*Commentary on an ADAR

Version 1.0

May 2021

Contents

Abbreviations and acronyms

Term	Definition
ADAR	applicant developed assessment report
ARTG	Australian Register of Therapeutic Goods
AUROC	area under the receiver operating characteristics curve
CCA	cost-consequences analysis
CEA	cost-effectiveness analysis
COS	core outcome set
CT	computed tomography
CUA	cost-utility analysis
DCAR	department contracted assessment report
EMSN	Extended Medicare Safety Net
ESC	Evaluation Sub-Committee
FDG	fluorodeoxyglucose
GPG	Greatest Permissible Gap
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HSROC	hierarchical summary receiver operating characteristic
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio
IV	intravenous
MAUI	multiattribute utility instrument
MBS	Medicare Benefits Schedule
MCID	minimal clinically important difference
MSAC	Medical Services Advisory Committee
NPV	negative predictive value
PASC	PICO Advisory Sub-Committee
PBAC	Pharmaceutical Benefits Advisory Committee
PBS	Pharmaceutical Benefits Scheme
PCOM	person (or patient) centred outcome measure
PET/CT	positron emission tomography/computed tomography
PICO	population, intervention, comparator, outcome
PPV	positive predictive value
PROM	patient reported outcome measure
PSA	probabilistic sensitivity analysis
PSM	proposed surrogate measures
QALY	quality-adjusted life year

Term	Definition
RCT	randomised controlled trial
SaMD	software as a medical device
TCO	target clinical outcomes
TG	Technical Guidance
TGA	Therapeutic Goods Administration

Preamble

These guidelines have been developed to provide advice to applicants and assessment groups on the health technology assessment (HTA) methods that are used throughout the Medical Services Advisory Committee (MSAC) assessment pathway. The guidelines may also give consumers and stakeholders an insight into the technical aspects of generating evidence for an assessment report; a summary of the process is also available on the MSAC website.[a] The guidelines apply to all requests for public funding that fall within MSAC's remit, including Medicare Benefits Schedule (MBS) services, national screening programs, blood products for the National Product List, and technologies that may be funded via other mechanisms. The current guidelines are to be used for all requests for new public funding (i.e. first-time applications and subsequent reconsiderations).

The MSAC guidelines and associated templates provide advice on analysing and presenting clinical and economic evidence. MSAC also considers other sources of evidence, including consumer-based evidence and input. Consumers are important in MSAC considerations. MSAC is committed to understanding consumer values, experiences and preferences, and integrating them into its consideration of a health technology. Consumer-based evidence and input provide additional context when interpreting the clinical and economic evidence, and the processes for providing such evidence to MSAC are described on the MSAC website.[b]

Further information on the processes for preparing and assessing requests for public funding via the MSAC pathway, as well as application forms and assessment report templates, are available on the MSAC website.[c] For ease of use, the application forms and templates are cross-referenced to these guidelines.

Purpose and roles of MSAC

MSAC is a nonstatutory committee established by the Australian Minister for Health in 1998. MSAC's role is to provide recommendations regarding funding of health technologies other than medicines. These health technologies include (but are not limited to) medical services (new and existing), other programs (blood products or screening programs), highly specialised therapies delivered as state-based services, and services provided with prostheses.

MSAC's recommendations are based on the strength of evidence about the comparative safety, comparative effectiveness, cost-effectiveness and total cost of the proposed health technology, and encompass:

- whether public funding should be supported for the health technology and, if so, the circumstances under which public funding should be supported
- the proposed MBS item descriptor and fee for the service where funding through the MBS is supported
- other matters related to the public funding of health services referred by the Minister for Health.

a www.msac.gov.au/internet/msac/publishing.nsf/Content/Factsheet-06

b www.msac.gov.au

c www.msac.gov.au

The recommendations are made to:

- the Minister for Health
- the Health Council
- Jurisdictional Blood Committee/National Blood Authority
- other committees as relevant.

The advice provided by MSAC can result in new items or amended items on the MBS, and may inform funding decisions for programs such as the national blood arrangements or National Blood Authority Immunoglobulin Governance Program.

There is no obligation on government to accept or implement the advice MSAC provides.

Membership of MSAC

MSAC is an independent expert committee comprising individuals from the fields of clinical medicine, health economics and consumer matters. The Minister for Health determines the size and composition of MSAC. Members are drawn from a wide range of experts, constituted from time to time to address the likely type of applications for the committee's consideration. The current membership of MSAC is available on the MSAC website.[a]

MSAC subcommittees

MSAC currently has 2 subcommittees: the PICO Advisory Sub-Committee (PASC) and the Evaluation Sub-Committee (ESC). MSAC also has an Executive Committee (made up of the chairs of MSAC, ESC and PASC, and the Deputy Chair of MSAC) to manage MSAC activities between formal committee meetings.

Department of Health

The Australian Government Department of Health (referred to in these guidelines as 'the Department') provides the Secretariat function for MSAC and its subcommittees. Applications for public funding are received by the Department, who refers applications to MSAC for independent advice. The Department is also responsible for:

- triaging which applications are suitable for MSAC assessment, contracting an assessment group to develop the PICO (population, intervention, comparator and outcomes) confirmation, coordinating communication between the applicants and assessment group, and providing policy and clinical advice
- organising, coordinating and covering the costs of developing a department contracted assessment report, liaising with an assessment group regarding requirements and timeframes
- organising, coordinating and covering the costs of a commentary to evaluate an applicant developed assessment report, liaising with an assessment group regarding requirements and timeframes.

a www.msac.gov.au

Overview of MSAC outcomes

The decision options available to MSAC are to recommend funding (including the circumstances of funding), defer their decision, or reject funding and/or recommend research (such as through the Medical Research Future Fund). MSAC is not able to recommend any interim funding.

In the event of a deferral or rejection, MSAC recommends the pathway for subsequent reconsideration (e.g. via ESC or directly to MSAC).

MSAC will also recommend a review period for some health technologies. This includes the expected timeframe for the review, the data requirements and the organisation responsible for submitting the data.

Further details regarding MSAC processes and pathways are addressed in guidance documents available on the MSAC website.[a]

Key factors influencing decision-making by MSAC

MSAC provides advice on the circumstances under which health technologies should be funded, subsidised or made available in the Australian health care system. Typically, this advice is to the Minister for Health and relates to the listing of a health technology on the MBS. However, MSAC's remit also includes providing advice for other funding arrangements.

In its considerations, MSAC is primarily informed by the strength and quality of the evidence for the following quantifiable factors:

- clinical need of patients and nature of the condition (Section 1)
- comparative health gain – assessed in terms of the magnitude and clinical importance of effect, and including both the effectiveness and the safety of the health technology (Section 2)
- comparative cost-effectiveness – results derived typically from a cost-effectiveness or cost-utility analysis (presented as incremental cost-effectiveness ratios), or from a cost-minimisation approach (Section 3)
- predicted use in practice and financial impact – presented as the projected annual cost per year to the Australian Government and/or to other funding bodies as relevant to the application (Section 4).

The impact of health technologies on the Australian population may not be limited to quantifiable impacts on health. For this reason, MSAC decision-making is also informed by additional, less-readily quantifiable factors:

- Equity – The advice to subsidise a health technology may have an impact on the equitable distribution of the health technology or health resources across different groups, such as those categorised by age, socioeconomic status or geographical location (Technical Guidance 29).
- Value of knowing – This refers to the effects derived from the use of a health technology that may not be characterised by improvements in health, for example, the potential benefits and harms associated with a knowledge of a prognosis or diagnosis (Technical Guidance 28).
- Presence of effective alternatives – This helps to determine the clinical need for the health technology (Technical Guidance 29.7).

a www.msac.gov.au

- Other relevant considerations – These include the impact on organisations, or the way in which organisational issues may create barriers or facilitators to the uptake of the new technology or efficiency of health care delivery, ethical concerns, and social aspects (Technical Guidance 29).

In making its decision, MSAC considers the best available evidence. This includes evidence provided in assessment reports, by experts or informed by consumer engagement. It also takes into account the likelihood of further information becoming available should a decision be deferred. For example, for technologies aimed at patients with rare diseases, the generation of evidence is more difficult than for common diseases; therefore, the willingness to make a decision in the face of uncertainty is greater. This is consistent with the principles of the National Strategic Action Plan for Rare Diseases.[1]

Purpose of the guidelines

These guidelines have been developed by the Department to assist in completing:

- an application form for consideration of a medical service/ technology by MSAC
- a PICO confirmation for a medical service/ technology considered by PASC
- an applicant developed assessment report (ADAR)
- a department contracted assessment report (DCAR)
- a commentary on an ADAR.

An assessment report is a document that captures the technical details relevant to the assessment of a technology for consideration by MSAC.

MSAC decision-making is informed by a range of factors, and an assessment report is not intended to capture all these factors. The MSAC process involves multiple inputs, of which the assessment report is the primary source for technical, typically quantifiable, evidence.

The technical components of an assessment of a health technology include:

- a clinical conclusion based on evidence of comparative health impacts
- an estimate of cost-effectiveness, as informed by the clinical conclusion
- an estimate of the utilisation of a technology, and the financial impact for the Australian Government or funder.

These are the core elements of the assessment report and of the guidelines, and reflect the key information relevant to MSAC decision-making. However, the assessment report also provides information on other relevant information that may inform MSAC decision-making.

These guidelines do not contain guidance on the incorporation of the views and perspectives of consumers, patients or members of the public obtained through public consultation activities. Such activities for the purpose of informing MSAC deliberations occur through other mechanisms that are detailed on the MSAC website. However, these guidelines do contain guidance on when published evidence of such data may be used to inform an assessment of other relevant considerations.

Navigating the guidelines

The types of technologies considered by MSAC are varied, but are broadly categorised into therapeutic technologies (such as consultative services, interventions and devices) and investigative technologies (such as medical tests).

For an assessment of a therapeutic technology, the relevant clinical TG subsections are in Section 2A (Technical Guidance 6 to Technical Guidance 8) and all the included subsections will be required for an assessment.

For an assessment of an investigative technology, the relevant clinical TG subsections are in Section 2B (Technical Guidance 9 to Technical Guidance 16). Not all TG subsections in Section 2B will be relevant for the assessment of a test. The relevant TG subsections will depend on the nature of the test and the available evidence, and are informed by an assessment framework (Technical Guidance 9). A guide to the relevant clinical TG subsections for assessing an investigative technology is presented at the beginning of Section 2B in Figure 6.

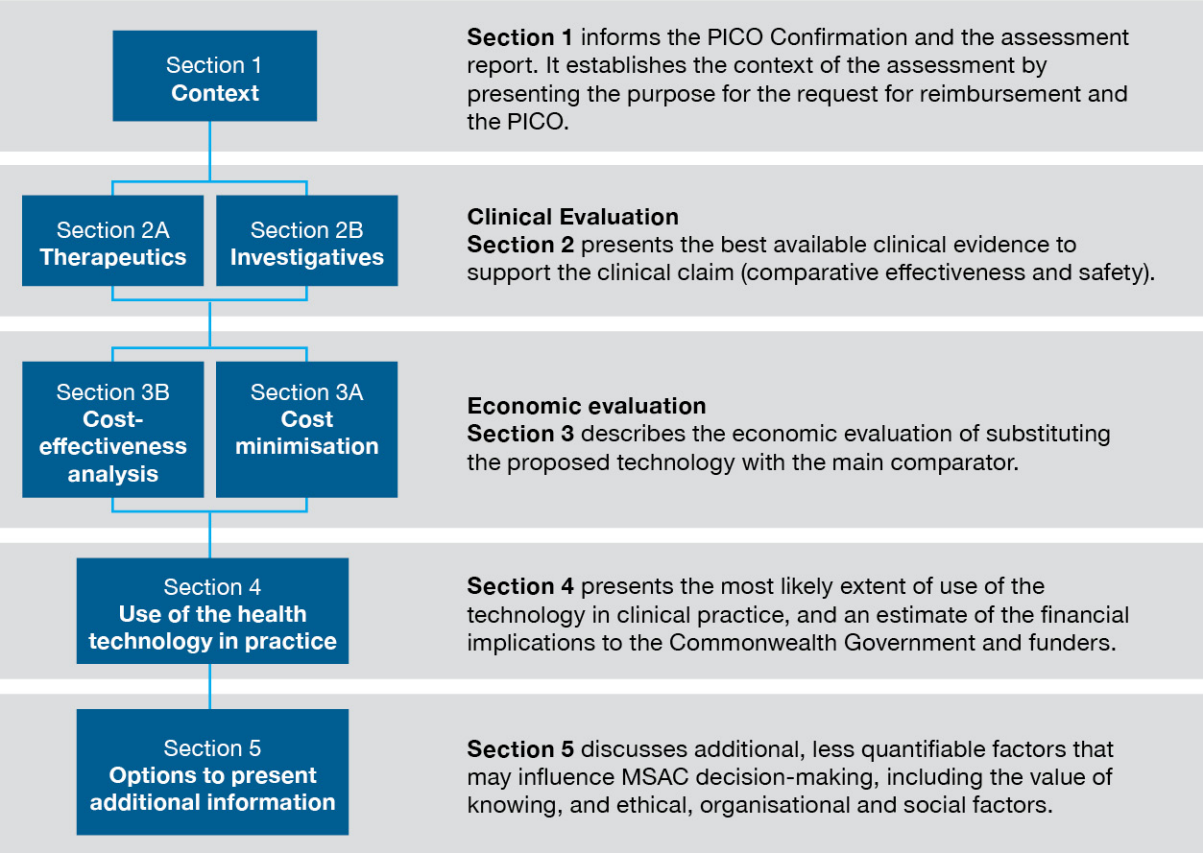


Figure 1 **Sections of the guidelines**

Presenting an assessment report

The information provided in these guidelines is intended to inform the creation of a PICO confirmation and an assessment report for a health technology. The appropriate formats for presentation of the PICO confirmation and assessment report are provided in templates on the MSAC website.[a]

a www.msac.gov.au

MSAC reconsiderations of a health technology

Health technologies that are not recommended for funding may be reconsidered by MSAC if new evidence can be provided to address the concerns raised by MSAC in response to the previous assessment report. The history of MSAC considerations for a health technology should be tabulated as described in Technical Guidance 4. The information requests in the guidelines should be followed for new information provided in subsequent assessment reports; however, information that is not in dispute should not be re-presented. Present a brief list of the components of the assessment report that are not in dispute, and delete sections of the template that are not required.

Care should be taken when presenting new information that alters the interpretation of results from previous assessment reports. If it is necessary to present a combination of old and new results so that new conclusions can be understood, consider having the text that MSAC has already seen in grey, and any new text in black.

An assessment report for the purpose of a reconsideration must clearly present and discuss how the new information addresses the main matters of concern to MSAC.

Glossary of terms used in the guidelines

This glossary defines terms used within the guidelines. Additional terminology is defined in the HTA advisory committee glossary.[a] This glossary may be partly or wholly incorporated in an updated version of the HTA glossary.

Term	Definition
Algorithm, clinical management	The set of possible clinical management options for a defined population over time, presented according to the subpopulations that receive each option. Often presented in simplified form as a flow diagram, or in more precise form as a decision analysis.
Assessment framework	The analytic framework or logic diagram that is used to illustrate the necessary steps that link the use of an investigative technology (commonly a test) in the target population and the consequences that this may have on health outcome gains.
Assessment questions	The questions addressed by the HTA to inform the overall public funding question.
Biomarker	A characteristic (usually measured by a test) by which a pathological or physiological process (e.g. disease, response to treatment) can be identified. A biomarker may be defined by the presence or absence of a characteristic, or it may be defined by a quantity of a parameter above or below a specified threshold.
Clinical utility	The net health benefit/harm derived from an investigative health technology across all those tested (including true positives, false positives, true negatives and false negatives).

a www.pbs.gov.au/info/industry/useful-resources/glossary

Term	Definition
Clinical utility standard	The test and method of interpretation used to allocate patients to alternative options in the key clinical studies generating direct evidence of health outcome gains. This replaces the previous term of 'evidentiary standard'. Test methodology involves test reagent(s)/kit, test platform, biospecimen type and preparation, what is tested (e.g. gene sequence, gene expression, staining, whether all cells or only some cells), body part(s) scanned, tracer used, qualitative or semi-quantitative interpretation of the image, and definition of the test/scan threshold result that differentiates between different clinical management actions (e.g. eligible for the targeted therapy or not).
Device	Used to imply 'medical device' Any instrument, apparatus, appliance, material or other article that: • is used for humans • is intended to diagnose, prevent, monitor, treat or alleviate a disease or injury, or modify or monitor anatomy or physiological functions of the body • generally achieves its purpose by a physical, mechanical or chemical action.
Direct from test to health outcomes evidence	Compare with linked evidence. Evidence showing the impact the test has on health outcomes.
Direct randomised trial	Compare with indirect comparison. A trial in which participants are randomly allocated to groups that receive either the proposed health technology or its main comparator.
Exchangeability	*Compare with transitivity*. An assessment of exchangeability in an indirect comparison or network meta-analysis considers whether there are any differences in the distribution of any characteristics across the relevant clinical studies that may confound the results of the comparison.
Exemplar	*Compare with facilitated*. The combination of intervention and population for which sufficient evidence is likely to be available for MSAC to decide its advice on public funding.
Facilitated	*Compare with exemplar*. A combination of intervention(s) and population(s) that is close enough to the exemplar to not require a full HTA. Instead, MSAC could decide its advice on public funding by accepting sufficient similarities between the facilitated combination(s) and the exemplar combination.
Germline variant	A gene change in a body's reproductive cell (egg or sperm) that becomes incorporated into the DNA of every cell in the body of the offspring. Germline variants are passed on from parents to offspring at the time of conception.
Health technology	A technology used in a health care system, for example, therapeutic services (such as medicines, procedures, blood products), medical devices, investigative medical services (such as diagnostic tests, imaging services, population screening tests), equipment and supplies, organisational and managerial systems, and programs of health delivery. For the purposes of some definitions in these guidelines, particularly in relation to existing health technologies, this usual definition is extended to include any medical service, placebo or watchful waiting instead of an active health technology. For ease of reading, the word 'technology' is used throughout the document, but applies to all types of technology or services.
Indirect comparison	*Compare with direct randomised trial*. An analysis that indirectly compares the proposed health technology to its main comparator by comparing one set of trials, in which participants were randomised to receive the proposed health technology or a common reference, with another set of trials, in which participants were randomised to receive the main comparator or the common reference.

Term	Definition
Investigative technology	A type of health technology that is claimed to generate clinically relevant information about the individual to whom the service is rendered. To achieve an improvement in health outcomes, this information must result in a change in the clinical management of an intermediate intervention. In this sense, investigative procedures can only indirectly improve health outcomes. Examples of investigative technologies are imaging, pathology, genetic testing, and clinical assessments for diagnosis, prognosis, staging, monitoring, prediction of treatment response, surveillance and cascade testing (also referred to as cascade screening). For ease of reading, the word 'test' is used throughout the document as an alternative term for 'investigative technology', but is intended to reflect the broad range of investigative technologies available.
Linked evidence	*Compare with direct from test to health outcomes evidence*. When evidence from studies of test accuracy is linked to evidence of change in management and evidence of treatment effectiveness to derive an estimate of the clinical utility of the test.
Machine learning algorithms	Mathematical models built on training data that are used to discover structure within data and/or to predict an output.
Multifactorial algorithms	Algorithms that combine multiple factors to determine a person's risk of a future event. Algorithms may be static (learning occurs before the dissemination of the technology) or dynamic (learning continues to occur after the dissemination of the technology).
Number needed to test	Number of people that need to be tested for one person to undergo the intended change in clinical management.
Penetrance	The proportion of individuals for whom traits or characteristics associated with a particular genetic variant will manifest in the phenotype within a specified period of time.
Proband	An individual (index case) in a family who is affected with a heritable disease or condition and has a relevant known pathogenic germline variant.
Somatic variant	A gene change that occurs after conception in non-germline cells, which is neither inherited nor passed on to offspring.
Standard, clinical reference	*Compare with standard, non-clinical reference*. A reference standard that detects a clinical disorder or clinical outcome of interest.
Standard, non-clinical reference	*Compare with standard, clinical reference*. A reference standard that detects a biomarker, parameter or analyte.
Streamlined	An abbreviated HTA approach used for facilitated population and intervention combinations.
Technology	A simplified term for all types of health technology or services.
Test	A simplified term for investigative health technology.
Test, cascade	A test of family members of a proband for the identified germline variant.
Test, diagnostic	A test used to inform or identify a disease, condition or injury.
Test, monitoring	A test used to observe a disease, condition or parameter over time.
Test, predictive	*Compare with test, prognostic*. A test that estimates differences in the proportions of individuals in a tested population who develop a disease or experience a clinical event over time according to different test results (such as test positive and test negative) if clinical management changes in response to one or more of these different test results. This is also known as treatment effect modification.

Term	Definition
Test, prognostic	*Compare with test, predictive*. A test that estimates differences in the proportions of individuals in a tested population who develop a disease or experience a clinical event over time according to different test results (such as test positive and test negative) without altering clinical management.
Test, screening	A test used to detect disease, abnormalities or associated risk factors in asymptomatic members of a population at risk.
Test, staging	A test used to classify the severity of a disease.
Test, triage	A test used to determine which patients require further tests.
Therapeutic technology	A type of health technology that is claimed to directly improve the health of people receiving it. Nothing else needs to be rendered to achieve the improvement in health outcomes. Examples of therapeutic technologies are devices, medicines, vaccines, procedures, blood products, programs or systems.
Transitivity	*Compare with exchangeability*. An assessment of transitivity in an indirect comparison or network meta-analysis considers whether there are important differences in the distribution of known treatment effect modifiers across the relevant clinical studies that are likely to confound the results of a comparison.
Value of knowing	Any consequence for the wellbeing of a patient, or their family members or carers, that does not arise from changes in health outcomes quantified in the clinical utility evidence. This includes concepts of benefits and harms of knowing, the benefits and harms of naming, and the impact on a patient's or family members'/carers' wellbeing through being able to plan nonhealth resources that come at a cost to the person, and sometimes funders, such as transport, accommodation, education, community care (of self and others, including dependants), equipment, income loss and insurance.

Section 1 Context

Introduction

This section provides guidance for establishing the context for an assessment of a health service/ technology. This includes describing the purpose of the application, developing the population, intervention, comparator and outcomes (PICO) criteria for use of the technology and the associated assessment questions, and justifying the proposed MBS descriptor and fee (where applicable). The structure of Section 1 is outlined in Flowchart 1.

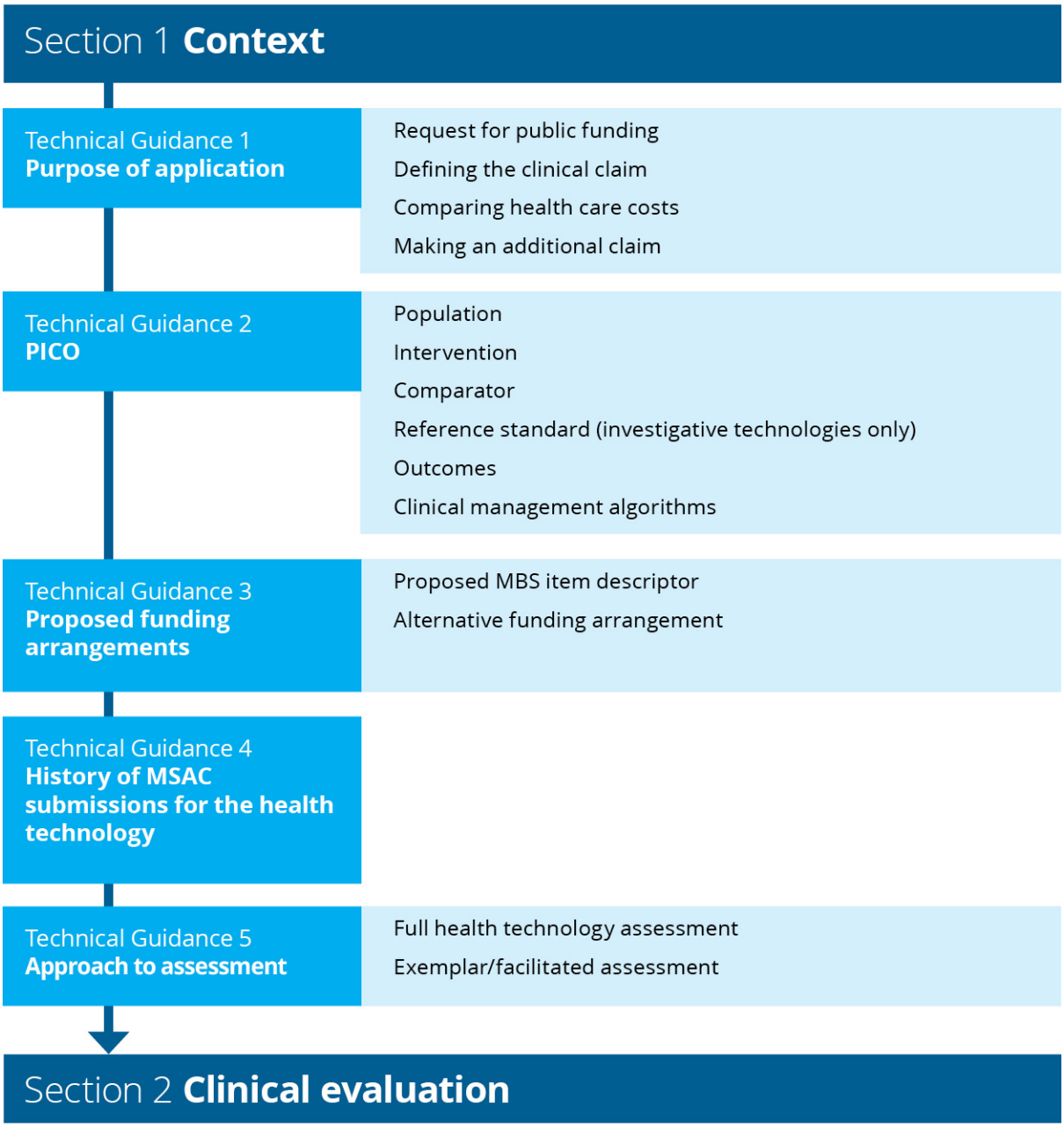
The MSAC application process may involve different pathways. For most applications that progress to an assessment report, the PICO Advisory Sub-Committee (PASC) first considers a PICO confirmation to focus the assessment report and ensure it is clinically relevant. In some circumstances (e.g. reconsiderations), a PICO confirmation may not be considered by PASC before the development of an assessment report.

Regardless of whether PASC formally considers a PICO confirmation, the development of the PICO and assessment questions is a pivotal part of the development of an assessment report. The information in this section will:

- act as a reference for the details that may be required for an application form
- provide guidance for preparing a PICO confirmation
- assist with establishing the PICO in an assessment report where no PICO confirmation has been required
- act as a reference for HTA groups performing a critical appraisal of an assessment report
- provide guidance on how to justify changes to an agreed PICO should the assessment report need to deviate from the PICO confirmation ratified by PASC.

Preparing an assessment report

In general, if an agreed PICO confirmation is available, this will be reflected in Section 1 of the assessment report. The assessment report should not deviate from the agreed PICO confirmation. If changes to the agreed PICO confirmation items must be made, both the agreed PICO confirmation item and the variation should be presented. The need for the new approach should be justified, noting that any deviations from the agreed PICO confirmation may affect confidence in the applicability of the evidence and/or analyses presented.



Flowchart 1 **Structure of the guidance for the context**

Technical Guidance 1 Purpose of application

KEY CONSIDERATIONS

- **Summarise the request for public funding (TG 1.1).**
- **Provide a claim based on the safety and effectiveness of the proposed health technology compared with the main comparator (TG 1.2).**

For investigative technologies

- **If the proposed technology is unlikely to be considered cost-effective on the basis of health gains, it may be possible to make a claim based on the additional value of knowing derived from the proposed investigative technology compared with the main comparator (TG 1.4).**

TG 1.1 Request for public funding

A clear purpose for the application is necessary to enable meaningful interpretation of the evidence presented in an assessment report. The application must be presented with reference to the patient population in whom use of the health technology is intended. The purpose for the application should be precise and include:

1. a clear definition of the proposed patient population
2. the intended use and outcomes of administering the proposed health technology, including a clear description of the technology and the proposed purpose of the technology (in the intended patient population)
3. a clinical claim for the proposed health technology in terms of its impact on health outcomes, that is, whether use of the technology results in superior effectiveness and/or safety compared to current management of the same condition, or whether it is claimed to be noninferior
4. a statement justifying the need for the health technology if there is no claimed health advantage, which may occur when
   a. the medical service involves an investigative technology that is claimed to have other benefits for an individual, family members or carers (Technical Guidance 28 – Value of knowing)
   b. the medical service involving the technology produces noninferior health outcomes but has additional practical and/or cost advantages over current clinical management
5. a statement clarifying whether funding is sought under the MBS or another funding source
6. a statement indicating whether other applications relating to the proposed health technology are in progress, for example, an application to the Therapeutic Goods Administration (TGA) or a Prostheses List application.

TG 1.2 Defining the clinical claim

The advice provided by MSAC is primarily based on the clinical effectiveness and safety, and the cost-effectiveness of a health technology compared with current practice. These are referred to as comparative clinical effectiveness and safey, and comparative cost-effectiveness, respectively. Acceptable cost-effectiveness means that the additional health or other benefits derived from the health technology are sufficient to justify any additional costs associated with use of the technology. Given MSAC's role and remit, MSAC determinations are more influenced by the health outcomes and health care spending associated with the technology within the health care system than impacts outside of the health care system.

In some circumstances, non-health-related outcomes (e.g. patient preferences and organisational issues that will affect implementation and use) may provide additional context for MSAC's decision-making. For more information on other relevant considerations, see Technical Guidance 28 and Technical Guidance 29.

Clinical claims for health technologies

The aim of most health technologies is to have a positive impact on the health of patients. A service involving a technology that is intended to replace an existing service will usually have a positive or neutral impact on health. In some circumstances, health technologies will result in a loss of health and will require the consideration of other factors for MSAC to make a favourable funding recommendation.

Health outcomes in this context refers to the aggregate of all patient-relevant health outcomes (positive and negative) – that is, the net clinical benefit. Separate claims may be reasonable for outcomes typically considered to measure *effectiveness*, and for outcomes that measure *safety* or *patient harms*. However, it is important to consider the net benefit of the technology in terms of its effectiveness and safety.

Appropriate clinical claims for health outcomes are:

- The use of the proposed technology results in superior health outcomes compared to the comparator/standard practice.
- The use of the proposed technology results in noninferior health outcomes compared to the comparator/standard practice.
- The use of the proposed technology results in inferior health outcomes compared to the comparator/standard practice.

Claims of inferior health outcomes are uncommon. Two examples where a claim of inferior health outcomes may be considered are:

- The health technology is less costly, and the magnitude of the health benefit lost is small.
- The health technology is more acceptable, or addresses equity issues, such that the uptake is expected to be greater than the current medical service. In this circumstance, it may be reasonable to explore a comparison against 'no medical service' for those that would not access the current service.

Clinical claims for investigative technologies

An investigative technology generates information about an individual in the target test population. This information is then interpreted, categorised and used for clinical decision-making. Clinical decisions may ultimately affect the health of the patient. The health impact of testing (i.e. the trade-off between benefits and harms across all patients who are tested) is defined as the clinical utility of the test.

While the impact of a test may ultimately be on health outcomes, the benefits of a test are often described in other ways, such as:

- an increase in the efficiency or ease of use of a test, with no change to the information derived
- an improvement in the efficiency or timing of information provided by a new test that replaces several sequential tests, but in comparison to the multiple tests, no new information would be provided

- a reduction in the adverse events associated with testing by using the new test, which may result in an improvement in health or may support a claim of noninferiority
- an increase in the acceptability or accessibility of a test, such that a broader population would access the test
- an improvement in the information provided, such that patients with a specific medical condition are more accurately categorised (which may or may not lead to improved health outcomes, depending on how each patient is managed).

The clinical claim can be informed by understanding the benefits of the test (relative to current care), and how this is likely to affect patient management, and ultimately patients' health outcomes (see Table 1). All health technologies, including investigative technologies (tests), are required to establish a claim that relates to health outcomes.

Table 1 Possible benefits of tests and suitable claims associated with these benefits

Comparative function	Possible benefits	Effect on management of patient	Health outcomes	Suitable clinical claims (health outcome gains)	Supportive evidence[a]
Test detects the same parameter as the comparator	**Replaces some or all current tests for the same condition. May replace no testing for some patients if new test increases coverage** • Faster, cheaper or more convenient • Smaller sample required, reduced re-biopsy rate, possibly safer • More accurate • More definitive or earlier result, can cease or reduce further testing • More feasible (panel vs sequential testing)	**No change in management**	**No change**	Noninferior to comparator	Evidence of no change in management. The downstream management will be the same for the same patients as the comparator.
		Change in management for at least some patients • Increase in coverage • Increase in compliance • Change in the patients identified (e.g. earlier in disease process) • Reduction in subsequent testing • Reduction in treatments for adverse events	**No change**	Noninferior to comparator	Evidence that overall health is noninferior
			Average improvement[c]	Superior to comparator	Evidence that overall health is superior
Test detects new parameter	**New test replaces current test** • Replaces test that detected a different parameter for the same purpose	**No change in management**	**No change**	Noninferior to comparator	Evidence of no change in management. The downstream management will be the same for the same patients as the comparator.
		Change in management for at least some patients • Better targets patients to appropriate treatments or subsequent management	**Average improvement[c]**	Superior to comparator	Evidence that overall health is superior

Comparative function	Possible benefits	Effect on management of patient	Health outcomes	Suitable clinical claims (health outcome gains)	Supportive evidence[a]
	New test or additional test[b] • Confirms diagnosis • Confirms diagnosis and provides additional information (e.g. predicts response to treatment) • Provides prognostic information (e.g. allows treatment planning, resource allocation or value of knowing) • Identifies new diagnosis or disease state • Monitors disease course	**No change in management**	**No change**	Noninferior to comparator	Evidence of no change in management. The downstream management will be the same for the same patients as the comparator. A new test that results in no additional health benefit but results in additional cost would require an additional claim associated with the nonhealth value of the test.
		Change in management for at least some patients • Better targets patients to appropriate treatments or subsequent management	**Average improvement[c]**	Superior to comparator	Evidence that overall health is superior

a Particularly for claims of noninferiority, evidence will be required to support no change in health *and* some evidence will be required to support the potential benefits of the test (such as speed or accuracy). Without some additional benefit of the test, the purpose for using the test rather than the comparator will be unclear. This may help avoid a proliferation of tests that do not provide additional information.

b For a new test (or an additional test) that results in increased costs, an improvement in health would typically need to be demonstrated. However, if the new test results in a change in management such that there are downstream cost offsets, a claim of noninferiority may be possible. Another possibility is that a new test may claim to be noninferior and the assessment of the test would include value of knowing, or other relevant considerations, to support the increase in cost.

c An average improvement in health is intended to reflect that, on balance, health for the whole tested population is improved. This accounts for both the health benefits and harms of the test, and the proportions of patients who will receive benefit or be exposed to harm.

Hierarchy of claims for investigative technologies

MSAC's Terms of Reference state that MSAC advises the Minister for Health on whether a medical service, health technology or health program should be publicly funded based on an assessment of the comparative safety, clinical effectiveness, cost-effectiveness and total cost. It is therefore logical that a claim regarding the comparative safety, clinical effectiveness and cost-effectiveness will be the most compelling for MSAC to consider. However, the committee is free to consider whether a test should be funded for reasons outside of these claims.

In some circumstances, investigative technologies may provide information that does not markedly affect health in a way that can be quantified, or they may affect personal wellbeing in ways that cannot be attributed to changes in the provision of health resources (Technical Guidance 28).

When deciding on the appropriate claim, it is important to consider a hierarchy of the claims that may be considered by MSAC (see Table 2). Where several claims are possible, claims that are higher in the hierarchy will be more informative for decision-making. The exception to this may be if a technology is likely superior in terms of health outcomes, but is no more costly than the comparator. In this case, it may be pragmatic to make a claim of noninferiority and pursue a more simplified economic approach.

In all circumstances, the clinical claim must relate to health outcomes. An assessment report may state that the clinical claim is accompanied by an additional claim for the value of knowing or other relevant considerations (see TG 1.4).

Table 2 Heirarchy of evidence and value for investigative technologies proposed by MSAC

Rank	Type of value supported by evidence	Claim
1	Direct from test to health outcomes evidence that shows an improvement in health outcomes Change in clinical practice for at least a proportion of investigated patients and linked evidence of an improvement in health outcomes. (A linked approach is more strongly supported if there is a clear codependency with a targeted therapy. Linked evidence may also be appropriate if the technology leads to outcomes such as stratification according to risk or prognosis, (such as staging of cancer) that lead to changes in clinical management for which the specific therapies might be less clear or more varied, but for which health outcome improvements could still be shown.)	Superior health outcomes
1a	Direct from test to health outcomes evidence that shows no change in health outcomes Linked evidence of no change in health outcomes	Noninferior health outcomes
2	Change in clinical practice for at least a proportion of investigated patients, without clear evidence of an improvement in health, but clear evidence or rationale that health is not diminished	Noninferior health outcomes
3	Change in family planning options (this would apply for testing of heritable mutations only)	Superior health outcomes (although may be restricted to intermediate outcomes)
4	Results more compelling (definitive, accurate, conclusive) than current investigations, thus leading to a reduction in current testing (e.g. diminishes the 'diagnostic odyssey')	Noninferior health outcomes (suitable if test cost is offset by avoiding subsequent tests). A claim of value of knowing is

Rank	Type of value supported by evidence	Claim
		required if the proposed test is more costly than the comparator test strategy.
5	Provides a basis for determining a clinical classification and thus informing prognosis or risk, but without changing clinical practice	Noninferior health outcomes, or value of knowing
6	Provides reassurance in a diagnosis or confirming the conclusions of other investigations.	Noninferior health outcomes, or value of knowing

TG 1.3 Comparing health care costs

The choice of a clinical claim, and the necessary approach to support that claim, is contingent on both the availability of evidence to substantiate the claim, and the cost of the new service relative to current practice. The 2 most common scenarios are:

- A proposed service that results in a greater cost to the health care system should lead to an overall improvement in health. A claim of superior health outcomes would be required, and the approach would seek to establish the magnitude of these benefits.
- A proposed service that is cost neutral or cost saving to the health care system should show at least no loss in health. A claim of noninferior health outcomes is possible, and in some cases, the approach required may be simpler than that for a claim of superior health outcomes. If the new service is expected to improve overall health but does not cost any more than the existing service, a claim of noninferiority would be sufficient to support the application.

TG 1.4 Making an additional claim for investigative technologies

In some circumstances, the health outcomes derived from the proposed investigative technology (and resulting changes to downstream management) may not justify the incremental cost associated with implementing the technology. In general, such technologies would not be considered cost effective, and would require additional evidence to support a positive funding recommendation.

For investigative technologies that would be regarded as cost-effective, *no additional claim should be made* (see Figure 2). However, other relevant considerations may remain an important component of the assessment report and should be presented (see Section 5).

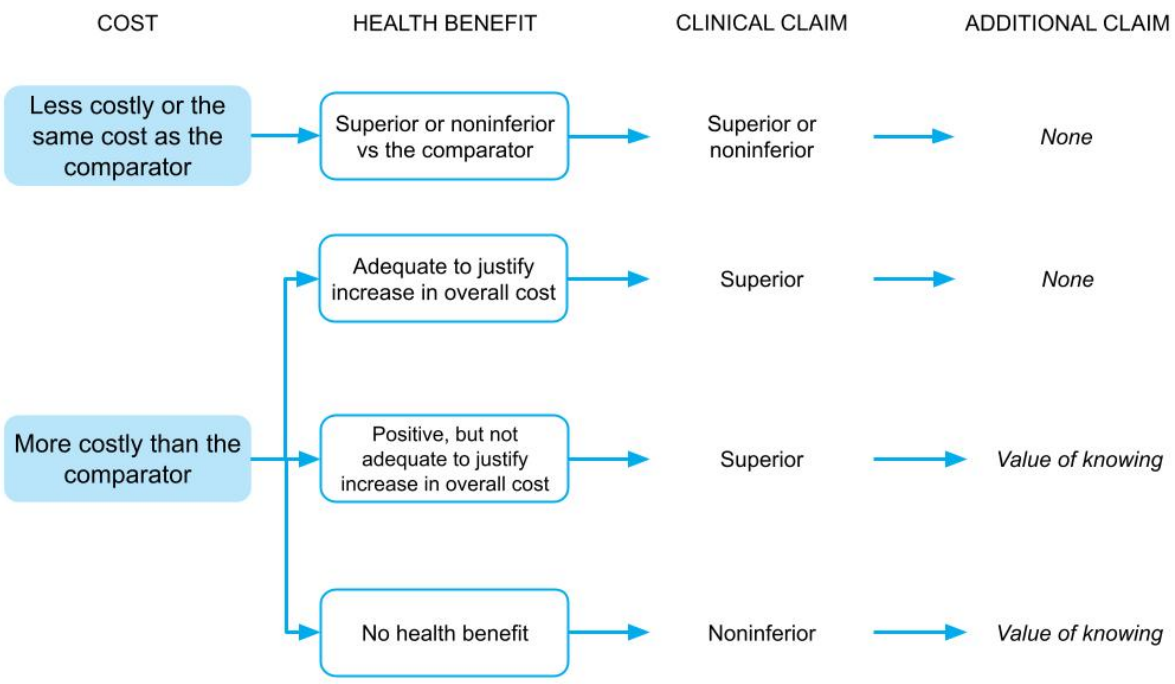


Figure 2 Deciding whether a claim in addition to the clinical claim is required to support the overall cost of the proposed investigative technology

It is envisaged that only investigative technologies would be required to make an additional (non-clinical) claim. Important additional nonhealth considerations for therapeutic technologies may also be presented, but would not invoke an additional claim.

If an additional claim is required, a clinical claim is made first (i.e. superior, noninferior or inferior health outcomes) and this is accompanied by a statement that the health technology results in other benefits versus the comparator. When presenting the additional claim, briefly state the nature of the additional benefits associated with the use of the health technology.

Value of knowing and other relevant considerations

Health technologies (therapeutic or investigative) may result in nonhealth outcomes, health outcomes that affect others, or health outcomes that may be difficult to quantify (such as quality of life related to knowing a diagnosis). These outcomes may be either positive or negative, and may involve a cost component.

Outcomes that are not captured in health outcomes claims may include the following (sometimes overlapping) categories:

- outcomes that affect others that may have an impact on quality of life, such as spillover effects on carers
- non–health care sector impacts, such as effects on educational attainment or attendance at work
- other utility outcomes, such as the value (benefits and harms) of knowing or naming a diagnosis.

These benefits, and how to present them, are discussed in Technical Guidance 28 and Technical Guidance 29.

When making a claim of value of knowing for an investigative technology, a parallel claim of health benefits (a clinical claim) *must* be made and tested with evidence. Usually, if a test is relying on a claim of value of knowing, the claim for an impact on health-related outcomes would be that the test is noninferior to current practice.

Appropriate claims for tests with a value of knowing are:

- The use of the proposed test results in noninferior health outcomes compared to the comparator/standard practice, but provides additional nonhealth harms/benefits to the patient or their families and carers.
- The use of the proposed test results in superior health outcomes compared to the comparator/standard practice and provides additional nonhealth harms/benefits to the patient or their families and carers.

The PICO to support claims of value of knowing must include outcomes that measure both the health-related claim (noninferiority/superiority) and the value of knowing claim. Outcomes for a value of knowing claim would include the types of options that become available to, or withdrawn from, an individual, or their family, as a consequence of information provided by a test. These options are not likely to be clinical (otherwise the claim would be for health outcomes), but may include the ability to make preparations, change behaviours or access support.

Technical Guidance 2 PICO

KEY CONSIDERATIONS

- **Define the target population (TG 2.1), the intervention (TG 2.2), the comparator(s) (TG 2.3) and the relevant outcomes (TG 2.5) to inform the assessment of the proposed technology.**
- **Provide a clinical management algorithm that describes clinical practice without the proposed technology, and a second algorithm that describes clinical practice with the proposed technology. Describe the main differences (TG 2.6).**

For investigative technologies

- **Define an appropriate reference standard against which the accuracy and performance of the proposed test is compared (TG 2.4).**

This guidance will inform the approach for developing a PICO confirmation, or for defining the PICO in an assessment report when no PICO confirmation is available. This TG subsection will also be relevant for determining whether an amendment to a PICO confirmation is required when preparing an assessment report.

Scoping searches for evidence about the health technology and the medical condition are required. Existing health technology assessments and systematic reviews may help inform the PICO. Relevant existing HTA reports or MSAC Public Summary Documents, and high-quality systematic reviews or key individual studies discovered during scoping searches should be made available to PASC. If no evidence is identified during these scoping searches, PASC should be made aware of the lack of evidence.

Multiple PICO sets may be required if:

- the health technology can be used across multiple populations
- the health technology can be used for different purposes

It is not always clear whether populations or test purposes are sufficiently different to require a separate PICO.

It is often the case that populations included in the same clinical study are sufficiently homogeneous to include in a single PICO. Cascade testing of family members is a clear example of when a separate PICO set is required.

If a test is intended to be used for diagnosis (to determine the presence of a disease) and also for monitoring (to inform adjustments to treatments), these are 2 separate uses (and also describe different populations) and separate PICO sets are required.

If multiple PICO sets are included, cross-reference to other PICO sets when details of PICO components are the same.

TG 2.1 Population

Provide an overview of the patient population, disease or condition that is targeted by the proposed health technology. Include relevant details of diagnosis, symptoms, prognosis, demographics and other issues relevant to the population targeted by the technology. State how potentially eligible patients are currently investigated, managed and referred within the Australian health care system in the absence of the proposed health technology. Provide detailed information on the natural history of the condition. This information is particularly important in circumstances where there is

no clear comparator, the comparator is no treatment or diseases are rare (and may have limited comparative evidence). The natural history should include whether the disease is stable, progressive, spontaneously remitting, fluctuating, episodic or probabilistic.

Characterise the Australian population for whom the health technology is intended, such as their age, sex, important comorbidities, and disease- or condition-related characteristics. Summarise the incidence and prevalence of the disease or condition in Australia using data from a reputable source, such as those listed on the Pharmaceutical Benefits Scheme (PBS) website: Sources of data for use in generating utilisation estimates.[a] Estimate the size of the population (and any relevant subgroups) expected to use the proposed health technology. For investigative technologies, provide the incidence and prevalence of the target population for the test (i.e. those who would receive the test) (see TG 11.9).

Compare the proposed population with the population defined in any relevant regulatory information (e.g. TGA indication) and/or key trial/study populations.

If the health technology is proposed for use in a subgroup of a population with a specific condition, describe the characteristics that identify the subgroup and a rationale for targeting the proposed subgroup. Explain which subgroups would be excluded from the target population.

If the health technology addresses health inequalities (such as those resulting from differences in access to care in rural and remote areas, or an area of unmet clinical need), the affected subgroups should be identified. Describe whether the proposed public funding will improve equity of access to the health technology, or whether it is likely to be used by those already receiving the services (through out-of-pocket expenses, or through state and territory funding).

If the proposed health technology is intended for use across multiple indications, tabulate these indications. In determining whether an indication is distinct, consider the following characteristics of the indication:

- differences in the target population
- differences in the disease, or location of the disease in the body
- differences in the mechanism of action of the proposed health technology.

Distinct populations and/or indications may require separate PICO sets.

If the proposed health technology is likely to be used for different purposes, and therefore has distinct indications, evidence to support each indication will be required (e.g. if it is used for diagnosis as well as monitoring). If different populations are to be assessed using evidence from one population (e.g. a population where there is a large amount of evidence) and assumptions made about how this evidence would apply to other populations (where there is little evidence), a biological rationale is required indicating that the disease/condition in the 2 populations would behave in a similar manner.

Codependent technologies

If an investigative technology being considered by MSAC is codependent, it is important to distinguish between the population eligible for testing and the population eligible for the treatment with the medicine or other therapeutic technology. A common error is to assume that the treated

a www.pbs.gov.au/info/industry/useful-resources/sources

population is identical to the tested population, whereas usually the tested population is broader than the treated population.

Genetic testing

If the test is for more than one indication, provide the clinical rationale for grouping the indications. For heritable conditions, provide the Online Mendelian Inheritance in Man (OMIM#) classification of the disease(s).

The relevant populations to describe for genetic testing are the testing population, the population with the pathogenic variant (test positive population) and a cascade testing population. Describe the genetic variants associated with the population of interest, and classify which of these are most prevalent, and which have the strongest clinical utility and/or cost-effectiveness argument. These are proposed to be the 'exemplar genes'. Rarer variants may be 'facilitated' (see 'Exemplar/facilitated assessment', TG 5.2). Given the high cost of many genetic tests, MSAC have an expectation that the number needed to test to identify one clinically actionable pathogenic variant is low. The precedent set by MSAC based on genetic testing for heritable variants in index patients with symptoms is a threshold of ≥10% for the pre-test probability of identifying a pathogenic variant in the target population. For cancer diseases, eviQ[a] is suggested as a suitable initial source of information on cancer genetics (such as information on frequency of variants in the population, and genes that are deemed clinically actionable).

When describing genetic tests or variants, italicise gene nomenclatures, but do not italicise gene products (proteins). Nomenclature guidelines are available from the Human Genome Variation Society.[b] This includes use of the term 'variant' rather than 'mutation'. Germline (heritable) variants are categorised using 5 classes:[2]

- Class 1 – benign
- Class 2 – likely benign
- Class 3 – variant of unknown significance
- Class 4 – likely pathogenic
- Class 5 – pathogenic.

Somatic variants are categorised using 4 tiers based on clinical significance:[3]

- Tier I – variants of strong clinical significance
- Tier II – variants of potential clinical significance
- Tier III – variants of unknown clinical significance
- Tier IV – benign or likely benign variants.

Describe the nature of the variants (such as deletions and copy number variations). Describe whether the variants identified are somatic or germline (heritable). When a condition is heritable, testing of an index case may affect the clinical management of family members. Cascade testing is the process of extending genetic testing to biological relatives of an index case with a pathogenic variant and is discussed in Technical Guidance 15. If cascade testing of first- and/or second-degree biological family members, or reproductive partners, would occur, this should be clearly described in a separate PICO set. For heritable conditions, describe the pattern of inheritance (e.g. dominant,

a www.eviq.org.au

b www.hgvs.org

recessive, X-linked). Describe the penetrance of the variants (i.e. what proportion of people with the variant express the phenotype associated with it).

Provide information on any differences in the variants found in different ethnic groups, and discuss those variants most likely to be relevant to the Australian population, including minority groups.

Populations not receiving the health technology

Health technologies may also have an impact on broader society if, for example, an infectious disease is detected, or if harms are caused to clinicians in the process of administering a treatment. Consider discussing these under 'Other relevant considerations'. However, if a key claim of benefit of a technology is the impact it has on the patient's family members/carers (e.g. improving their quality of life, allowing them to participate in the workforce), consider including studies reporting on these outcomes (in family members/carers) in the assessment report (clearly distinguished from outcomes for the patient).

TG 2.2 Intervention

Proposed health technologies

Describe the key components of the proposed health technology including whether it is an investigative or therapeutic technology (or both).

Provide details of how the proposed health technology is expected to be used, including frequency of use, mode of delivery, clinical setting, specialist training and provider type. Describe the required infrastructure for use of the technology, and whether the health system is currently able to provide this. State whether the proposed health technology is currently funded (in the public or private setting) in Australia for the same or another clinical indication.

Therapeutic technology

Describe the mechanism of action and the pathological process(es) the technology is claimed to address. Be explicit about whether the proposed therapeutic technology will be used in addition to existing therapies (add-on) or as a replacement for an existing therapy, or displace an existing therapy to a later line of treatment. If the technology can be used at different points in the management of the patient (i.e. lines of therapy), clearly describe when the technology will be used.

Investigative technology

Describe the purpose of the test; for example, diagnostic test, prognostic test, staging test, predictive test, monitoring test, surveillance test, cascade test, screening test (see HTA advisory committee glossary[a] for definitions). Tests used for multiple purposes usually require multiple PICO sets. Be explicit about whether the proposed test will be used in addition to existing tests (add-on) or as a replacement for an existing test, or used prior to existing tests (triage test). Describe any other elements of the test strategy, and how the proposed test would be incorporated into this. Describe how the test is ordered, how the test is performed, how the test results are interpreted (including any cut-off points), how the test result is communicated to the tested population, and how the test result is used to inform any clinical decisions.

As the health benefit of a test is indirect (i.e. it only influences health indirectly through the information provided by the test results), describe the downstream consequences of the proposed test that support the clinical claim (e.g. if a health benefit is claimed, how the benefit is achieved, and what treatments follow positive test results *and* negative test results). Discuss the biological

a www.pbs.gov.au/info/industry/useful-resources/glossary

plausibility for the impact of the test on patient health outcomes (noting that for 'black-box' tests this may not be possiblea). If the downstream consequences of the test are likely to change in the near future due to the availability of other technologies, consider including these in the assessment, even if they are not yet established in the Australian health care system.

Other relevant considerations

Identify contextual factors that could modify the effectiveness, clinical utility, test accuracy or safety of the health technology (such as the 'learning curves' of service providers). Discuss whether there are likely to be any implementation issues (e.g. a change in the specialty that delivers the technology, sample storage requirements, education and training requirements, changes in access to care, communication between and within organisations). Provide details of any quality assurance program or training program already in place or required.

If the technology is investigative, discuss whether the biomarker is correlated with factors such as ethnicity and sociodemographic status, and whether there are therefore ethical considerations that must be considered.

Codependent technologies

Describe any additional elements of the health technology for which funding is being sought (e.g. is it codependent with a medicine, such that only patients with a specific genetic variant determined by the proposed test will be eligible for a specific medicine).

If a codependent test–medicine combination is being assessed, ensure that the 'intervention' describes what treatment biomarker-positive patients *and* biomarker-negative patients would receive. In addition, it is important to describe the treatment that each of these groups would have received in the absence of the proposed test.

Proposed health technologies that include a medical device

For health technologies that use a medical device (such as surgery to place an implant, or adjustment of a pulse generator), identify devices currently on the Prostheses List that may be suitable for the service. State whether the current application to MSAC is combined with a Prostheses List application. Consider whether it is reasonable for similar devices to be in or out of scope for the assessment, based on substantial clinical equivalence, and regardless of whether or not they are on the Prostheses List. Justify the use of evidence gathered with an earlier model of the device, and consider how such evidence is transferable to the current model of the device.

If the health technology involves a particular brand of device, and the intention is for the proposed funding to be brand-specific, provide justification for why a brand-specific approach is more suitable than a device-agnostic approach (i.e. why other devices would not achieve the same results).

Multifactorial algorithms

If a prognostic or predictive test uses a combination of variables to predict a clinical endpoint, describe how the variables were chosen, and how the algorithm that combines these variables was developed and validated. The biological rationale and potential clinical application of the proposed test should be made clear.

In some circumstances, it may not be possible to describe the algorithm development and biological rationale because the algorithm is either commercial-in-confidence, or developed by machine-learning and considered a 'black box' (i.e. it is unclear what components are actually used in the

a For more information on 'black-box' algorithms, see TG 15.5.

algorithm). For a fixed algorithm, the dataset that was used for training the algorithm should be clearly described, including whether it was a convenience sample consisting of some positive and negative cases (diagnostic case control), or whether the tested cohort represents the characteristics of the target population in real-life practice.[4]

If the algorithm is not fixed (i.e. it is a self-learning algorithm), the methods of quality assurance should be described (i.e. how it can be ensured that the algorithm does not become biased or reduce in performance).

Additional information requests for multifactorial algorithms can be found in TG 15.5.

Genetic testing

Describe the type of genetic testing being performed to identify the variants described in the 'Population' section.

Describe the scale of gene analysis proposed, using the following MSAC-endorsed classifications:

- monogenic testing – limited mutation testing or whole gene testing
- small gene panel – assaying 2 to ≤10 genes
- medium gene panel – assaying 11 to ≤200 genes
- large gene panel – assaying >200 genes, but remaining sub-exome
- non-targeted – whole exome sequencing or whole genome sequencing.

Describe any possible alternative testing scenarios (such as alternative timing, or alternative combinations of genes/variants).

Describe the type of samples required (e.g. cheek swabs, tumour tissue, blood). Describe whether testing is qualitative or quantitative, and if quantitative, include the proposed allele frequency threshold and the rationale for this. Describe whether there is any need for confirmatory testing if a variant is identified, and the methods used for any supplementary testing. Classify the interpretive complexity (low/medium/high), taking into account both qualitative aspects (e.g. level of expertise required, complexity of bioinformatics pipelines, software requirements) and quantitative aspects (e.g. time component of labour required, cost of software licences). Include sufficient information to enable an estimate of the resources needed to adequately interpret the test results.

For any heritable variants, describe if cascade testing of family members or testing of prospective reproductive partners is warranted (see TG 15.1).

TG 2.3 Comparator

Select the comparator(s) in the context of the Australian population with the targeted condition, the current alternative health technologies for that condition in Australia, and the technologies most likely to be replaced (or added to) in clinical practice. A single comparator will be appropriate in most circumstances. The comparator(s) should be selected based on the technology most likely to be replaced or added to in clinical practice, rather than on the availability of evidence.

Therapeutic technology

Most comparators will involve one of the following:

- a current MBS-listed therapeutic technology – If the proposed therapeutic technology is likely to replace an existing MBS-listed service, the relevant comparator would be the existing therapeutic technology.
- standard medical management (reflected in studies as no treatment, placebo or sham treatment) – If the proposed therapeutic technology does not replace a current therapeutic technology, or is used in addition to a current therapeutic technology, the comparator would usually be standard medical management. Standard medical management may include the use of medicines, medical services, best supportive care or conservative management.
- current PBS-listed medicine(s) – If the proposed therapeutic technology is likely to replace pharmacological management of the target population, the relevant comparator would be the current PBS-listed medicine(s).

Investigative technology

Most comparators will be one of the following:

- a current MBS-listed test (or multiple existing tests/test strategies)
  - If the proposed test is likely to replace an existing MBS-listed test, the relevant comparator would be the existing test.
  - If the proposed test is likely to be used in addition to an existing MBS-listed test, the relevant comparator would be the existing test with no additional testing, and the intervention should be the proposed test plus the existing test (or plus or minus the existing test if the proposed test is a triage test).
- no testing and standard medical management – If the proposed test does not replace a current investigative technology, the comparator would usually be standard medical management and no testing.

The expectation is that the chosen comparator is a health technology with established cost-effectiveness. When the comparator is funded on the MBS, PBS, or Life Saving Drugs Programme (LSPD), it may be reasonable to assume that the cost-effectiveness of the comparator has been established, even if a formal cost-effectiveness analysis has not been performed. However, where the comparator is funded under a different source, and the cost-effectiveness of the comparator is unknown, the cost-effectiveness of both the comparator and the intervention may need to be established. The requirement for assessing the cost-effectiveness of the comparator should be discussed with PASC. If the comparator is known to be cost-ineffective, an additional comparator (what would be used in the absence of the first comparator) would also be required. Cost-effectiveness of the comparator would only need to be established in very few cases.

In situations where the health technology proposed for public funding is already established practice (i.e. it has already 'diffused'), the comparator should be what was used before the introduction of the health technology. If other health care changes have occurred in addition to the introduction of the proposed health technology, the comparator should reflect what would be expected to occur in the absence of the proposed health technology. The comparator for the budget impact analysis (Section 4) should always be current practice, regardless of the comparator used to determine the safety, effectiveness and cost-effectiveness of the health technology (i.e. if the intervention has already diffused, the budget impact analysis would assess the impact of cost-shifting from the current funding source to the proposed funding source, and any impact of an increase in utilisation).

Justify the selection of the comparator. The comparator should be clearly identifiable in the clinical management algorithm. Identify factors that may affect the main comparator in the future, such as the introduction of other near-market health technologies. If another health technology is reasonably expected to enter the Australian market for the same targeted population, it is optional to include it as a supplementary comparator. This is not a common occurrence, however, given the length of time between the development of the PICO confirmation and consideration of an assessment report by MSAC. Applicants should be aware that near-market comparators may also apply to MSAC, and the 2 interventions could be compared against each other. Procedural fairness allows both applicants to provide a comparison against the other.

If multiple comparators are identified, describe whether different comparators are used for different subpopulations of the overall target population. Include details of the subpopulations in the population section (TG 2.1). Multiple comparators may also be required if the proposed health technology is intended for more than one target population; this may require multiple PICO sets.

If multiple comparators have been identified in the PICO confirmation, present a comparison of the proposed health technology with each comparator. In the absence of a PICO confirmation, the usefulness of comparisons against multiple comparators for MSAC decision-making may depend on:

- the evidence supporting the choice of each of the comparators
- the risk that the proposed health technology is less effective than the comparator in a subpopulation
- the size of the subpopulation as a proportion of the overall target population (if there are comparators with a small market share, it may be appropriate to mention, but not focus on them in the assessment report).

Avoid comparing the proposed health technology against a 'basket' of comparators, where the effect of the proposed technology against individual comparators cannot be derived. If a basket of comparators is required because there is demonstrable ambiguity for the choice of the appropriate comparator (or comparators), and the evidence presented in the literature involves a comparison against the same basket of comparators, it is crucial to describe the applicability of the basket to the Australian setting.

Outline the funding arrangements for the comparator (i.e. provide details of any MBS items and other key health care resources required to deliver the comparator).

Codependent technologies

For codependent test–medicine combinations, be explicit about the comparators for the different components of the pairing. For example, if a new medicine requires a test to determine eligibility, but is compared against standard practice that does not require a test, 'no testing' will be the appropriate comparator for the test, whereas 'standard practice' would be the comparator for the medicine.

TG 2.4 Reference standard (relevant for investigative technologies only)

If a linked evidence approach is used, the results of the proposed test strategy will need to be compared to that of the comparative test strategy. If the concordance between the 2 test strategies is not exceedingly high, it should be determined which is the more accurate strategy. To do this, the 2 test strategies should be compared against a reference standard.

The reference standard is a test or series of tests that are used to determine the presence or absence of the target condition or clinical information of interest. Ideally, the reference standard is

the best available, clinically accepted, error-free procedure to do so. If there are any disagreements between the reference standard and the proposed test, it is assumed that the proposed test is incorrect. Thus, the choice of an appropriate reference standard is a very important determinant in establishing test accuracy.

The reference standard need not be a viable substitute for the proposed test. For example, if the purpose of the test is to determine the extent of a cancer to plan a surgical resection, the reference standard may involve the histological examination of the resected tissue by a pathologist.

If the purpose of the test is to predict a future health outcome, the reference standard is likely to be the health outcome (e.g. for prognosis, the reference standard may be the likelihood of cancer recurrence within 5 years; or for a predictive test, the reference standard would be whether biomarker-positive and biomarker-negative patients respond differently to a targeted treatment versus standard care). However, if there is an established way to measure the analyte or biomarker of interest (such as *BRCA* variant status), this could be used as a non-clinical reference standard.

There will be some instances where a reference standard does not exist (or where the proposed test is considered to be the reference standard). In these cases, the accuracy of the proposed test itself will need to be demonstrated by direct from test to health outcomes evidence showing a health benefit resulting from use of the test, or by comparison against a suitable clinical utility standard.

Clinical utility standard

A special type of reference standard is the test that was used in the generation of direct from test to health outcomes evidence, which establishes the clinical utility of a test. This test (including the method of acquiring the sample, testing characteristics and interpretation of the results) is called the clinical utility standard.

The description of the test should explicitly state the test reagent(s)/kit, test platform, biospecimen type and preparation, what is tested (e.g. gene sequence, gene expression, staining, whether all cells or only some cells), body part(s) scanned, tracer used, qualitative or semi-quantitative interpretation of the image, and definition of the test/scan threshold result that differentiates between different clinical management actions (e.g. eligible for the targeted therapy or not).

If the proposed testing methods are not identical to the clinical utility standard (i.e. do not overlap completely with the tests linked to health outcomes), a comparison of all proposed tests against the clinical utility standard is required.

Box 1 Example of using a clinical utility standard

An applicant developed assessment report (ADAR) for a codependent test–drug is submitted to MSAC and the Pharmaceutical Benefits Advisory Committee. The ADAR provides evidence that a branded assay 'Eleasa' is predictive of treatment response to the targeted treatment 'Erloxiline'. The Eleasa assay (including the type of sample taken, test method and interpretation of results used in the key trial) is the 'clinical utility standard'. In Australia, laboratories may use the Eleasa assay, or a range of in-house assays. To demonstrate that use of the in-house assays will have similar health consequences as the Eleasa assay, concordance between the in-house assays and the clinical utility standard is required. In this case, the comparison would describe the number who would be test positive or test negative with the clinical utility standard and with the proposed test, so that the implications of discordance can be discussed.

If there is no overlap (i.e. if the clinical utility standard is not available in Australia), a comparison between those tests proposed for use in Australia and the clinical utility standard is of critical importance to allow MSAC to make a decision.

For more information on the clinical utility standard, see TG 11.2.

TG 2.5 Outcomes

The outcomes chosen influence the scope of evidence included in the assessment report. The assessment phase should identify whether evidence is available for the chosen outcome measures.

Identify the patient-relevant health outcomes for the target population, disease or condition, and decide which are the most critical outcomes to assess to address the clinical claim (TG 1.2). The outcomes that will be most influential for MSAC decision-making are those that are patient relevant and demonstrate the safety and effectiveness of the technology compared to the comparator.

Clinical outcomes should be limited to those that would reasonably influence MSAC decision-making. The GRADE (Grading of Recommendations, Assessment, Development and Evaluations) approach recommends that no more than 7 key outcomes (including benefits and harms) should go into summary of findings tables, based on the amount of information that decision-makers can balance at once.[5]

Patient-relevant health outcomes (relevant for both therapeutic and investigative technologies) include:

- outcomes that are directly relevant to the patient, reflecting improvements in quality or length of life – Surrogate or intermediate outcomes are acceptable if they have been validated as being able to predict patient-relevant outcomes.[6] If known at the PICO confirmation stage, provide validated examples of transformation from the surrogate to a patient-relevant outcome.
- outcomes that relate to the direct safety of the health technology or comparator (e.g. harms from biopsy or radiation) or the indirect safety (e.g. harms caused by learning curve or insufficient training, lack of equipment maintenance, inappropriate patient selection)
- outcomes that relate to the effectiveness or safety of any downstream interventions
- outcomes that are expected to change if the proposed health technology is publicly funded
- outcomes that are expected to be no different if the proposed health technology is publicly funded (required in the assessment of noninferiority).

Different outcomes may be necessary if the proposed technology is intended to be used across different indications (e.g. diagnosis and predisposition testing) or across subpopulations with different characteristics. Clearly state which outcomes are relevant for each indication or subpopulation. Differences in indications, purpose or populations may require multiple PICO sets.

Family and societal outcomes include (Technical Guidance 28 and Technical Guidance 29):

- outcomes that relate to benefits or harms of the intervention or comparator beyond the index patient (i.e. psychological or physical benefits or harm to family members, fetus, carers, health care professionals, health care setting, public or the environment)
- outcomes that reflect broader consequences to the health care system, including changes to health care resource use (hospital days, patient throughput, impacts on other services or medicines)

- outcomes that relate to health disparities (e.g. equity of access, areas of unmet clinical need).

Given resource constraints, prioritise which outcomes are most important to include. PICO confirmation developers should explore a wide range of outcomes, and then seek feedback from PASC and other stakeholders about which are the most important.

For superiority claims, define the minimal clinically important difference (MCID) for the key outcomes. For noninferiority claims, define and justify the noninferiority margins of key outcomes. For more information on MCID and noninferiority margins, see Appendix 5.

Identify any other relevant considerations that will affect the implementation of the health technology, such as patient acceptability or compliance, privacy concerns, cybersecurity concerns and biosecurity. Some of these concerns will be relevant to devices or services that transmit or store data, genetically modified organisms and biologicals. For more information on other relevant considerations, see Technical Guidance 29.

Investigative technology

The harms and benefits of treatment or intervention following a test (for both test positive and test negative populations) are relevant outcomes.[7] Direct from test to health outcomes evidence provides relevant harms and benefits (as long as the outcomes reported are relevant). If an investigative technology requires a linked evidence approach to demonstrate clinical utility, test performance and change in management are required outcomes.

Test accuracy outcomes are described in Technical Guidance 11 and include:

- concordance of the range of tests used in Australia with the clinical utility standard (the test used in the key trial demonstrating clinical utility of the test)
- outcomes related to the diagnostic accuracy of the test (demonstrating the accuracy of biomarker or disease detection compared to a reference standard)
- outcomes related to the longitudinal accuracy of the test (demonstrating how accurately the test estimates the health outcome of interest)
  - prognosis (with reference to a health outcome at a later time point)
  - predictive (with reference to response to treatment)
  - monitoring of disease (with reference to changes in a health state)
  - monitoring of treatment response.

Change in management outcomes are described in Technical Guidance 12 and include:

- outcomes related to changes in diagnostic thinking and subsequent testing
- outcomes related to changes in preventive or therapeutic strategies
- outcomes related to adherence to preventive or therapeutic strategies
- outcomes related to referral patterns and/or the frequency and timing of follow-up.

Health outcomes resulting from a change in management are described in Technical Guidance 13 and include:

- outcomes demonstrating the clinical utility of a change in management (e.g. mortality, morbidity, quality of life)
- outcomes assessing safety of the downstream implications of testing.

Tests may also provide additional value to patients or to their family members and carers, and discussion of these outcomes could supplement an assessment of the clinical utility of the test. In cases where clinical utility cannot be demonstrated (e.g. where no change in patient management occurs as a result of the test information), and an additional claim is made regarding the value of knowing, these outcomes will be key to consider. Examples of value of knowing outcomes include:

- ending the patient's diagnostic odyssey
- reproductive planning
- long-term planning (e.g. education, career, housing, finances)
- increased/decreased sense of control
- psychological (positive or negative) impact on index patient
- stigmatisation or discrimination
- access to the National Disability Insurance Scheme
- greater understanding of future health care needs
- the ability to connect with others in the same situation.[8]

For more information on the value of knowing, see Technical Guidance 28.

Codependent technologies

For test–medicine codependent technologies, specify which outcomes are relevant for assessing each component. Relevant outcomes for the test will include test accuracy and performance compared with the clinical utility standard, and relevant outcomes for the therapeutic component will be health outcomes. Define health outcomes for both test positive and test negative populations.

TG 2.6 Clinical management algorithms

Prepare an algorithm (flowchart) that depicts current management or investigations, plus management of the disease or condition in the Australian target population in the absence of the proposed health technology (i.e. the comparator). Prepare a second algorithm that depicts the eligible patients and the circumstances of use of the proposed health technology if the MBS listing or other public funding is implemented as requested. The 2 algorithms may be combined into a single flowchart, if appropriate.

The clinical management algorithms provide clarity about:

- the target population
- intended use of the proposed health technology
- the comparator technologies that would be replaced, added to or displaced
- possible changes in patient management due to the proposed health technology
- changes in resource use.

Algorithms must be consistent with the population, intervention and comparator discussed earlier in this Technical Guidance subsection, and should inform the structure of the economic model (Technical Guidance 18).

Indicate whether the proposed health technology can be used at different points in the clinical management algorithm, or present multiple algorithms for different clinical indications or uses. Justify the positioning of the proposed health technology in the clinical management algorithm.

Use the following sources to inform the algorithm(s) (in order of preference):

- a literature review of relevant published clinical management guidelines – Independently developed, up-to-date evidence-based clinical practice guidelines developed for the Australian setting are preferred. If possible, verify that the identified guidelines are currently in use and adhered to. Aspirational guidelines may not reflect current practice, and guidelines developed in an area that is rapidly evolving may no longer be relevant
- existing studies on the management of the condition
- an expert panel and/or a well-designed survey (if current guidelines or literature are not available)
- expert clinical opinion – see Appendix 9 for further advice on expert input.

Identify the following characteristics in the algorithm(s):

- diagnostic criteria and/or prior tests to determine the target population (eligible patients), including tests required to support any proposed continuation criteria
- important characteristics of eligible patients (such as risk factors, severity of disease and remaining treatment options)
- circumstances of use of the proposed health technology, including who is providing the service, and whether special training or specialised facilities are required (provide a justification for these below the flowchart)
- treatments provided, including any required prior tests, medical services or treatments, required co-administered therapies, and consequences for subsequent therapy options; give particular consideration to whether a proposed health technology is likely to replace a currently available option, or whether it is likely to displace that option to a later line of therapy
- health care resource provision, both before and after the point in the flowchart at which the proposed health technology is introduced.

Extend the clinical management algorithms to the expected end of the disease or condition process, or until the algorithms for the proposed health technology and the main comparator(s) are expected to be the same.

Summarise the differences between the current and proposed clinical management, as depicted in the algorithm(s).

Technical Guidance 3 Proposed funding arrangements

KEY CONSIDERATIONS

- **Describe and justify the proposed funding arrangement.**
- **For funding of the proposed technology sought through the MBS, take one of the following approaches:**
  - **Identify an existing MBS item that can be used unaltered.**
  - **Amend an existing MBS item descriptor.**
  - **Propose and draft a new MBS item descriptor.**
- **State and justify the proposed fee.**

Applications should state and justify the funding arrangement being proposed (i.e. whether a new MBS listing is requested, an amendment to an existing MBS listing, funding for a package of care, or whether another form of public funding is sought, and why).

Outline the expected changes to health care resources if a patient receives the intervention rather than the comparator. This is discussed in more detail in Technical Guidance 22.

TG 3.1 Proposed MBS item descriptor

If public funding is sought through the MBS, the health technology may be funded under an existing MBS item, an amended MBS item, a new MBS item, or it may require a combination of these options. Follow the guidance relevant to the circumstances of the proposed health technology in this subsection.

Existing MBS item

Provide the existing MBS item and descriptor that the health technology is proposed to be covered under.

Amended MBS item

Provide the existing MBS item and descriptor, and the draft changes to the MBS item (use highlighting or strikethrough to clearly present proposed changes). Report the nature of proposed change(s), which may include one or more of the following:

- an amendment to the way the health technology is clinically delivered under the existing item(s)
- an amendment to the patient population under the existing item(s)
- an amendment to the schedule fee of the existing item(s)
- an amendment to the time and complexity of an existing item(s)
- access to an existing item(s) by a different health practitioner group
- minor amendment(s) to the item descriptor that does not affect how the health technology is delivered
- an amendment to an existing specific single consultation item
- an amendment to an existing global consultation item(s)
- other change (describe).

New MBS item

Table 3 shows the components of a new MBS item.

Table 3 Proposed MBS item

Component	Description
Category	[proposed category number] – [proposed category description]
Group, Subgroup	[proposed group number] – [proposed subgroup]
Proposed item descriptor	[provide proposed item descriptor]
Proposed fee	[provide the proposed fee]

Draft an MBS item descriptor that defines the target population and the health technology that would be eligible for MBS funding. The objective of the MBS item descriptor is to specify criteria that would reasonably ensure that the performance and costs of the proposed health technology are consistent with the conclusions in the assessment report. A broad MBS item descriptor may permit access to a population for whom the performance of the health technology is limited or unknown, and may result in a higher cost to government. The descriptor should therefore be specific enough to minimise the usage beyond its intended population ('leakage'). If clinical guidelines or practice in some settings (such as the public hospital system or internationally) use the technology in a broader population than is proposed for MBS funding, consider whether this will have implications on the potential for leakage.

Discuss the relevance of the criteria included in the proposed MBS item descriptor (such as the health practitioner group that may access the item, the population and the intervention). State whether other items are expected to be used in conjunction with the item number (such as anaesthetic item numbers). Identify possible eligibility criteria that have been omitted from the MBS item.

Specify who is able to request the service, and how often the service may be claimed (e.g. times per year, or if there is a once per lifetime limit). Specify where the service will be provided (e.g. as an in-hospital service on an admitted patient, an in-hospital service on an admitted or non-admitted patient, or out-of-hospital/outpatient service). Where relevant, state the applicable type of procedure: Type A, B or C.

- Type A procedures are overnight procedures
- Type B procedures are same-day hospital procedures
- Type C procedures are often out-of-hospital procedures that do not normally require admission.

Specify any exclusion criteria (such as items that cannot be co-claimed) and any exemptions.

MSAC prefers technology-agnostic language for item descriptors, allowing for a variety of health technologies (particularly tests) to use the same item number without the need to amend the MBS item descriptor. Use of tradenames should therefore be restricted to cases where it can be demonstrated that generic language may result in dissimilar health technologies being able to use the proposed MBS item.

Multiple MBS items may be required to plan and administer a technology, or to use over the life of a technology; for example, for the implantation and removal of devices. Draft as many item numbers as required.

If the costs associated with using the proposed technology in different populations will differ, separate MBS items will be required to allow for different fees.

If cascade testing of biological relatives or testing of reproductive partners is relevant, draft additional MBS items to cover these indications.

MBS item fee

Explain how the proposed MBS fee has been derived. In circumstances where the health technology is proposed to be covered by an existing or amended MBS item, explain why its fee is appropriate.

A comparison of MBS item fees of similar services may provide some context for the proposed fee. However, greater justification for a fee will usually be required. Typically, the health technology should be costed with reference to relevant input costs. These may include the costs of the time taken for the provider to perform the service (before, during and after the service), and direct service costs such as costs of the time taken for identified 'non-rebated employees' to be involved in providing the service, costs of identified consumables, and costs of shared use of identified reusable equipment. Any consumables (such as sterile sheaths and IV giving sets) that must be used by a practitioner to conduct a procedure are generally included within the MBS item fee. There is also precedent that wet laboratory consumables may be included in MBS fees for pathology services.

If the proposed health technology is already in use in Australia, provide a summary of the fees charged for this service.

Include information about any Extended Medicare Safety Net risk, where an out-of-hospital benefit cap may need to be applied. This usually applies to out-of-hospital services where practitioners are charging large amounts. Where MBS fees are higher than the Greatest Permissible Gap (GPG) threshold, include the likely GPG rebate amount (for out-of-hospital services).

Genetic tests

MSAC-endorsed gradations for panel sizes listed in TG 2.2 correspond to different fee levels, so benchmarking against MBS items that fall into these categories would be appropriate.

Genetic counselling represents best practice for patients or family members undergoing genetic testing for a heritable disease, and this should be discussed under other relevant considerations (Technical Guidance 29). However, genetic counselling cannot be funded separately via the MBS. Consequently, it is suggested that genetic counselling be captured within an appropriate consultation item claimed by the professional responsible for the care of the patient (usually the professional requesting the test). The cost of genetic counselling should be incorporated into the economic modelling and estimates of financial impact. Genetic counselling is not included as a criterion in the MBS item descriptor.

TG 3.2 Alternative arrangement for funding

Describe the arrangements for public funding sought through mechanisms other than the MBS. Explain the process of identifying the target population, and methods for restricting use to the intended population.

Explain the proposed fee or the amount to be charged.

Technical Guidance 4 History of MSAC submissions for the health technology

KEY CONSIDERATIONS

- **Document the history of applications and committee considerations for the proposed technology.**
- **For reconsiderations, describe the matters of concern raised by the committee and the changes in the current assessment intended to address the concerns.**

Tabulate the dates of previous committee considerations for the proposed health technology and indication (see Table 4).

Table 4 MSAC submission history

Committee	Meeting date(s)
PICO Advisory Sub-Committee	[add]
Evaluation Sub-Committee	[add]
Medical Services Advisory Committee	[add]

For reconsiderations, present a table that summarises the issues raised by MSAC, and show how the current assessment report addresses the issues, with cross-referencing to the relevant sections of the current assessment report (see Table 5).

Summarise the key matters from the previous MSAC considerations and cross-reference to the MSAC Public Summary Document, where possible. Highlight key points from the Public Summary Document and cite the relevant paragraph/page. Identify any important matters of concern raised by the Evaluation Sub-Committee (ESC) that were not over-ruled by MSAC. Issues raised by ESC that are not specifically mentioned by MSAC remain outstanding.

Table 5 Summary of key matters of concern

Component	Matter of concern	How the current assessment report addresses it
[Identify the relevant section of the previous assessment report, e.g. comparator, clinical claim, economic evaluation]	[Cite paragraph of the MSAC PSD (use abbreviated referencing in tables), identify matter of concern]	[(Addressed/Not adequately addressed/not addressed) Comment and/or cross-reference to where matter is addressed in the executive summary or main body.]
Example text: Clinical place in therapy	Example text: MSAC suggested the descriptor should reinforce that psychotherapy must have been previously trialled (PSD, p.2).	Example text: Addressed. Restriction amended to reflect MSAC comments.

Component	Matter of concern	How the current assessment report addresses it
Example text: Clinical effectiveness	Example text: MSAC noted there was other available evidence that could be informative on the relative effectiveness that was not presented in the reconsideration, including the EUnetHTA 2017 and Ontario Health 2016 reports (PSD, p.3)	Example text: Addressed. The efficacy results from EUnetHTA 2017 are now applied in the economic modelling as this is the more recent of the 2 reviews requested to be reviewed by MSAC.

EUnetHTA = European Network for Health Technology Assessment; MSAC = Medical Services Advisory Committee; PSD = Public Summary Document

Technical Guidance 5 Approach to assessment

KEY CONSIDERATIONS

- **Most applications considered by MSAC and its subcommittees will require a full HTA to inform the effectiveness, safety and cost-effectiveness of the proposed technology relative to the main comparator(s) (TG 5.1).**
- **For some genetic tests involving the assessment of multiple, related genetic variants, a simplified approach for some of the less common genetic variants may be reasonable:**

  – Nominate exemplar genes, that is, the genetic variants that are more prevalent and have the strongest evidence linking their identification through the proposed test with health outcomes (TG 5.2).

  – Describe the relationship between the remaining genetic variants and the exemplars. Explain how the underlying principles required to support a facilitated approach are satisfied (TG 5.2).

Part of the process of consideration of an application by PASC should be to establish whether all components of the proposed listing should be evaluated with a full HTA, or whether the exemplar/facilitated approach may be incorporated.

TG 5.1 Full health technology assessment

Most health technologies assessed by MSAC will require a full HTA to be performed (including a systematic review of the safety and effectiveness, and determination of the cost-effectiveness and budget impact of the technology).

For more information on methods used in literature searching, assessing the risk of bias, GRADE and extracting the key characteristics of studies, see Appendix 2 to Appendix 5. The methods for assessing therapeutic technologies are well established, and therefore quite stable. However, methods for assessing investigative technologies are less established, and new guidance is provided in the appendixes on using risk of bias tools to assess a range of tests with different uses, and how to adapt the GRADE approach for different linked evidence components.

TG 5.2 Exemplar/facilitated assessment

An exemplar/facilitated assessment is an alternative approach that is intended to simplify the assessment of gene-related investigative technologies. The principles of an exemplar/facilitated assessment involve describing similarities between technologies for which there is adequate evidence to inform a decision and technologies that have less established evidence. While this concept is not unique to genetic/genomic technologies, at the time of writing these guidelines, the best examples of this assessment approach have been in the genetic/genomic space.

Defining the exemplar and facilitated genes

A gene panel or whole exome/genome analysis seeks to identify pathogenic variants in genes that may result in disease. There may be limited evidence to inform the clinical utility associated with the testing for many variants. However, certainty of decision-making is improved when there is evidence to support the clinical utility of testing for the *most prevalent variants* or for *variants associated with management changes that have large impacts on health*.

A proposed exemplar gene is one where:

- the comparative safety and effectiveness (in terms of health outcomes) of testing for pathogenic variants in the gene have been established or are expected to be established in the current assessment
- the impact that testing of the gene has on health outcomes (the clinical utility) is derived from direct from test to health outcomes evidence, or a linked evidence approach
- the extent of the involvement of pathogenic variants of the gene in the disease (prevalence) of the target population is known.

A proposed facilitated gene is one where:

- the gene is included on the same panel (or the same whole exome/genome sequencing analysis) as the proposed exemplar gene(s)
- the reason for testing for pathogenic variants in the gene is the same as that for the exemplar(s) (e.g. to provide a diagnosis for the same disease and/or used in the same population)
- the pathogenic variant(s) of the gene is less prevalent than the exemplar(s)
- the requested fee or unit cost of the panel would not increase with the inclusion of the gene
- there is a clear basis to conclude that the identification of pathogenic variants would not result in increased costs or decreased health outcomes compared with testing only for the exemplar gene(s).

Box 2 Example of an exemplar/facilitated relationship

A panel for detecting heritable hypertrophic cardiomyopathies contains 10 or more genes. The most common genes implicated in heritable cardiomyopathies are *MYH7* and *MYBPC3*, accounting for about 50% of identified hypertrophic cardiomyopathies. These genes are included in the proposed panel. The link between pathogenic variants in these genes and disease is well established. The evidence supporting improvements in health of individuals identified with pathogenic variants in these genes is weak. There is some evidence that screening family members of patients identified with pathogenic variants of *MYH7* or *MYBPC3* may result in early identification of cardiomyopathies, and a reduction in unnecessary monitoring in those that are not carrying the genetic variants. These genes are the exemplars.

In addition to *MYH7* and *MYBPC3*, pathogenic variants in many other genes have also been identified, with some panels including 40 genes or more. The prevalence of these pathogenic variants is much less than *MYH7* and *MYBPC3*. There is evidence that these genes are implicated in hypertrophic cardiomyopathy, but it may not be as strong due to the rarity of the pathogenic variants. The inclusion of these genes onto the panel would not increase the utilisation of the panel, as the genes would only be tested in patients with suspected hypertrophic cardiomyopathy, which is the same indication as for the exemplar genes. The management of patients with hypertrophic cardiomyopathy is unlikely to change for different identified genes, and hence the downstream costs and consequences are likely to be the same. These genes are appropriate to propose as facilitated genes.

Principles of an exemplar/facilitated assessment

If the cost-effectiveness of testing of the exemplar gene(s) can be justified, testing of additional genes (with less evidence, but minimal impact on cost) need not be assessed through a full HTA approach. Instead, an exemplar/facilitated assessment may be used, which reduces the information required for proposed facilitated genes.

If including additional genes on a panel results in an increased uptake of testing, or an increase in the unit price of testing, a full HTA approach is required to establish that the increases in cost are associated with an improvement in health. If the testing of additional genes results in marked changes to downstream health impacts or costs, it is not appropriate to nominate such genes as facilitated genes and an exemplar/facilitated assessment cannot be used.

An exemplar/facilitated assessment is intended to provide assurance that the inclusion of additional genes will not result in a loss of health, or a marked increase in test costs or downstream costs.

Description of an exemplar/facilitated assessment

Pre-assessment phase (application form and PICO confirmation)

- Determine which genes harbour the pathogenic variants responsible for disease. Of those genes, identify those with the highest prevalence and those with the strongest evidence to support a link between testing for the gene and health outcomes. From this list, nominate the exemplar genes.
- If the panel or sequencing covers a range of diseases, first identify the most prevalent diseases and those where testing is likely to have the largest impact on health outcomes. For each of these 'exemplar diseases', nominate the exemplar genes.

The PICO confirmation should clearly present and justify the choice of the exemplar genes. This is necessary to guide the assessment.

Assessment phase

A full HTA quantifying the clinical utility of testing the exemplar gene(s) is required (see Section 2B for guidance on the assessment of investigative technologies).

Present evidence to support the principles of the exemplar/facilitated assessment approach. Tabulating the relationship of the exemplar gene(s) and the facilitated genes according to the principles of the approach is useful (see Table 6).

Table 6 Example of tabulating exemplar and facilitated genes

Characteristic	Exemplar gene(s)	Facilitated genes
Intervention	*BRCA1/2* testing	*STK11*, *PTEN*, *CDH1*, *PALB2* and *TP53*
Population that would be tested	Diagnosed with breast cancer or ovarian cancer with a >10% chance of having a pathogenic mutation (based on clinical and family history)	Same as exemplar No increase in number of tests performed
Prevalence of the pathogenic variants in disease	Patients are eligible for testing if clinical factors (e.g. cancer type, age, history) mean that the post-test probability of a pathogenic variant is >10%	The combined yield of the facilitated genes accounts for about 40% of the positive tests (with *BRCA1/2* accounting for about 60%)
Unit price of the testing	$1,200	Same as testing for the exemplar genes
Correlation with disease (clinical validity)	Established	Some evidence to support a link between the facilitated genes and disease
Clinical utility	Established	Currently no change in management
Cost-effectiveness	Approved	No additional costs and no loss in health

Addressing the principles of the exemplar/facilitated assessment approach is evidence based. Evidence is required to support claims of:

- prevalence of the pathogenic variants of the facilitated genes
- correlation between the facilitated genes and disease
- change in management (and, if relevant, the impact on health)
- if relevant, the costs of downstream changes to management.

In a full HTA approach, a systematic search is required to inform the direct from test to health outcomes evidence or the components of the linked evidence approach. A systematic search is unlikely to be useful for less prevalent pathogenic variants of facilitated genes. Identifying evidence to support the clinical validity of the facilitated genes may include reference to gene databases (e.g. ClinVar). Evidence to support change in management may include a survey of relevant clinical specialists.

Present a stepped economic evaluation that first assesses testing only for the exemplar gene(s), with subsequent steps including the increases in costs and additional benefits associated with the inclusion of the facilitated genes on the panel.

Possible facilitated approach 1

If the inclusion of facilitated genes on the panel does not result in any expansion of the population (and has no additional downstream costs), the size of health benefit associated with the facilitated genes does not need to be quantified. There must be some logic that the benefits of testing for the facilitated genes are likely to outweigh any harms. In this situation, the cost-effectiveness and financial impact of listing a panel with only the exemplar genes is not likely to differ with the inclusion of additional facilitated genes (see Figure 3).

Box 3 Examples of facilitated genes with the same population

1. If *BRCA1/2* testing is considered cost effective in patients with breast or ovarian cancer, adding additional variants in *STK11*, *PTEN*, *CDH1*, *PALB2* and *TP53* genes to the same panel can be done without quantifying the benefit of testing for these additional genes in the same population.
2. A large gene panel is proposed for opportunistic preconception carrier screening of prospective parents for heritable variants that can cause serious diseases. The population is asymptomatic prospective parents. Increasing the number of genes from a small number to a large number will not alter the number of people likely to get tested, unless there is family history of a particular condition that is only represented by genes tested for on the broader panel. The assessment would determine the most commonly identified variants, the variants associated with the most serious diseases, and the variants with the highest quality evidence supporting a change in health outcomes following testing. These variants inform the choice of the exemplar genes. An assessment of the clinical utility or value of knowing for the exemplars would be performed using a full HTA approach. Separate assessment of the clinical utility or value of knowing for the facilitated genes would not be required.

Possible facilitated approach 2

If listing of the facilitated genes results in an expansion of the population who would use a gene panel, the additional financial impact requires that the cost-effectiveness is considered (see Figure 3).

One of the underlying principles for exemplar/facilitated assessments is that the exemplar listing is cost-effective. If the costs associated with including the facilitated genes (without any benefits attached to their inclusion) can be incorporated into the economic model of the exemplar genes, and the panel remains cost-effective, the size of the expected health benefit of adding the facilitated genes need not be quantified.

If incorporating the costs of the facilitated genes due to the expansion of the population into the economic model reduces the cost-effectiveness of the proposed listing, the benefits of including the facilitated genes on the gene panel need to be estimated.

If it is logical (based on evidence, expert clinical opinion or biological plausibility) that the health benefits of testing for the facilitated genes are the same or similar to those for the exemplar genes, the benefits of the exemplar genes can be assumed to apply to the facilitated genes. Costs and benefits can therefore be incorporated into an economic model.

Box 4 Example of facilitated genes and facilitated populations

A panel for heritable variants associated with colorectal and endometrial cancer was proposed. However, there are many different types of colorectal and endometrial cancer.

The bulk of the clinical evidence identified was for patients with clinical signs/symptoms of Lynch syndrome and familial adenomatous polyposis (FAP). These could be considered the exemplar conditions. Two economic models were created: one for Lynch syndrome, and one for familial polyposis (including FAP, and MUTYH-associated polyposis [MAP]). The exemplar genes were *MLH1*, *MSH2*, *MSH6* and *PMS2* for Lynch syndrome; *APC* for FAP; and *MUTYH* for MAP. Less common conditions (which were facilitated) were juvenile polyposis syndrome, Peutz-Jeghers syndrome and hereditary mixed polyposis syndrome. Five additional facilitated genes were requested to be on the panel. The financial analysis incorporated the cost of testing patients likely to have these syndromes, although they were not incorporated into economic modelling.

MSAC accepted the clinical and cost-effectiveness evidence for Lynch syndrome and FAP, and considered it reasonable to rely on the same evidence base to include the requested testing of additional genes to detect variants associated with the identified rare syndromes associated with colorectal or endometrial cancer (MSAC 1504 final PSD).

Unsupportable scenario

When any of the underlying principles of the exemplar/facilitated assessment approach are not met, a full HTA is required to support the inclusion of the facilitated genes on the proposed panel.

Box 5 Example of unsupportable scenario

In 2018, MSAC supported public funding for genetic testing to confirm the diagnosis of Alport syndrome (a heritable form of kidney disease). In 2019, an application was received to amend the MBS listing from patients with suspected Alport syndrome to all forms of heritable kidney disease, with the clinical claim that the advantages and disadvantage of genetic testing are identical to those for Alport syndrome. An exemplar/facilitated assessment was considered, to determine if an abbreviated assessment of the broader population could be performed, using Alport syndrome as the exemplar condition. However, one of the underlying principles of a facilitated/exemplar assessment is that the facilitated listing is for a smaller population than the exemplar listing. As genetic testing for heritable kidney disease (other than Alport syndrome) would be performed more frequently if publicly funded than genetic testing for Alport syndrome, the expanded population could not be considered as a facilitated assessment.

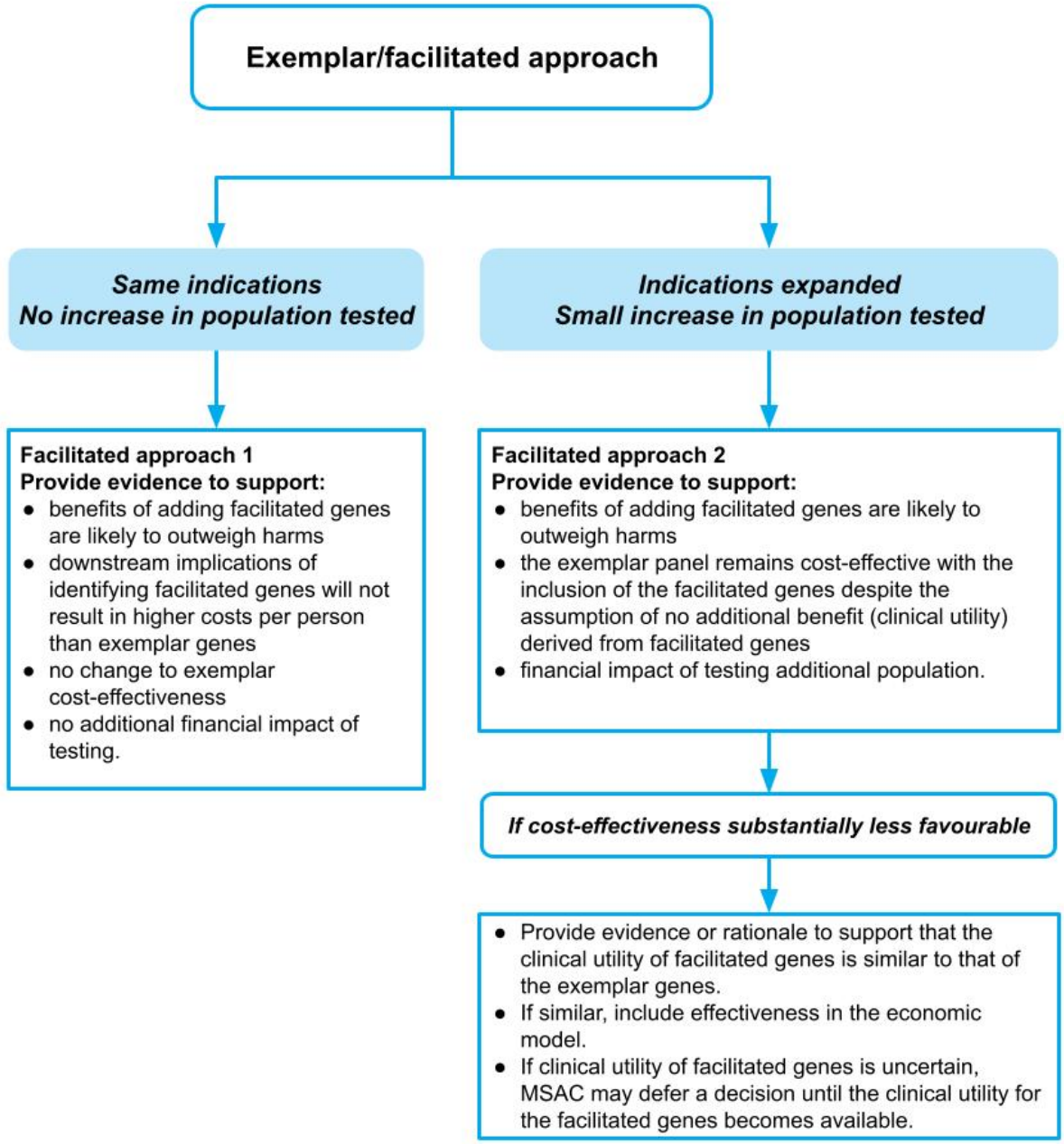


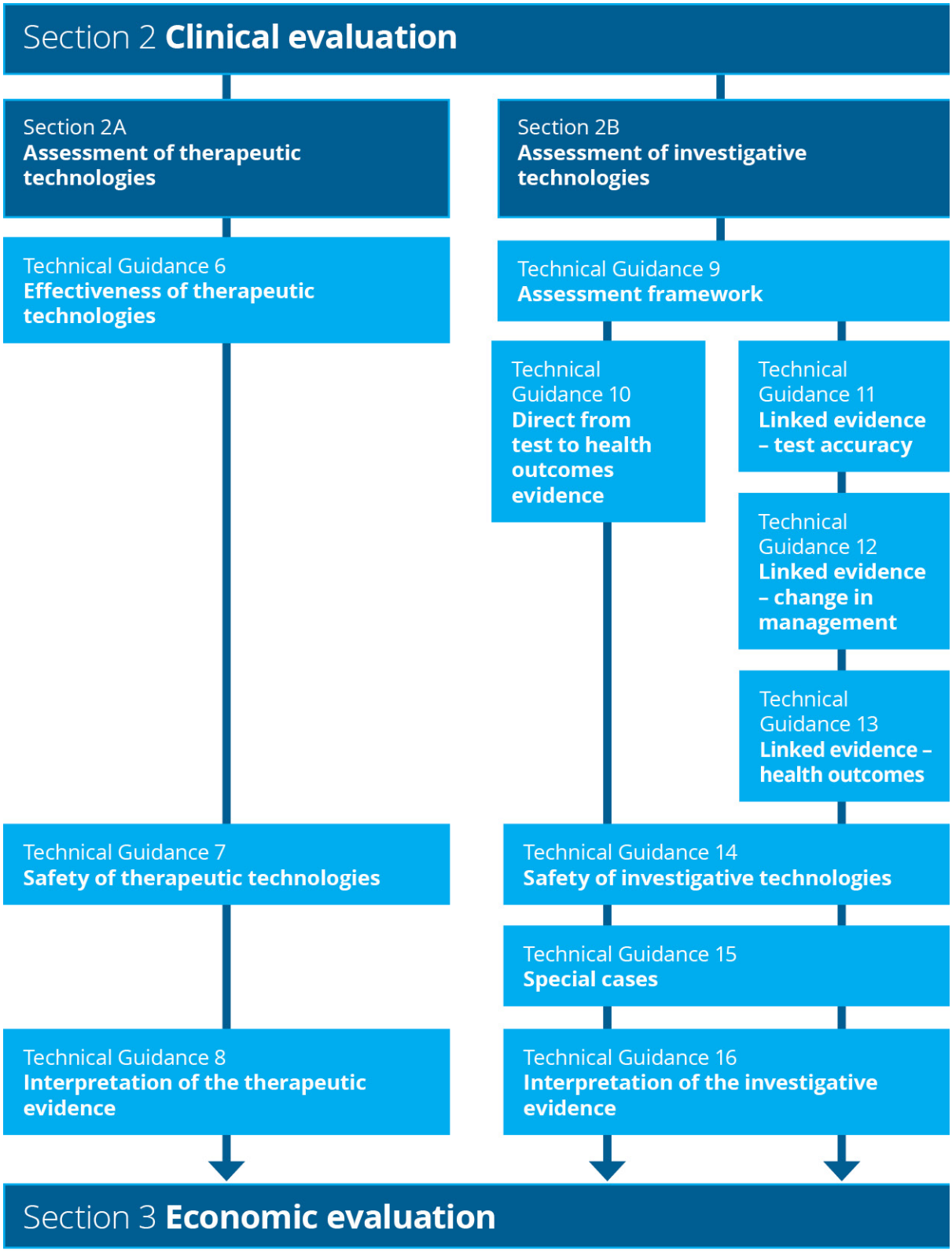
Figure 3 **Facilitated approaches, based on whether population is identical to, or slightly different to, exemplar listing**

Section 2 Clinical evaluation

Introduction

Section 2 provides guidance on evaluating the clinical evidence for the proposed health technology compared to the main comparator in the context of the requested listing. Clinical evidence for an application is derived from a literature search, critical appraisal and synthesis of the best available evidence.

The structure of Section 2 is outlined in Flowchart 2. Guidance for the clinical assessment of a health technology is given separately for therapeutic technologies (Section 2A) and investigative technologies (Section 2B). Guidance on methods that are common across both therapeutic and investigative technologies is presented in Appendix 2, Appendix 3, Appendix 4 and Appendix 5.



Flowchart 2 **Structure of the guidance for the clinical evaluation**

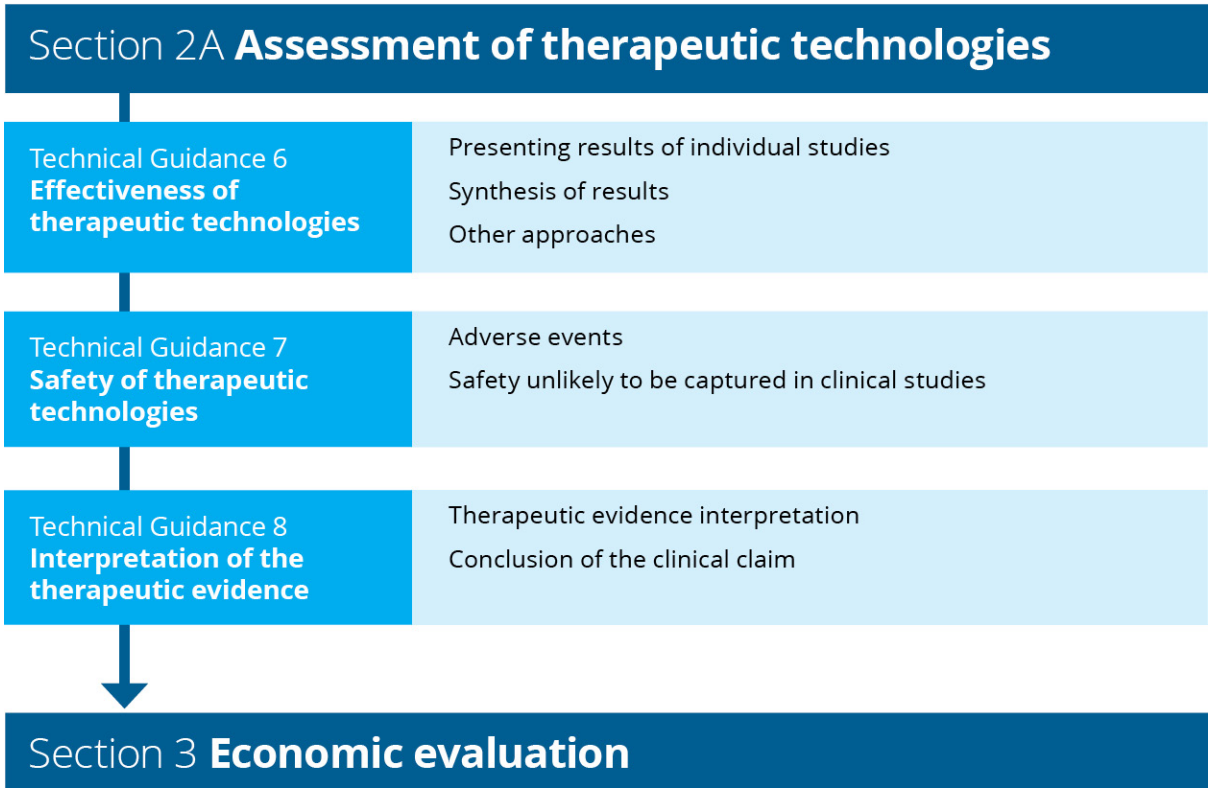
Section 2A Assessment of therapeutic technologies

Therapeutic technologies are a type of health technology that is expected to, or claimed to be able to, directly improve the health of people receiving it. Nothing else needs to be rendered to achieve the improvement in health outcomes. Examples of therapeutic technologies are devices, medicines, vaccines, procedures, programs or systems.

The assessment of a therapeutic technology is well established. The approach includes:

- a systematic literature search for relevant evidence (Appendix 2)
- an assessment of the risk of bias of the included evidence (Appendix 3)
- presentation of characteristics of the included studies (Appendix 5)
- an overall assessment of the certainty of the evidence for each outcome (Appendix 4).

The structure of Section 2A is outlined in Flowchart 3. The TG subsections in Section 2A describe the assessment and presentation of the effectiveness and safety of a therapeutic technology.



Flowchart 3 **Structure of the guidance for assessment of therapeutic technologies**

Technical Guidance 6 Effectiveness of therapeutic technologies

KEY CONSIDERATIONS

- **Present the results for individual studies numerically (tabulated) and/or graphically rather than as narrative summaries (TG 6.1).**
- **If appropriate, present a meta-analysis and/or subgroup analysis of the aggregated data (TG 6.2).**
- **Discuss the magnitude of the treatment effect (or pooled treatment effect, if appropriate) with respect to statistical heterogeneity and the strength of the evidence (TG 6.2).**
- **If an indirect comparison is necessary, discuss the likely effect of potential confounders, consistency of treatment effects and heterogeneity of any meta-analysis on the indirect estimate of effect (TG 6.3).**
- **If an adjustment for treatment switching is necessary, describe the mechanism for switching, the baseline characteristics of switchers vs nonswitchers, any differences between switchers and nonswitchers, and the reasons for switching (TG 6.3).**

The objective of including a systematic overview of evidence for a therapeutic technology in the assessment report is to efficiently synthesise and present the most relevant study results for decision-making. The ***assessment*** report should contain key results and discuss the results as they relate to the clinical questions identified in the PICO process. Extensive tables are placed in the technical report and referenced in the main body of the assessment report.

The structure used to present the study results can be adjusted depending on the quantity and nature of the evidence that needs to be presented. Typically, subheadings representing each outcome identified in the PICO process in descending order of patient relevance is an appropriate method for organising the results.

Study results presented in an assessment report should include the following:

- results from individual studies (e.g. trials, studies or meta-analyses identified in the literature search)
- meta-analysis of results (if appropriate)
- indirect comparisons (if appropriate)
- discussion of the quality and certainty of the included evidence
- summary of supplementary evidence (if appropriate)
- discussion of the overall evidence base in the context of the risk of bias (Appendix 3), quality (Appendix 4), confounding (Appendix 5) and applicability to the proposed target population (Technical Guidance 2 and Technical Guidance 3).

Present all information for each outcome separately to improve readability.

TG 6.1 Presenting results of individual studies

It is important to ensure that the results from individual studies are reported in addition to pooled estimates. Relevant details of individual study results include:

- the number of patients at risk (or the number that provide data)
- the number of patients experiencing the event (if appropriate)

- the percentage of patients with the event, and means (standard deviation) or medians (interquartile range) within groups, as appropriate
- confidence intervals (CIs) of the outcomes within groups
- relative and absolute differences between groups, and their CIs.

MSAC prefers the results to be presented numerically and/or graphically (e.g. as a forest plot) rather than narrative summaries, when a meta-analysis is logical and feasible. When it is not feasible or appropriate to meta-analyse individual studies, alternative numerical and/or graphical summaries should be compiled.

An example of results from a meta-analysis is presented in Figure 4. No pooled estimate is provided due to the heterogeneity of indications and endpoint definitions. An example of poor-quality data (within patient, before and after case series using different measures between studies) is shown in Figure 5. This is an example of a graphical presentation when data are not able to be meta-analysed.

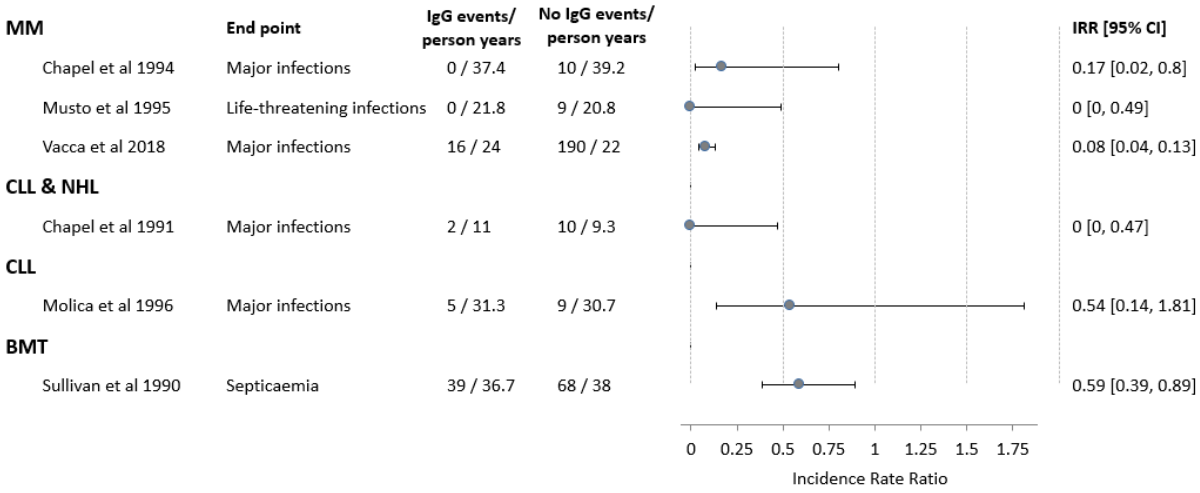
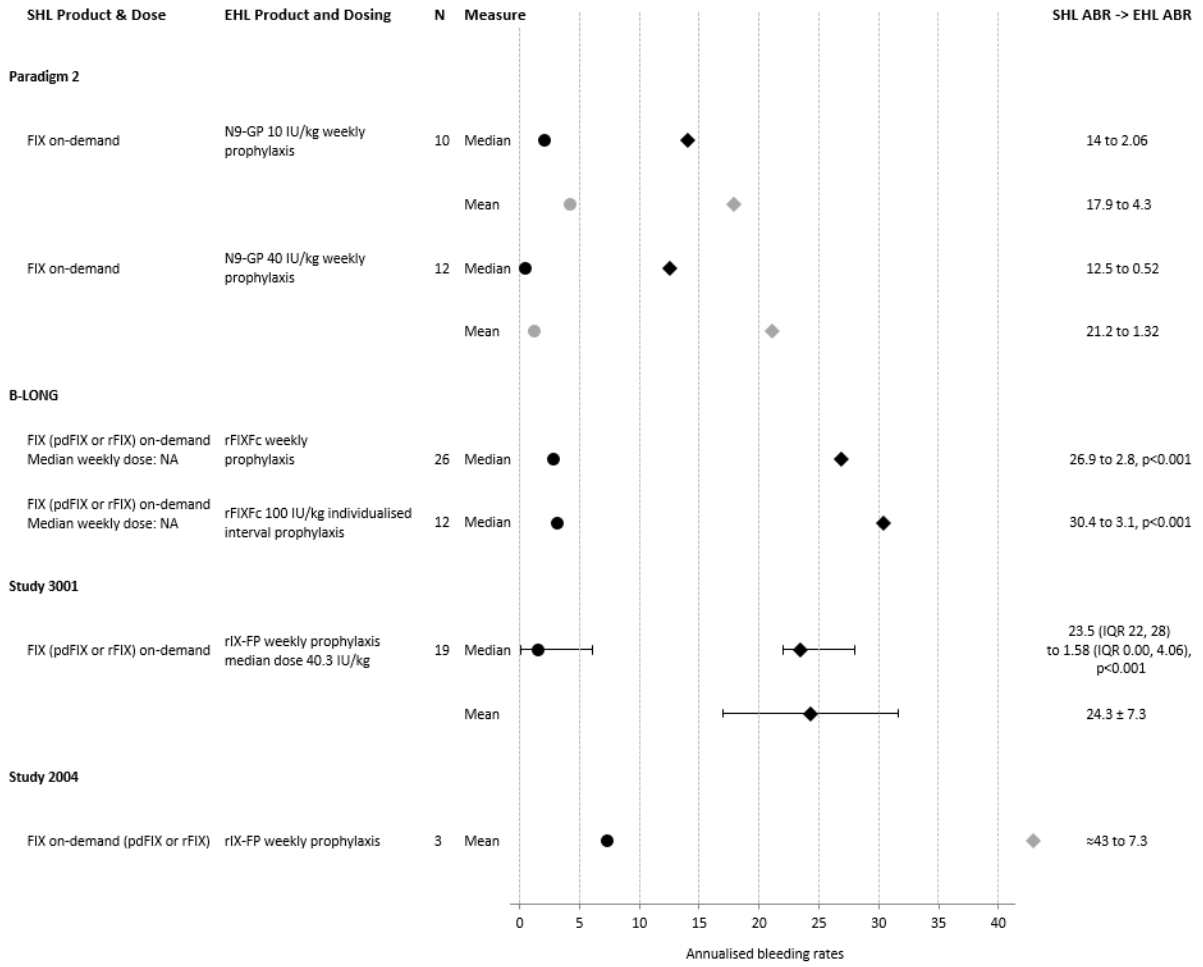


Figure 4 **Example of a forest plot showing effectiveness of intravenous immunoglobulin (IgG) at reducing serious infections in different subgroups of patients from included randomised controlled trials**



ABR = annualised bleeding rate; EHL = extended half-life; IQR = interquartile range; IU = international units; SHL = standard half-life

Shaded data are commercial in confidence.

Figure 5 Example of poor-quality data from studies comparing annualised bleeding rates after SHL on-demand versus EHL prophylaxis in adolescents and adults

The format of tables and figures used to present results will need to be adapted to the data available and the type of outcome.

In addition to individual study results, the following information may be relevant to present, to aid interpretation of tabulated data:

- the timing of the outcome assessment (e.g. change from baseline at 6 weeks)
- whether studies measured the same outcome at more than one time point (e.g. 6 weeks, 12 weeks, 3 months) – Discuss whether the treatment effect differs across other time points.
- a discussion about the appropriateness of thresholds used to translate continuous outcomes into dichotomous or categorical outcomes
- the statistical method used for any analysis – State the method used in a footnote to the table. Report any covariates used in the statistical analysis. If a statistical method adjusts for covariates, present the results of an unadjusted analysis in the table footnote. If required, discuss the appropriateness of the statistical analysis, or the impact of different studies applying different methods.

- whether any assumptions are required to support the statistical method (e.g. assumption of proportional hazards for a Cox proportional hazards model), and whether the assumptions have been tested – State whether there is a risk that the assumptions supporting a statistical approach are invalid.
- justification for the choice of any noninferiority margins that were applied.[9]

Dichotomous data

Dichotomous data are presented as the number of patients with the event in each arm, as a proportion of the total number of patients in the arm (i.e. n/N). For comparative studies, this permits a relative risk and a risk difference to be generated. This is the appropriate treatment effect to present if studies present a relative or absolute treatment effect adjusted for covariates.

Continuous data

Many studies measure a continuous variable at baseline and again at a prespecified time point. The treatment effect from such studies can be reported in several ways, including mean difference, weighted mean difference and analysis of covariance. Consider the appropriateness of pooling continuous data across studies if different covariates have been applied, or if time points differ.

Time-to-event data

Time-to-event data can be presented by studies in several formats. As well as presenting the individual study estimates for a time-to-event outcome, it is important to describe the method for analysing the time-to-event data. Many methods rely on underlying assumptions, which should be presented alongside results to aid interpretation.

Ordinal or categorical data

Attempt a similar approach to that described for continuous data if the trial results are available as ordinal or categorical data (e.g. a Likert scale for patient-reported outcome measures). Expert biostatistical advice will be helpful in such circumstances, particularly if meta-analysis is applied.

Multiattribute utility instrument data

Results from a multiattribute utility instrument (MAUI) are commonly reported across several time points. Report detailed MAUI results in the technical appendix. Report MAUI results for each time point and each arm within the study. Include the number of patients eligible to respond, and the number who actually responded, which is important for interpreting the results. Clearly state if compliance rates are not reported, as this undermines the confidence in the results.

In the main body of the report, the difference between the arms (with 95% CI) should be expressed as the integrals between the mean utility weights obtained over time up to the median (or other relevant time point) follow-up in the study. Explain alternative approaches used for comparing MAUIs.

State which scoring algorithm has been used to map the MAUI to utilities. Discuss the applicability to the Australian setting.

When providing an interpretation of the MAUI results, discuss the consistency or inconsistency with any concomitantly used patient-reported outcome measure in the same study.

Subgroup analysis

If only some of the participants from the whole study population are relevant to the target population, present a subgroup analysis to show the treatment effect of the proposed therapeutic technology in the relevant population.

Ensure that the participant characteristics and treatment details have been extracted (as per Appendix 5) for the whole study population and each of the relevant subgroups.

Provide the following information to support a subgroup analysis:

- clarification for why the proposed health technology should not be available to the patients in the complement of the subgroup, and why the study enrolled a broader population
- the plausibility of a variation in treatment effect for the subgroup, as it relates to the biological or clinical rationale for using the therapeutic technology. An unexplained variation is difficult to interpret in the absence of such an explanation (cross-reference to any discussion of biological plausibility that has been provided in the context section of the assessment report).
- whether the subgroup analysis was prespecified and whether randomisation was stratified by the subgroup. If the subgroup was defined using a threshold (such as a level of marker in the blood, or a severity score), justify the choice of the threshold. Discuss the impact of varying the threshold on the outcomes.
- the number of subgroup analyses originally conducted and any statistical adjustment for multiple comparisons.

Present the analysis of the treatment effect for the subgroup and compare this with the complement of the subgroup. Tabulate results side-by-side to improve readability. Test for interaction between the subgroup and its complement to support and quantify the association between the treatment effect and the covariate defining the subgroup.

In assessment reports that rely on a meta-analysis, subgroup analyses for individual studies may be presented in the technical appendix (and referenced); however, the information required to support a subgroup analysis should be provided in the main body of the report.

If a subgroup must be extracted for only some of the studies included in the assessment report, present a subgroup analysis prior to performing a meta-analysis. Use a random effects meta-analysis for pooling data, if feasible.

Presentation of nontrial evidence

Although randomised controlled trials are considered the best study design to minimise systematic error, if these are not available or not sufficient (e.g. because they are too small for replicability of results to be certain or they are not directly applicable to the target setting in Australia), other study designs should be presented. MSAC's preference for study designs is as per the hierarchy in the NHMRC levels of evidence for interventions.[10]

Treat real-world data according to the type of evidence they produce. For example, if registry data are able to provide information on both the intervention and comparator, a cohort study design may be appropriate; however, if data are only able to provide information on a single treatment, handle the study as a single-arm case series.

MSAC has a preference for comparative study designs (such as cohort studies with a concurrent or historical control) over noncomparative studies (such as case series). However, there will be situations where no comparative studies have been performed (such as for rare diseases, where trial

evidence is difficult to generate). Indirect comparisons are preferred over presenting only single arm data, which do not allow MSAC to consider comparative safety and effectiveness. Order the results by outcome measure and attempt to make a narrative comparison of the intervention and comparator. Conclusions based on an unadjusted indirect comparison would need to be tentative due to the high potential for the comparison to be influenced by factors other than the treatments being compared. Larger differences between the results for the intervention and the control provides greater certainty of the *direction* of effect, even if the size of effect is unclear.[11]

Modelling has suggested that confounding (from nonrandom allocation) is unlikely to explain associations with a relative risk of more than 2 or less than 0.5.[12] If there is no comparison treatment, compare the intervention against natural history (the trajectory of health outcomes in the absence of treatment).

If nontrial evidence is used, the following information should be presented:

- the quality of the source data (for both the proposed intervention and the main comparator)
- the methods for generating an estimate of comparative safety and effectiveness
- quantitative estimates of the comparative safety and effectiveness of the proposed intervention, and confidence or credible intervals
- the uncertainty in the estimates due to data quality or sources of bias, transitivity of the evidence across the data sources, and applicability of the sources to the Australian target population
- additional evidence or methods to reduce the uncertainty of the results.

TG 6.2 Synthesis of the results

An MSAC assessment report requires the synthesis of results from the available evidence. The appropriate approach will depend on the nature of the data.

Meta-analysis

If appropriate, present a meta-analysis of the aggregated data. If a meta-analysis is performed, state the software used and describe the methods. Select a method that is appropriate for the format of the outcome data (e.g. dichotomous, continuous, time-to-event). Document and reference the methods used so that they are reproducible and verifiable.

Present a forest plot that includes the estimate of the individual study treatment effects, and the pooled treatment effect. In a table note, report whether any studies that reported results for the relevant outcome were excluded from the meta-analysis, and explain the reasons for exclusion.

Discuss the methods used for pooling time-to-event data, or outcome measures that are derived using statistical approaches that control for covariates.

Report results for statistical heterogeneity (Cochran *Q* with a chi-square test for heterogeneity and the I^2 statistic). Discuss any heterogeneity identified in the meta-analysis with reference to the study characteristics (Appendix 5), outcome definitions (Appendix 5) and study design (Appendix 3). Use an appropriate method to test for the risk of publication bias and comment on the findings.

If multiple studies report on the same or similar outcome, but it is inappropriate to perform a meta-analysis, explain why. Describe the results narratively and nominate the most relevant studies on the basis of study quality and applicability.

The discussion of the synthesis should contain the following elements:

- a statement of the direction of the treatment effect
- an estimate of the magnitude and precision of the treatment effect
- discussion of the consistency of the treatment effect across studies
- discussion of the strength (certainty) of the evidence base.

The discussion for each outcome should reference the concerns identified relating to the search, risk of bias and study characteristics. In addition, discuss the applicability of the study populations or their characteristics to the target Australian population.

Magnitude and direction of treatment effect

Discuss the estimate of the magnitude of the treatment effect (or pooled treatment effect, if appropriate) in the context of its clinical relevance (minimal clinically important difference). If relevant, discuss the comparative estimate of effectiveness in the context of a nominated noninferiority margin.

Comment on the consistency of the treatment effect across key subpopulations (e.g. by patient or disease characteristics), if available.

Narrative synthesis

If studies are too heterogeneous to meta-analyse, the results should be synthesised narratively. Incorporate the same elements as for the discussion of the synthesis from a meta-analysis, as listed above.

If the evidence is poor, discuss whether better data are likely to become available in the future, or whether factors such as the rarity of the condition, or the diffusion of the intervention and its acceptance as 'best practice' will prohibit (future) trials from being performed.

Description of strength of the evidence

After presenting individual study results and meta-analysis (if appropriate), discuss the precision, consistency, directness (applicability of the population, intervention, setting, comparator) of the comparison, and the risk of publication bias across the evidence base (per outcome). For each outcome, discuss the overall strength of the evidence base. The use of a GRADE table is optional. Do not present extensive GRADE tables in the assessment report.

TG 6.3 Other approaches to presenting evidence

Indirect comparison

An indirect comparison may be appropriate if no direct randomised controlled trials are identified in the systematic literature search. Identify relevant studies following the guidance provided in Appendix 11. If more than one randomised trial has the same intervention and a common reference, separately pool the treatment effect before performing the indirect comparison.

Describe the method(s) used for the indirect comparison, such as the Bucher single pairwise method,[13] matching-adjusted indirect comparison,[14] simulated treatment comparison,[15] network meta-analysis or mixed treatment comparison.

More complex methods, such as network meta-analyses, may be presented as supplementary analyses. Where there are multiple common comparators in the network, present the results of

pairwise comparisons for each link in the network. Although some methods consider nonrandomised studies in a network, avoid including nonrandomised studies if possible. Where nonrandomised studies must be included, present the results of the network meta-analysis both with and without the nonrandomised studies.

Avoid unadjusted indirect comparisons (such as naive comparisons between single arms), or indirect comparisons where differences in trial characteristics may affect the transitivity of the trials in the comparison. These comparisons are difficult to interpret and reduce the confidence of MSAC in decision-making. Where patient-level data are available for at least one study in the comparison, use matching-adjusted indirect comparisons or simulated treatment comparisons to correct for trial differences and improve the transitivity of the comparison.

When considering complex approaches (e.g. matching-adjusted indirect comparisons, simulated treatment comparisons, network meta-analyses, mixed treatment comparisons), balance the additional information requests and challenges these approaches may present with any reduction in uncertainty they may deliver. In the technical report, provide sufficient detail to allow the analysis to be repeated. Provide an explanation of the method, the statistical code (including programming code for statistical software such as Stata, R, SAS or WinBUGS), the assumptions required for the approach used, and how the assumptions were validated. For methods that require individual patient data (matching-adjusted indirect comparison or simulated treatment comparison), attach these data in a spreadsheet. In the main body, compare the results with those derived from a simple indirect comparison method, and explain any difference.

When presenting the results of an indirect comparison, include the following:

- the number of studies included in the indirect comparison, and whether any studies identified in the systematic literature search were excluded (and why)
- an assessment of the balance of potential confounders across arms in individual trials
- an assessment of the heterogeneity of studies included in the analysis
- a comparison of the event rates across the common reference arms of pairwise comparisons. Discuss the implications of differences in the event rates. If event rates indicate a difference in baseline risk across trials, discuss whether the relative treatment effects are consistent with baseline risk.
- justification for the choice of outcome measure (e.g. odds ratio, relative risk, absolute risk difference). The choice of outcome measure should minimise the variation in the comparative treatment effect within each and between all sets of included randomised trials.
- conversion of the indirect estimate of relative treatment effects to an absolute risk difference, where a relative outcome measure is nominated and the desired outcome is an absolute risk difference
- the indirect estimate of effect presented as relative risk and/or odds ratio (or the ratio of hazard ratios) with its 95% CI (or if previously justified, the absolute risk difference)
- in cases where trials have been excluded, sensitivity analyses in which these trials are included (if possible).

Adjustment for treatment switching

Adjustments to correct for the influence of treatment switching on the treatment effect may rely on assumptions that are difficult to validate. Evidence without treatment switching is preferred.

In circumstances where participants in the control arm of the included study switch and receive the proposed therapeutic technology, it may be reasonable to statistically adjust the treatment effect to remove the effect of the proposed therapeutic technology on subsequent endpoints. If switching to the proposed therapeutic technology reflects clinical practice, the appropriate comparator would include subsequent treatment with the proposed therapeutic technology, and no adjustment is necessary.

If an adjustment for treatment switching is necessary, describe the mechanism for switching. If switching (or the extent of switching) does not reflect clinical practice, provide the following:

- baseline characteristics of switchers vs nonswitchers (and discuss the differences)
- reasons for switching.

Several methods are available for adjusting the survival estimate for treatment switching.[16] Using simple methods may be acceptable when the adjusted estimate of the comparative treatment effect is clearly toward the null. If complex methods are used, provide details of the approach used, the assumptions made (and how they have been tested), and a comparison across more than one method.

Provide a discussion and interpretation of the results.

Where there is a largely uncontaminated estimate of an outcome that occurred before switching, discuss whether the outcome is a valid surrogate, and translate the surrogate to the final outcome.

Combining an adjustment for treatment switching with the use of subgroups or indirect comparisons will result in a high degree of uncertainty and should be avoided. If such an approach is necessary, ensure that the results of any analyses are unlikely to overstate the benefit of the proposed therapeutic technology.

Technical Guidance 7 Safety of therapeutic technologies

KEY CONSIDERATIONS

- **Report key adverse events from each included study, separately and, where appropriate, as a meta-analysis, and discuss the implications for any disparities across groups (TG 7.1).**
- **The absolute number of adverse events in each category may be a more appropriate estimate for the economic or financial analysis (TG 7.1).**
- **If the included evidence is not sufficient to capture long-term or rare adverse events, or adverse events in patients with comorbidities or receiving concomitant treatments, an extended assessment of the safety of new therapeutic technologies is required (TG 7.2).**

An assessment of the impact of a health technology on health outcomes includes an assessment of relative safety versus the main comparator. The assessment of safety has 2 key parts for therapeutic technologies:

- assessment of the direct and more immediate impacts of use of the health technology (often captured to a varying degree in the included clinical studies)
- assessment of longer term or rarer safety events unlikely to be captured in clinical studies.

In some cases, safety outcomes, or harms, can be difficult to distinguish from effectiveness outcomes. For example, one of the advantages of laparoscopic surgery compared with open surgery may be reduced blood loss, and because blood loss is an established complication of surgery, it may be regarded in some studies as an effectiveness outcome. Where such outcomes represent the key outcomes from clinical studies, they may be presented alongside effectiveness outcomes. Guidance for presenting effectiveness outcomes is provided in Technical Guidance 6.

TG 7.1 Adverse events

Identify the key safety events in the included studies, and determine whether any important safety events have been omitted. If the omission is related to poor reporting, additional study evidence will be required. If the omission is related to the rarity of the safety outcome, or insufficient follow-up in the clinical study, refer to the guidance on the extended assessment of safety in TG 7.2.

When reporting safety from a clinical study, as a minimum, the following categories of adverse events should be considered:

- any adverse event
- any adverse event resulting in discontinuation of the treatment
- any serious adverse event
- any adverse event resulting in death
- each and every other type of adverse event where the frequency or severity differs substantially across groups.

Where additional adverse events are to be reported (e.g. treatment-emergent adverse events, adverse events of special interest), explain the importance of the adverse event and interpret the result.

Report adverse event data as both the number of patients reporting an adverse event in each category and the absolute number of adverse events in each category. The absolute number of

events in each category may be a more appropriate estimate for costing adverse events in an economic or financial analysis because the number of patients who experience an adverse event will not capture patients who experience 2 events in the same category.

For each important adverse event, present the results as for dichotomous data, and include relative risks and risk differences with their 95% confidence intervals across the groups. Present each study separately. Where appropriate, meta-analyse the results using a random effects model and provide an interpretation.

Analyse the relative adverse event rates (events per period at risk), if the average period at risk per participant varies substantially between treatment groups (e.g. using a straight Poisson regression or a negative binomial approach). Present the assumptions associated with statistical analyses and how they were tested.

TG 7.2 Safety unlikely to be captured in clinical studies

The assessment of safety beyond clinical studies is necessary for new therapeutic technologies, or therapeutic technologies used for a new indication. This assessment may therefore be relevant for assessment reports of investigative technologies if they lead to a change in management that involves a therapeutic technology.

Ideally, the estimate of the relative safety of a health technology is derived from high-quality comparative studies. Clinical trials are often inadequate for providing data on comparative harms for several reasons:

- Trials tend to enrol patients who are healthier, or have fewer comorbidities or concomitant medications, and they have more stringent monitoring than would occur in the target population.
- Trials are usually underpowered and of insufficient duration to detect long-term or rare adverse events.
- Adverse events in clinical trials designed to emphasise efficacy results are often underreported.[17]

Discuss whether the included evidence base is adequate for identifying:

- less common adverse events or safety concerns
- adverse events that may occur in the longer term
- harms that may occur due to differences between the target population and the more selected population that may be enrolled in a clinical trial.

If the included evidence is not sufficient to capture long-term or rare adverse events, or adverse events in patients with comorbidities or receiving concomitant treatments, present additional evidence. Describe the search strategy for identifying nonrandomised studies of the proposed health technology, or registry data. If appropriate, include evidence for safety of the proposed health technology in other indications. Where the proposed health technology is delivered in combination with an implantable device, provide an assessment of the safety of that device. Sources of safety information may include device registries, regulatory databases, complaints registries and post-market surveillance studies.

Technical Guidance 8 Interpretation of the therapeutic evidence

KEY CONSIDERATIONS

- **Summarise the overall evidence base, taking into account the level, quality and statistical precision of the evidence, the consistency of the results across the clinical studies and across subgroups, the strength or certainty of the evidence, and its applicability to the Australian setting (TG 8.1).**
- **Summarise any other uncertainties in the evidence, and other relevant factors that may have an influence on decision-making, particularly implementation and ethical factors (TG 8.1).**
- **The conclusion of the clinical claim of the therapeutic technology should be a simple and unequivocal statement about the superiority, inferiority or noninferiority of the intervention to the comparator that is supported by evidence provided in the submission (TG 8.2).**

The objective of summarising the overall evidence base in the assessment report is to describe the results as they apply to the clinical claim in the specific context of the Australian setting.

TG 8.1 Therapeutic evidence interpretation

Provide a summary of the overall evidence base (without repeating evidence from other sections). Consider:

- the level of the evidence, taking account of the directness of the comparison
- the quality of the evidence
- the clinical importance and patient relevance of the effectiveness and safety outcomes
- the statistical precision of the evidence
- the size of the effect
- the consistency of the results across the clinical studies and across subgroups
- the strength or certainty of the evidence
- the applicability of the evidence to the Australian setting
- any other uncertainties in the evidence, including missing outcomes or populations
- other relevant factors that may influence decision-making, particularly implementation and ethical factors.

TG 8.2 Conclusion of the clinical claim

The interpretation of the clinical data presented in Section 2 is crucial in determining the success of the submission. It is important to classify the health outcomes of the proposed health technology in relation to its main comparator, that is, whether it is superior, inferior or noninferior to the comparator.

The conclusion of the clinical claim for a therapeutic technology should be a simple and unequivocal statement that is supported by evidence in the submission.

Example:

The use of [proposed health technology] results in superior/noninferior/inferior effectiveness compared with [comparator].

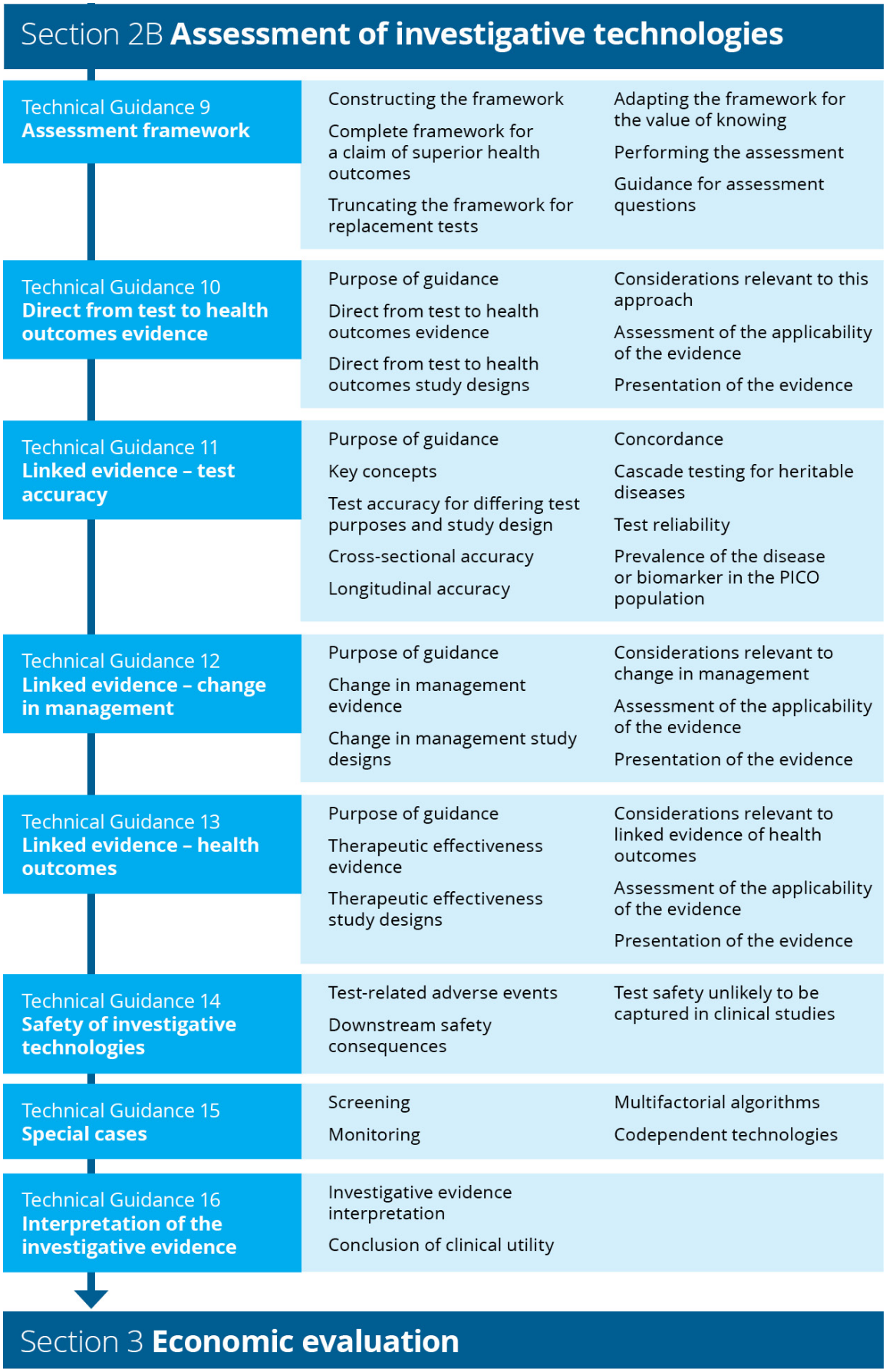
The use of [proposed health technology] results in superior/noninferior/inferior safety compared with [comparator].

Section 2B Assessment of investigative technologies

Investigative technologies are a type of health technology that is expected to, or claimed to be able to, generate clinically relevant information about the individual to whom the service is rendered. To achieve an improvement in health outcomes, this information must result in a change in the clinical management of the individual. In this sense, investigative procedures can only indirectly improve health outcomes. Examples of investigative technologies are imaging, pathology, genetic testing, and clinical assessments for diagnosis, prognosis, staging, monitoring, prediction of treatment response, surveillance and cascade testing. For ease of reading, the word 'test' is used throughout the document as an alternative term for 'investigative technology', and is intended to reflect the broad range of investigative technologies available.

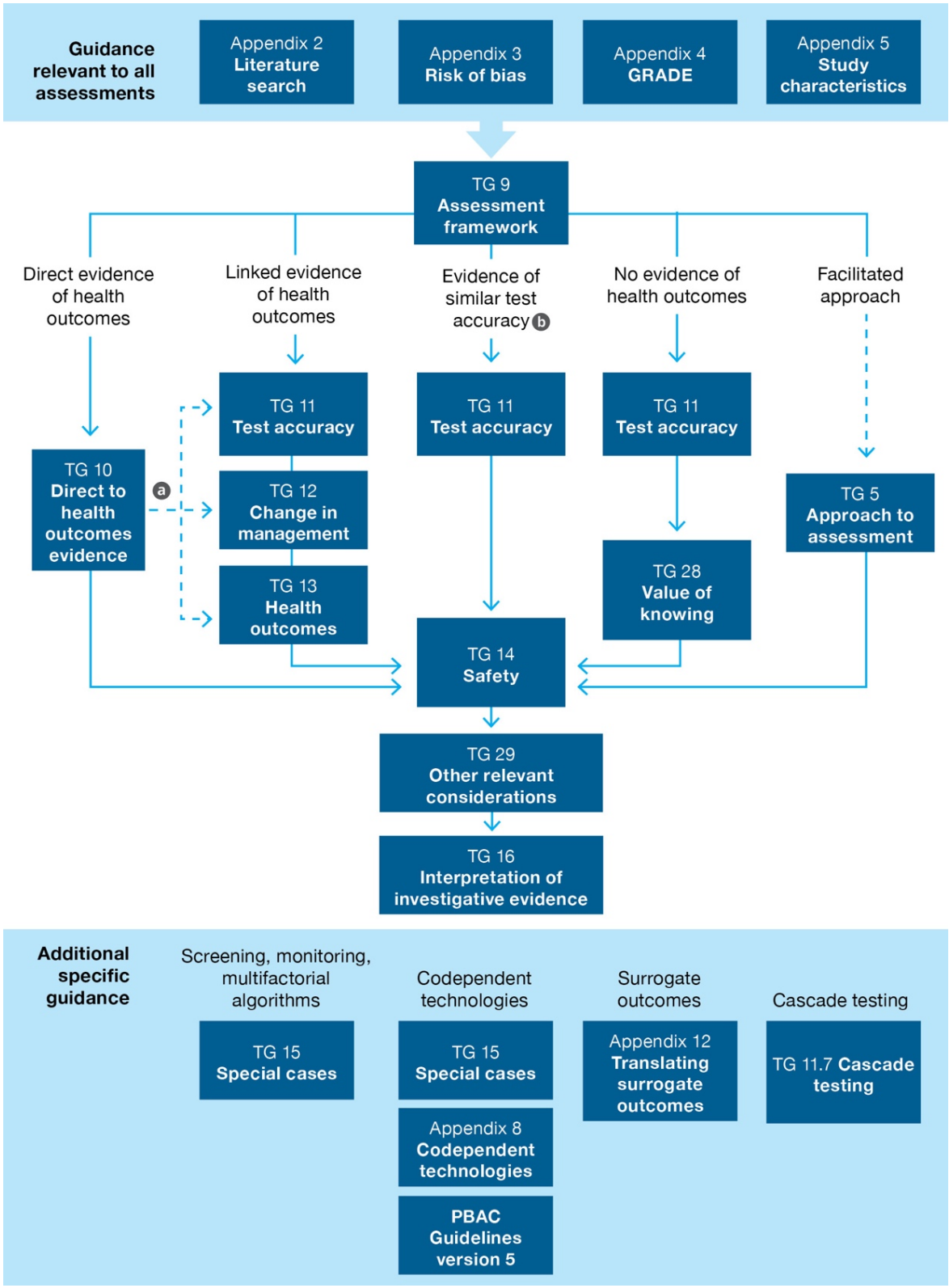
The clinical component of an assessment of a health technology determines whether the technology is inferior, noninferior or superior to its comparator in terms of health outcomes. In some circumstances, the assessment will incorporate other types of benefits (value of knowing) or other considerations. The method for assessing the clinical claim for tests will vary according to the type of evidence that is available. The available evidence, and the purpose of the test, will also determine which TG subsections in these guidelines are relevant to for the assessment.

The structure of Section 2B is outlined in Flowchart 4.



Flowchart 4 **Structure of the guidance for assessment of investigative technologies**

Figure 6 describes the TG subsections that would be relevant for different approaches taken in an assessment of an investigative technology. Technical Guidance 9 explains the types of evidence that may be available or that will be required.



a Linked evidence guidance may be relevant to technologies with direct from test to health outcomes evidence if the applicability of the test, management decisions or treatment outcomes is in question.

b An assessment that relies on a comparison of test accuracy alone is permissible only if this comparison is against a currently reimbursed comparator.

Figure 6 **Navigation of the clinical components of the guidelines for investigative technologies based on the type of evidence required to assess the clinical claim**

Technical Guidance 9 Assessment framework

KEY CONSIDERATIONS

For investigative technologies

- **An assessment framework is required to determine the evidence required to verify the clinical claim linking a test result with health outcomes (TG 9.1).**
- **A complete assessment framework is required for a claim of superior health outcomes, but the assessment framework may be truncated or adapted for claims of noninferiority or for value of knowing (TG 9.2–9.4).**

The link between a test and health outcomes is rarely clear. The introduction of tests may have limited or no impact on health outcomes, and there may not be a reliable link between the accuracy of a test and the magnitude of health outcomes.[18] For a test to have an impact on health outcomes, it must inform a sequence of actions. The method adopted in these guidelines to describe each step from testing to health outcomes is the 'assessment framework'.

An assessment framework is required for the evaluation of investigative technologies.

This TG subsection describes the method for performing an assessment of the health impacts of a test. The actual assessment of a test is described in subsequent TG subsections. However, additional components required for a health technology assessment are not included in the assessment framework, including other relevant considerations (ethics, social, organisational, patient or consumer input), health economics and financial implications.

If the technology can be used for different purposes (e.g. diagnosis and monitoring), more than one assessment framework may be required.

TG 9.1 Constructing the assessment framework for a health technology

The assessment framework describes the evidence required to verify the clinical claim, and to support the economic approach. The purpose of creating an assessment framework is to help ensure that the assessment provides all the information that is useful to MSAC, without providing unnecessary information. The assessment framework is not a management algorithm, although for simple testing strategies it may resemble one. The structure of the framework describes conceptually the steps between the target population and the final health outcomes. The framework used will be influenced by whether the test is an add-on test or a replacement test (complete or partial). An add-on test will have additional costs, which must be justified by either superior health outcomes or strong logic for value of knowing, while a replacement test may be able to use a truncated framework.

The concept of an assessment framework is based on the process used by the US Preventive Services Task Force (USPSTF) to guide their assessments of preventive services and health promotion.[19] Many of the USPSTF services have been tests or screening programs; however, the frameworks are flexible and can be adapted to represent the steps between any intervention and the consequent health outcomes. The use by MSAC of the 'linked evidence approach' is consistent with these assessment frameworks.

The general structure of an assessment frameworks is a diagram that includes populations, outcomes (or information derived at each step), and actions or inferences that link the boxes. The diagram is accompanied by annotated questions corresponding to items in the diagram.

The number of steps in a framework reflects the number of actions or inferences that need to be taken between the decision to test and the final health outcomes. The components of the framework are consistent with the PICO elements described in the PICO confirmation, or as amended in Section 1 of the assessment report.

As a minimum, the initial framework should contain (letters refer to Figure 7):

Ⓐ a brief description of the test population

Ⓑ the name of the test

Ⓒ a link between the test and the final health outcomes (direct from test to health outcomes evidence)

Ⓓ the information provided by the test

Ⓔ change in management (change in subsequent testing and/or treatment)

Ⓕ outcomes (if surrogate outcomes are included in the PICO, provide these as a step prior to final outcomes)

Ⓖ adverse events associated with actions (the test or subsequent management decisions).

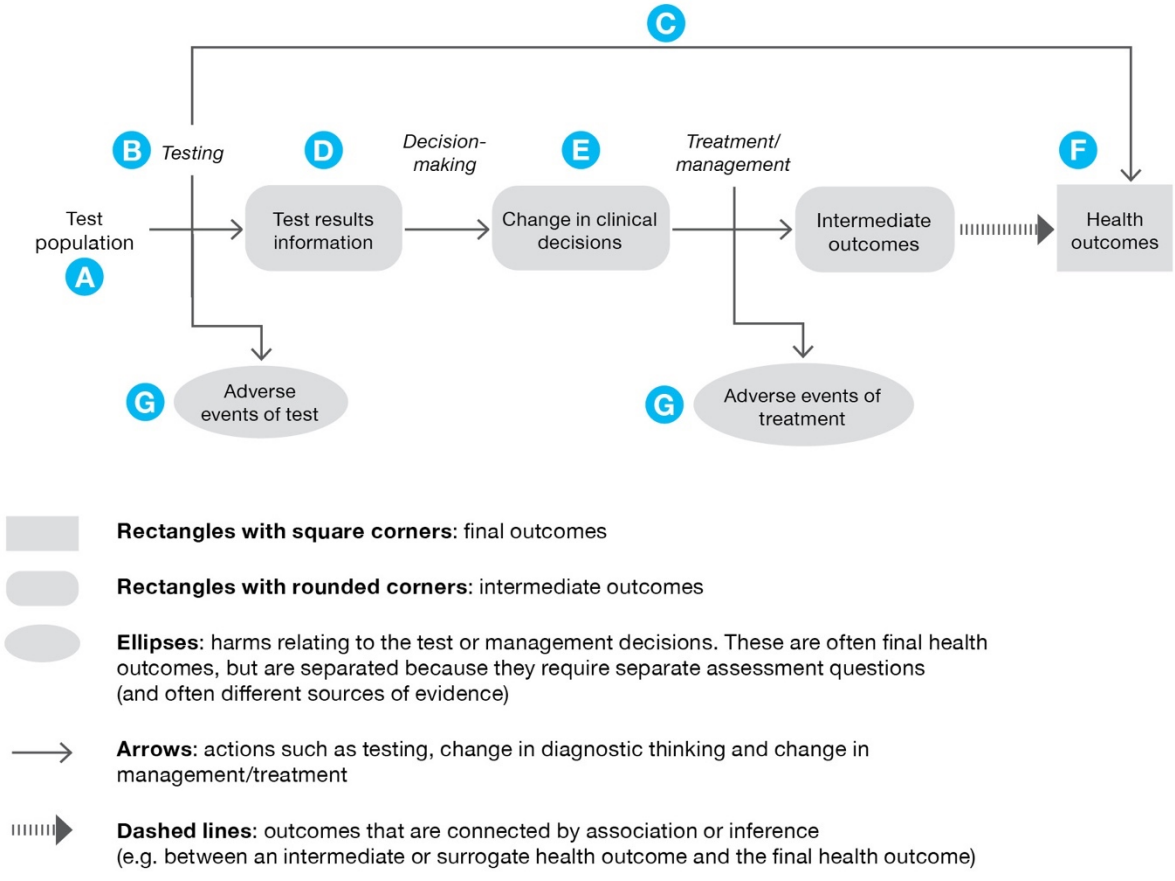


Figure 7 Components of an assessment framework (adapted from the USPSTF procedure manual 2015)[19]

Once the initial framework has been constructed, add in any additional steps that occur between the point of testing and final outcomes. Common inclusions are subsequent tests, particularly if the initial test is a screening test. Lines are drawn between actions and outcomes, with lines that omit steps representing more direct linkages.

Annotated questions may be derived from any part of the assessment framework; however, questions are necessary for each of the arrows that link the different steps. Assessment questions explicitly request evidence that compares the proposed test with the comparator. Each link is numbered to enable easy reference to the assessment questions.

Additional components that may be relevant include nonhealth outcomes (such as ethical implications, efficiency outcomes or value of knowing outcomes).

TG 9.2 Complete assessment framework required for a claim of superior health outcomes

An example of a generic assessment framework is given in Figure 8 and may be adapted as required.

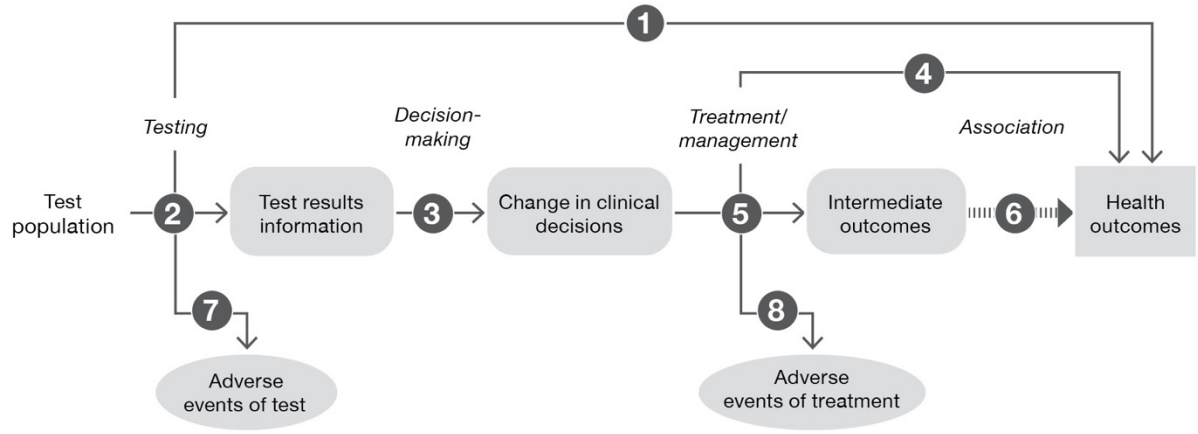


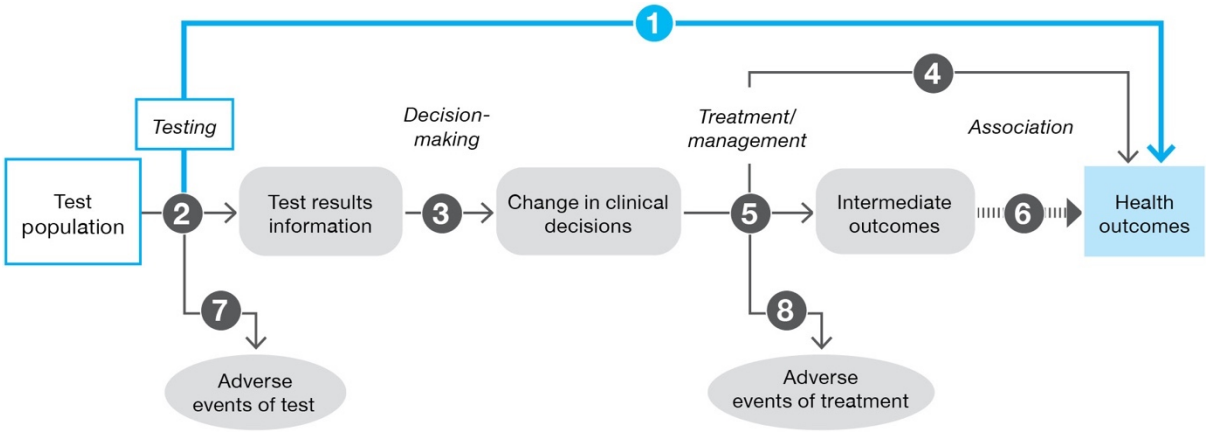
Figure 8 **Generic assessment framework showing links from test population to health outcomes**

Assessment questions for a claim of superiority (Figure 8)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

1 Does the use of the test strategy in place of the current test strategy (comparator) result in the claimed superior health outcomes?

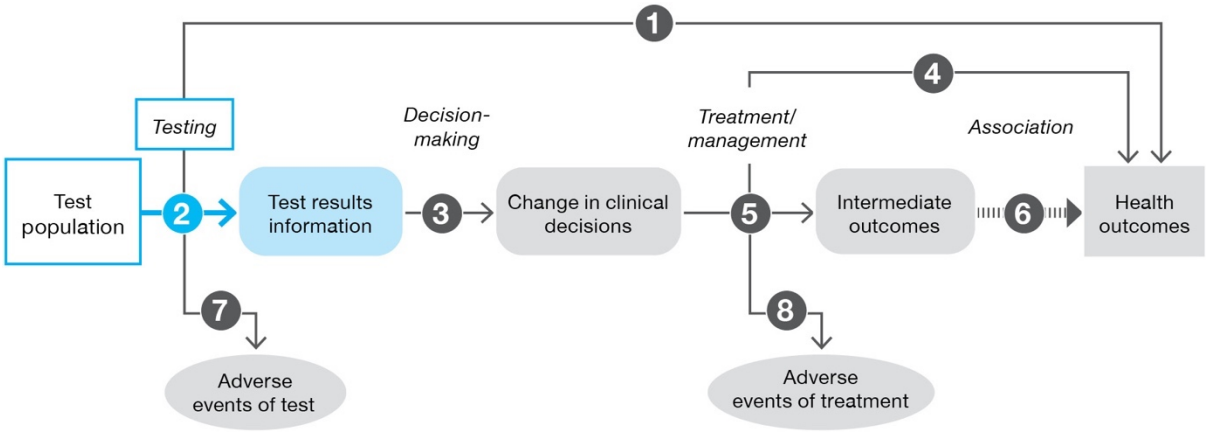
If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard?



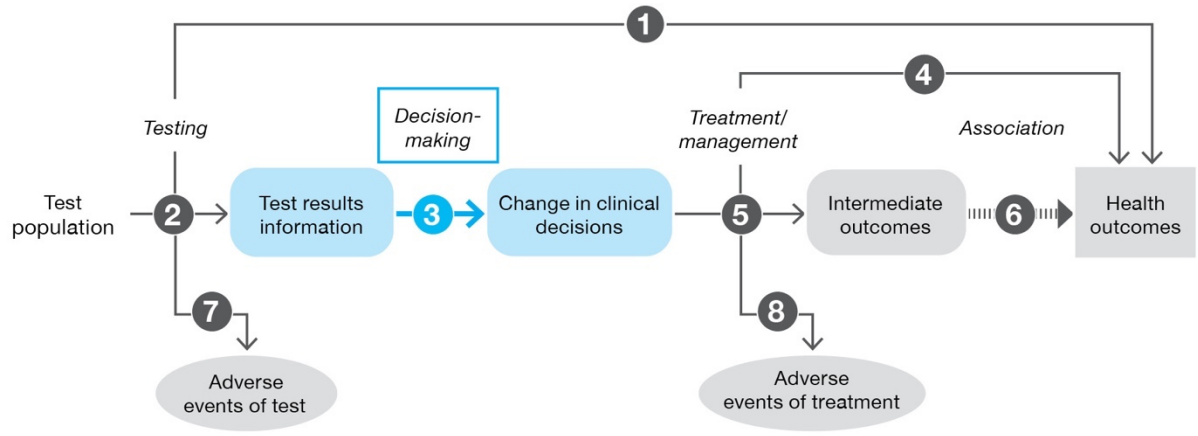
LINKED EVIDENCE

2 How does the information from the proposed test differ from that of the comparator? What is the concordance of the findings from the proposed test relative to the comparator? What is the accuracy of the proposed test (against a relevant reference standard) compared with the comparator?

If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard?

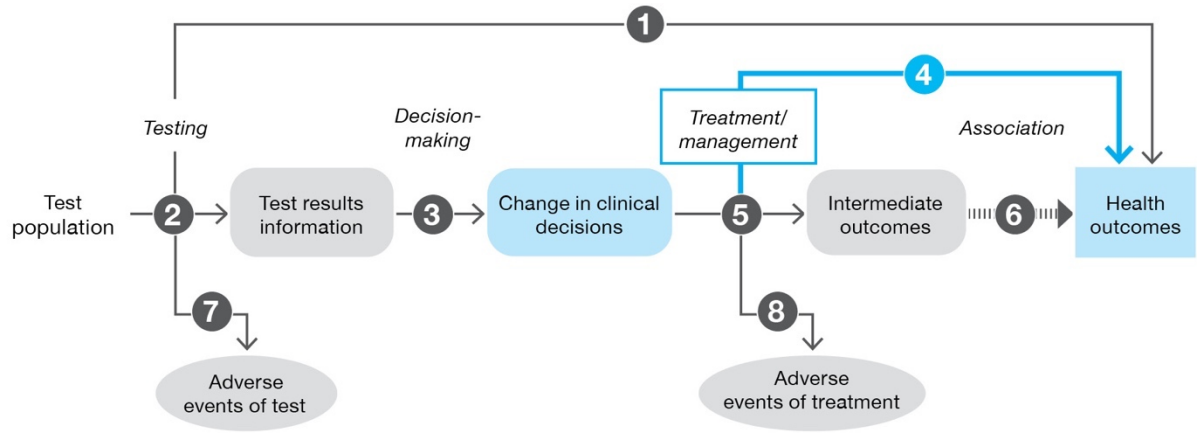


3 Does the availability of new information from the proposed test lead to a change in management of the patient (compared to the information gained from the comparator)?

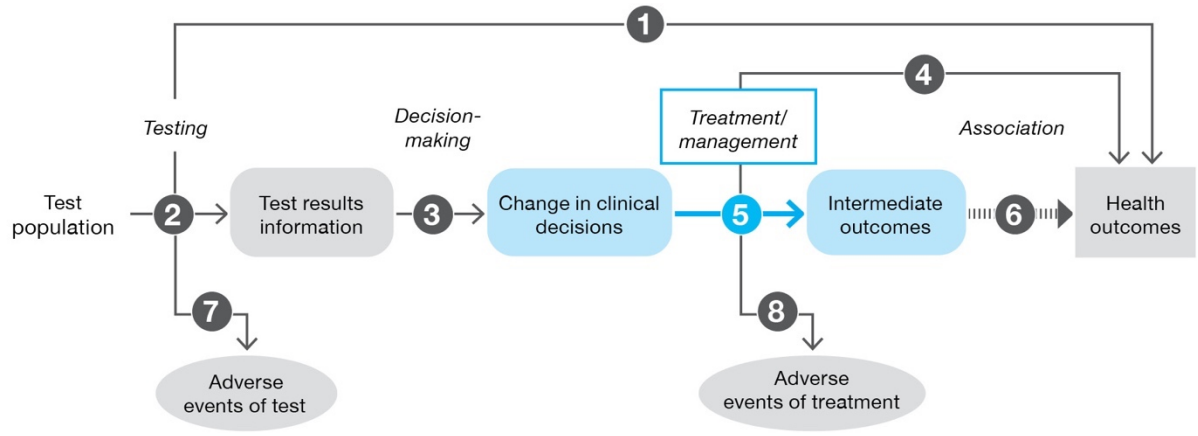


4 Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention), result in the claimed health outcomes?

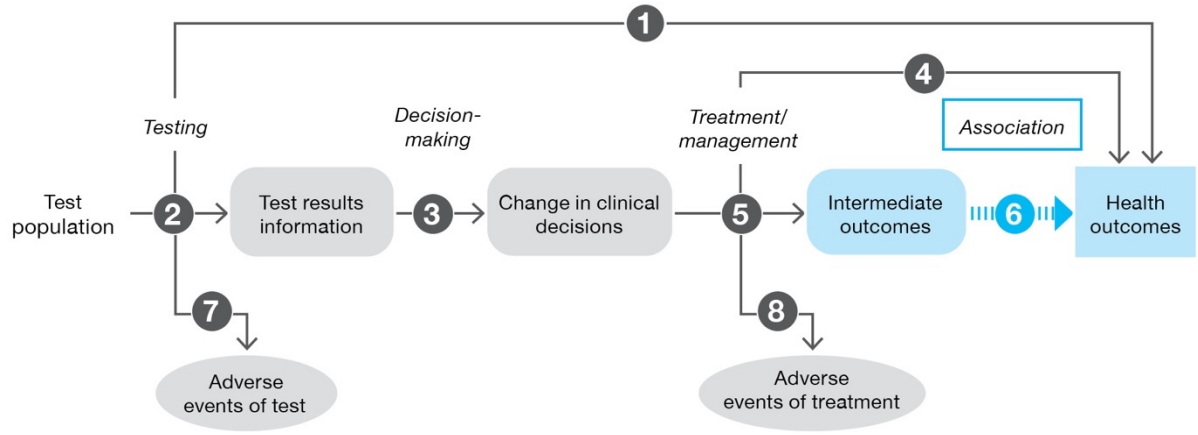
Has the treatment/management been provided to a population with the same spectrum of disease that the proposed test identifies? Is it biologically plausible that the treatment/management will be as effective in the population with this spectrum of disease? *When the proposed test detects more patients than the comparator test, it is likely that it is detecting earlier or less severe disease. The established treatment effect evidence (based on the comparator test) may be for a more narrowly defined (usually higher risk) positive population for whom the expected benefits may differ from those in the population identified using the proposed test.*



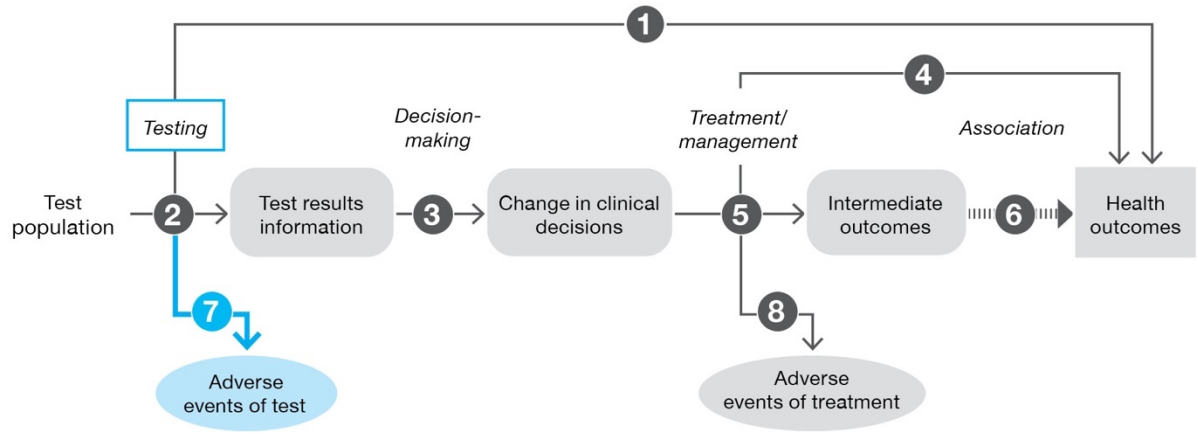
5 Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention), result in the claimed surrogate outcomes?



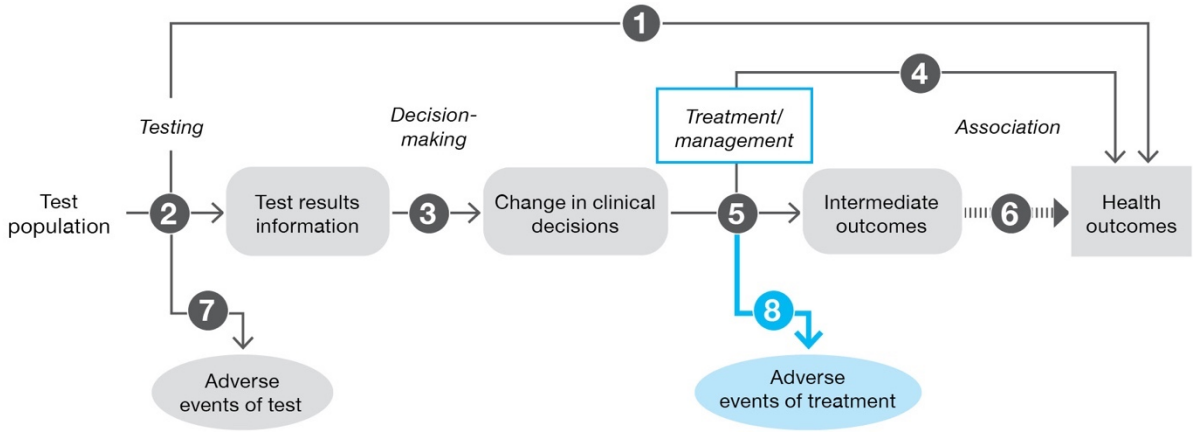
6 Is the observed change in surrogate outcomes associated with a concomitant change in the claimed health outcomes, and how strong is the association?



7 What are the adverse events associated with the proposed test strategy and the comparative test strategy?



8 What are the adverse events associated with the treatments/interventions that lead from the management decisions informed by the test and by the comparator?



Not all steps in the assessment framework need to be presented. For example, if there is adequate direct from test to health outcomes evidence to support the assessment, presentation of linked evidence may not be necessary. If there is evidence of the effect of treatment on final health outcomes, the step including surrogate outcomes may be removed.

Even though the availability of evidence for any step of the framework may be uncertain, these steps should not be removed from the framework. In the assessment it should just be made clear that evidence was not available to address the assessment question. However, not all steps will need to be explored if shorter paths to health outcomes are possible.

Additional relevant questions could relate to subgroups in the population, the prevalence of a disease, concordance of the tests used in Australian practice against the clinical utility standard, feasibility/efficiency of the testing procedure, value of knowing, ethical issues etc.

See Appendix 1 for guidance on adapted assessment frameworks for more definitive tests; tests for monitoring; multifactorial, black-box and self-learning algorithms; and universal screening tests.

TG 9.3 Truncating the framework for replacement tests

The generic framework (Figure 8) identifies the steps between the test and the final, patient-relevant health outcomes. When the proposed test replaces a current test, the full linked evidence approach can be shortened in some circumstances. In this case, it would be appropriate to make a claim of noninferiority (the proposed test results in noninferior health outcomes compared with the comparator test).

For example, if a new test is proposed to replace an existing test that is listed on the MBS, the question is whether the proposed test is equivalent to its predecessor (in regards to, for example, analytical performance, diagnostic accuracy, interpretability of results, feasibility and cost).

The full linked evidence approach cannot be shortened if:

- the proposed test is an add-on test (a new test or used in addition to current testing) such that it is expected to provide additional/different information

- the proposed test is claimed to result in different health outcomes (superior or inferior)
- there is insufficient evidence to support a claim of noninferiority based on a shortened assessment framework – This is particularly relevant for triage tests where there may be a lack of information on the health consequences of being ruled out from subsequent confirmatory testing.[20]

In these circumstances, evidence of the impact of the test on health outcomes is required to substantiate the clinical claim and the magnitude of the difference in health outcomes, and to inform the economic analysis.

If the clinical claim is that the test is noninferior in terms of health outcomes, the framework may be truncated in some circumstances.

Same test accuracy

For claims of noninferior health outcomes, where the claim is based on the proposed test providing the same information as the comparator, the approach *may be* reduced to a comparison of the information provided by the test and by the comparator.

If the proposed test reports on the same parameter, the concordance of the proposed test and the main comparator is required. If the tests are concordant, it may be reasonable to infer that there would be no difference in management, and health outcomes would be noninferior. Where the accuracy of the comparator test is uncertain or poor, concordance should be accompanied by test accuracy derived for both tests against an appropriate reference standard.

A more likely scenario is that the proposed test will not be absolutely concordant with the comparator that is currently used in Australia. In these circumstances, it may be reasonable to pursue a claim of noninferiority and adopt a truncated approach (rather than a full linked evidence approach) if adequate evidence can be provided to support that the proposed test is more accurate than the comparator test. This evidence would include a discussion of the true classification of cases that are discordant between the proposed test and the comparator, and a discussion of the downstream implications of the different test results. The goal of the assessment of a test with a small improvement in accuracy over the comparator will be to establish that the proposed test results in health outcomes that are no worse.

When the proposed test is likely more accurate than the comparator (and results in discordant results), the assessment should provide:

- sensitivity and specificity of the proposed test strategy, and of the current testing strategy, as derived by a comparison with an appropriate reference standard. If no reference standard is available, justify how the discordance is known to represent an improvement in accuracy.
- quantification of how the tests differ in terms of true/false positives and true/false negatives
- an explanation of why the tests differ in their detection of the parameter
- a discussion of the basis on which evidence-based management decisions were established.

Most importantly, the assessment should present the implications of downstream management decisions for the discordant cases; that is, discuss the impacts of the proposed test detecting fewer false positives or false negatives. If the proposed test identifies more positive patients than the comparator, clearly establish that there is no change in the spectrum of the disease (see Technical Guidance 13 for a discussion on the implications of a change in the spectrum of the disease).

An example of an assessment framework truncated at test accuracy is given in Figure 9.

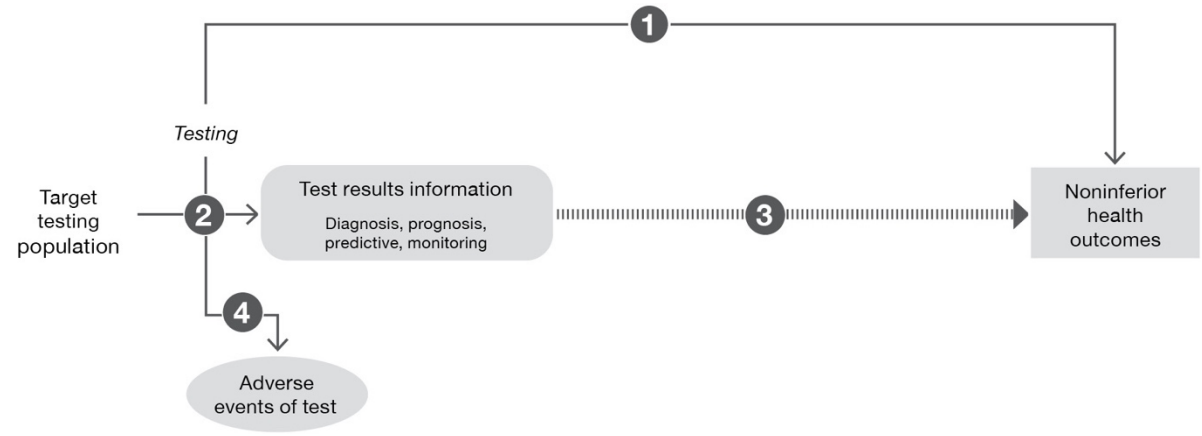
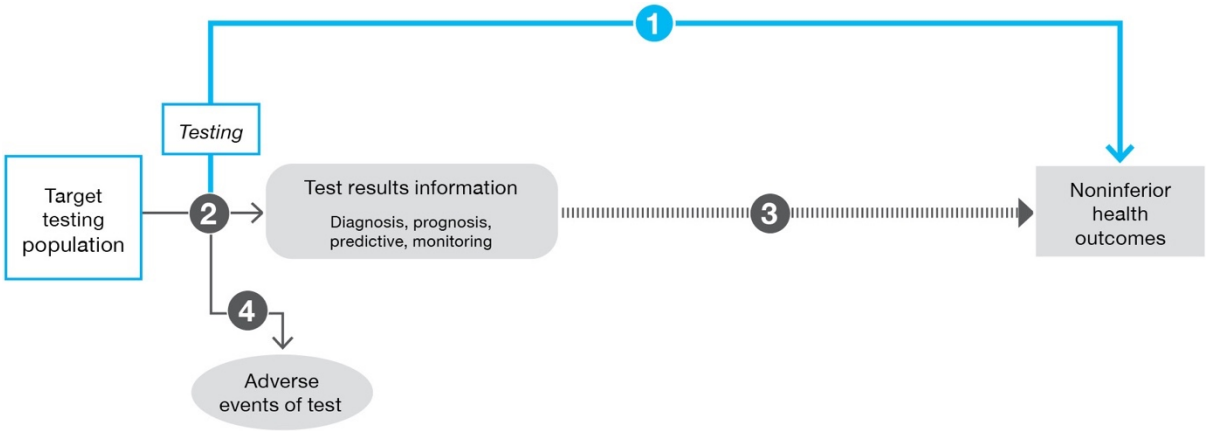


Figure 9 **Assessment framework that has been truncated at test accuracy (concordance, test accuracy) with the inference that identical test accuracy will result in the same health outcomes**

Assessment questions for a claim of noninferiority based on comparative accuracy of test results (Figure 9)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

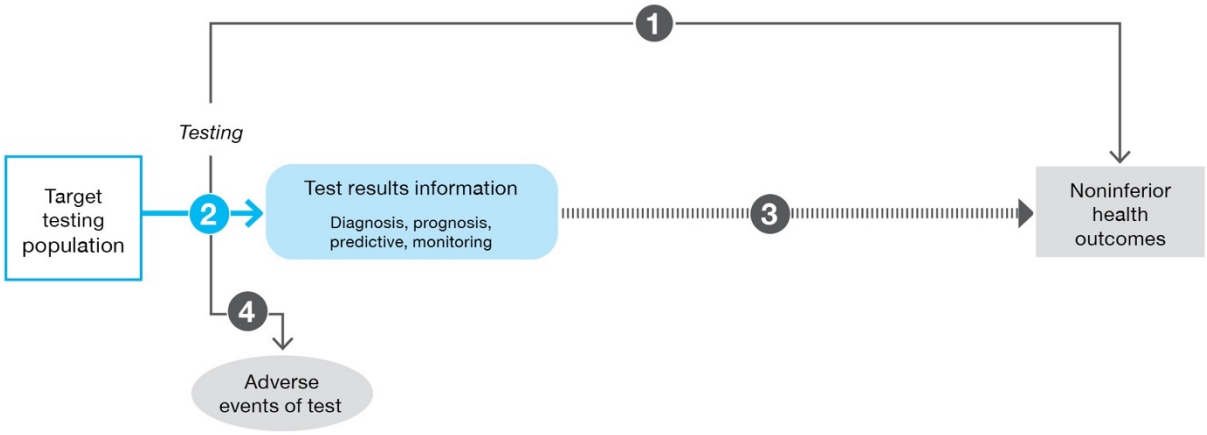
1 Does the use of [the proposed test] in [the target population] result in [key health outcomes, e.g. survival, quality of life] that are no worse than [the main comparator]? (If adequate direct from test to health outcomes evidence is available, go to Assessment question 4).



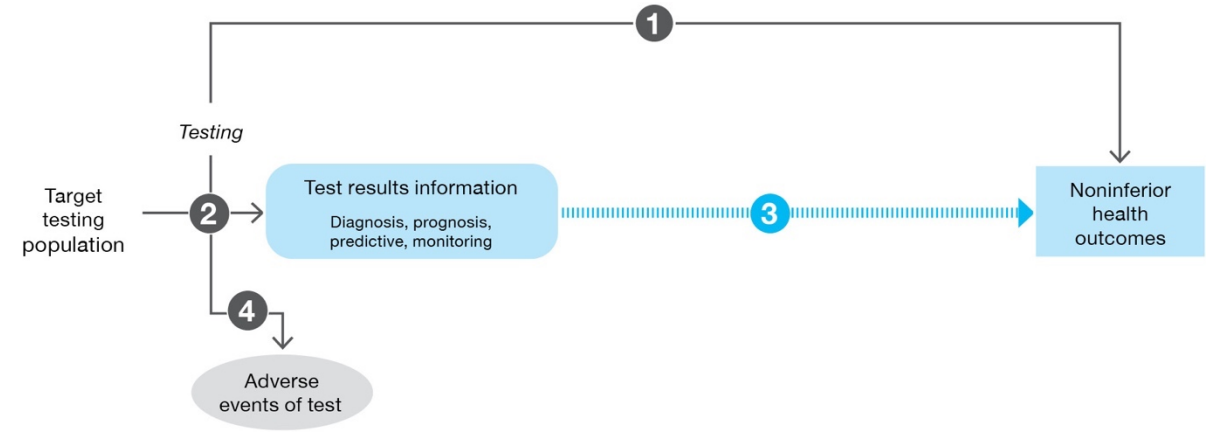
LINKED EVIDENCE

2 Is [the proposed test] in [the target population] concordant with [the main comparator]?

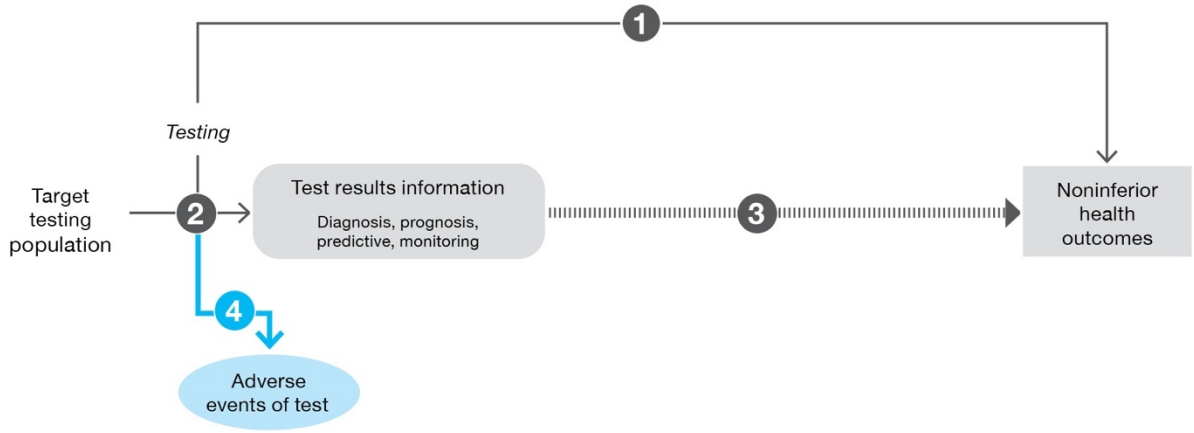
If concordance is unavailable, are the rates of false-positive and false-negative patients similar in accuracy studies? Is there additional evidence to support that estimates of similar test accuracy indicate that the proposed test and comparator test are categorising patients similarly?



3 *Inference that similar test results from both proposed test and comparator will result in the same management decisions, and noninferior health outcomes*.



4 What are the harms of [the proposed test] and of [the main comparator]?



Assessment question 2 may require judgement and justification. Presentation of a 2-by-2 table showing the proportions who are test positive and negative for each test is required. Where tests are almost 100% concordant, it is important to explain the cause of the nonconcordance; for example, are nonconcordant cases a consequence of identifying more or fewer patients, and is there evidence (compared with an appropriate reference standard) of their true test result? For tests where nonconcordance arises because the test is less accurate, a claim of noninferiority using a truncated approach cannot be pursued. For tests where nonconcordance arises because the test is more accurate, consider a full linked evidence approach to explore the impact of the threshold at which nonconcordance would invalidate the claim of noninferior health outcomes; using a truncated approach would be contingent on the causes of nonconcordance.

Other circumstances requiring/suitable for truncated frameworks

Assessment frameworks may also be truncated at different steps if there is evidence that the interpretation of the proposed test information is the same, or results in the same action as the comparator. This may be relevant if the proposed test reports a different parameter to current practice, but the clinical decision-making remains the same. It may also occur for some triaging tests if the triage test has 100% sensitivity, such that the final categorisation of patients is the same, but the triage test permits a proportion to avoid a more costly or invasive test.

See Appendix 1 for truncated frameworks for claims of noninferiority based on change in management, and use of triage testing.

TG 9.4 Adapting the framework for the value of knowing

A clinical claim relating to health outcomes must be made for all tests. For tests that rely on an additional claim relating to the value of knowing (i.e. benefits or harms related to things other than using the test information to change the management of patients), add an additional link relating to value of knowing in the framework (Figure 10). This can be done for a test that is claiming noninferior or superior health outcomes.

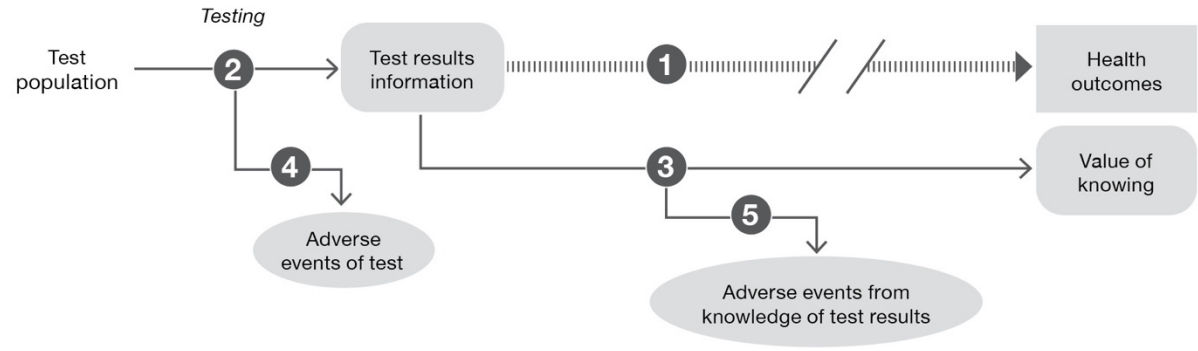
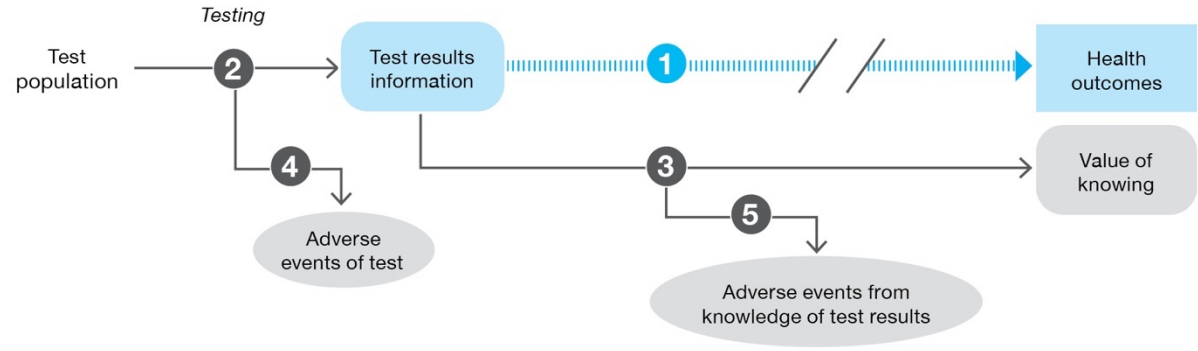


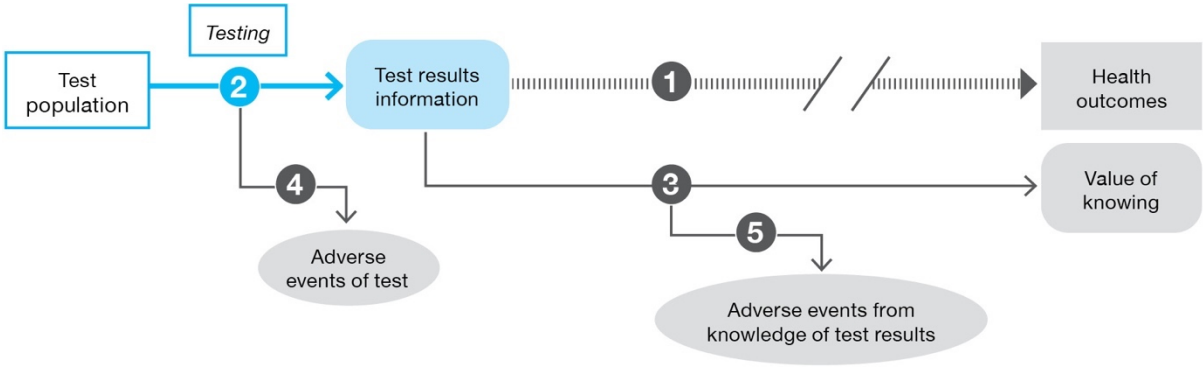
Figure 10 Assessment framework representing a claim of the value of knowing

Assessment questions for establishing a claim relating to the value of knowing (Figure 10)

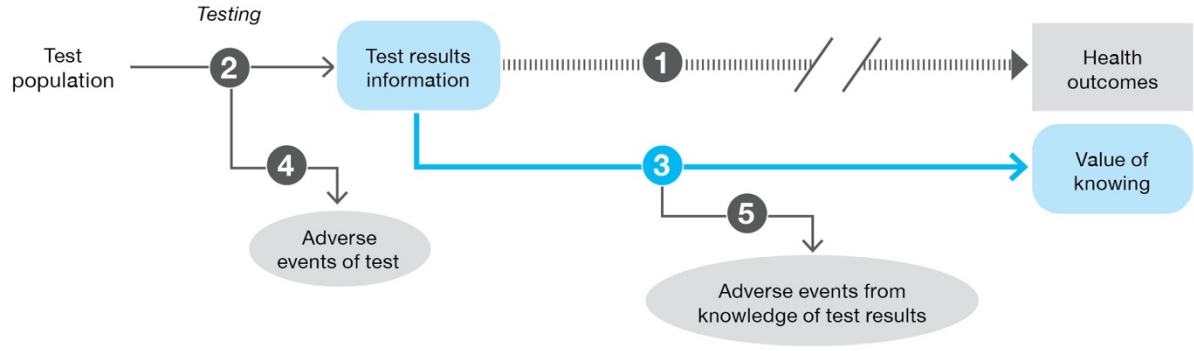
1 Does the use of the proposed test in place of the current test (comparator) result in superior health outcomes, or outcomes that are no worse than the comparator?



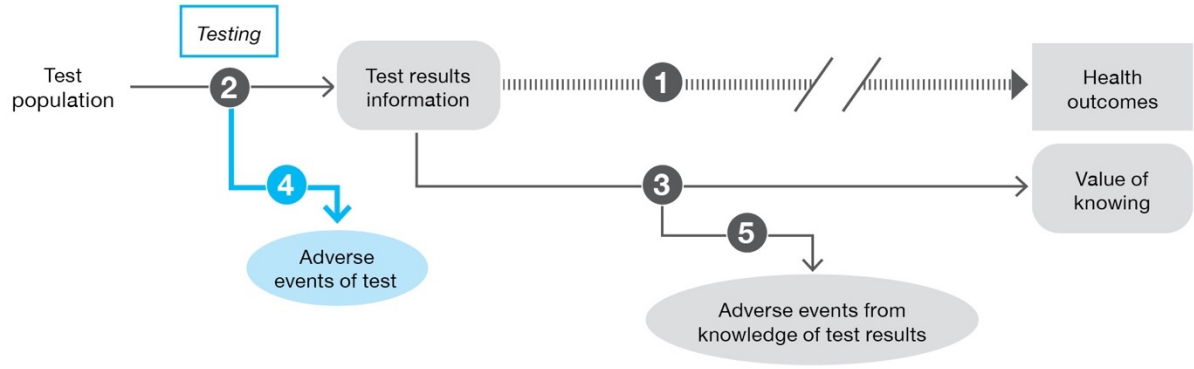
2 How does the information from the proposed test differ from that of the comparator?



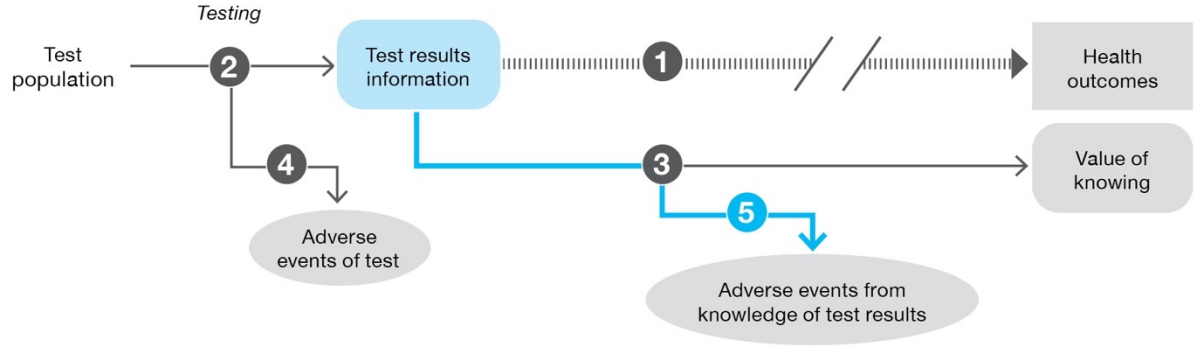
3 What beneficial impacts does the availability of information from the proposed test compared with the comparator have on nonhealth outcomes related to the value of knowing? What positive behavioural changes or actions (taken by the individual, families, carers or others) are associated with knowledge of the test results, compared with the information provided by the comparator?



4 What are the adverse events associated with the proposed test and the comparator?



5 What adverse effects/harmful impact does the availability of information from the proposed test compared with the comparator have on nonhealth outcomes related to the value of knowing? What negative behavioural changes or actions (taken by the individual, families, carers or others) are associated with knowledge of the test results, compared with the information provided by the comparator?



TG 9.5 Performing the assessment

Generate the assessment framework

Step 1: Prepare and report the PICO.

Step 2: Construct the assessment framework.

Construct the assessment framework based on the PICO. Where there are multiple comparators or multiple populations, the evidence required may still be captured by a single assessment framework (although it would inform multiple sets of assessment questions), or more than one assessment framework may be required.

If permitted by a clinical claim of noninferiority, truncate the preliminary framework, and present the final framework and the arguments necessary to support the truncation.

Step 3: Generate the assessment questions from the assessment framework.

Capture the assessment questions related to the assessment framework and record these in the PICO summary tables.

Address the assessment questions

Step 1: Address the direct from test to health outcomes evidence assessment questions (Technical Guidance 10).

For frameworks based on a claim of superior health outcomes, or that cannot be truncated to an earlier step, report on the available evidence for direct outcomes. Also report if no direct from test to health outcomes evidence is identified (or direct from test to health outcomes evidence is inadequate). Adequate direct from test to health outcomes evidence will not require the support of subsequent steps from the linked evidence approach.

Step 2: Address the test accuracy (Technical Guidance 11).

Compare the test accuracy against an appropriate reference standard. If there is a key trial providing direct from test to health outcomes evidence for a particular test (the clinical utility standard), assess how the range of tests used in Australia performs compared with the clinical utility standard.

Step 3: Address change in management (Technical Guidance 12).

For frameworks based on a claim of superior health outcomes, or that cannot be truncated to steps earlier than a change in management, report on the evidence for a change in management. Where the clinical claim relies on a change in management, and evidence for this step cannot be identified, the assessment will not be able to establish the magnitude of the benefit of the test in terms of health outcomes and so a claim of superiority is not appropriate. The assessment report should include advice from MSAC, and/or incorporate an assumption relating to change in management and highlight this as a major area of uncertainty. The method of deriving the change in management used in place of an evidence-based estimate must be described.

If there is evidence of no change in management, a claim of superiority in terms of health outcomes is unlikely to be appropriate (unless there are marked safety benefits). A claim of noninferiority may still be possible.

Step 4: Address the impact of change in management on outcomes (Technical Guidance 13).

Report on the expected health-related outcomes associated with the management decisions. If evidence for final health outcomes is not available, present evidence for validated surrogate outcomes, and translate the surrogate outcomes to final outcomes in a subsequent step.

Step 5: Discuss the consistency and transitivity of the evidence across the framework.

Direct from test to health outcomes evidence is preferred to linked evidence. Uncertainty increases with the number of steps between the decision to test and the final health outcomes.[20] As with any linked evidence approach, one source of uncertainty is the transitivity of the evidence across each step.

Step 6: Discuss the applicability of the evidence across the framework to the target population and setting of use of the technology.

Step 7: Discuss social, ethical, legal and organisational issues associated with implementation of the test.

Step 8: Summarise the results.

Provide an overarching summary of each of the steps. Discuss uncertainties in each of the steps, and the implications of these uncertainties on the final estimate of health outcomes.

TG 9.6 Location of guidance for assessment questions

The location of guidance in this document for each of the steps in the framework (Figure 11) is shown in Table 7.

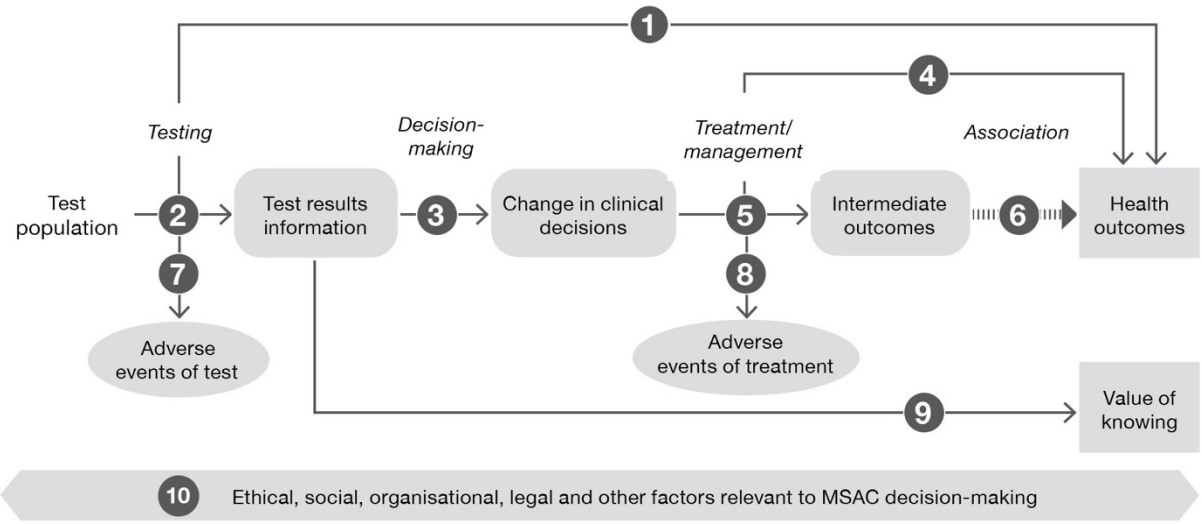
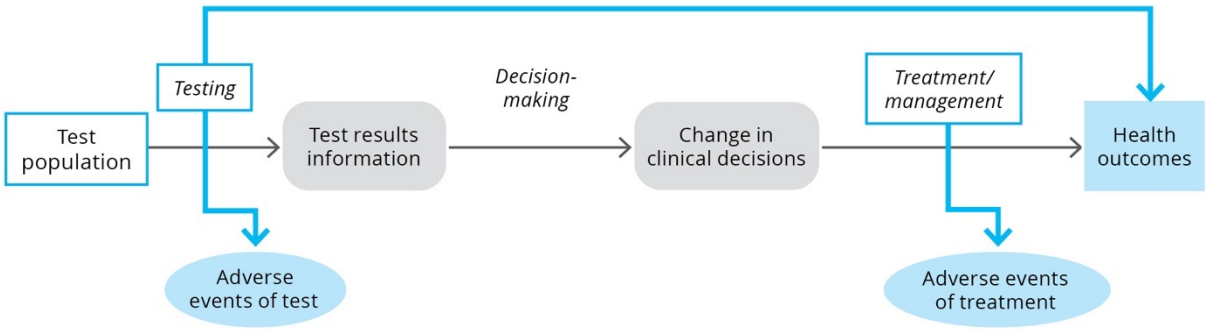


Figure 11 **Steps in the assessment framework (numbers refer to Table 7)**

Table 7 **Location of guidance in this document for each step in the assessment framework**

Step in the framework	Description	Location of guidance
1	Presentation of direct evidence of the impact of testing on health outcomes	Technical Guidance 10
2	Test accuracy, sensitivity/specificity, concordance	Technical Guidance 11
3	Evidence of change in management	Technical Guidance 12
4	Evidence of the effect of treatment/management on health outcomes	Technical Guidance 13
5	Evidence of the effect of treatment/management on surrogate outcomes	Technical Guidance 13 and Appendix 12
6	Evidence of the association between a change in the surrogate outcome and the target final outcome	Appendix 12
7	Evidence for the safety of the test	Technical Guidance 14
8	Evidence for the safety of downstream management decisions	Technical Guidance 14
9	Evidence for the value of knowing test information and its impact on nonhealth outcomes	Technical Guidance 28
10	Evidence for social, ethical, legal and organisational impacts associated with implementation of the test	Technical Guidance 29

Technical Guidance 10 Direct from test to health outcomes evidence



KEY CONSIDERATIONS

For investigative technologies

- **Direct from test to health outcomes evidence may not provide results that are transferable or generalisable to the target population; studies are often underpowered to detect a difference in health outcomes and additional components of linked evidence may be required (TG 10.4).**
- **Present the direct from test to health outcomes evidence in the same way as for a therapeutic health technology. Describe any applicability concerns of the direct from test to health outcomes evidence. Explain how additional evidence, if required, has been used to address issues with applicability (TG 10.6).**

TG 10.1 Purpose of the guidance

The assessment of investigative technologies involves identifying the impact of a test on health-related outcomes and, in some cases, on the broader value of knowing. The evidence required to establish an impact on health outcomes can involve either a direct from test to health outcomes evidence approach, or a linked evidence approach. This TG section relates to a direct from test to health outcomes evidence approach and discusses:

- a definition of direct from test to health outcomes evidence
- study designs that can be used to inform a direct from test to health outcomes evidence approach
- considerations relating to this type of approach, including limitations and gaps that may arise from particular study types
- how to assess the applicability of direct from test to health outcomes evidence to clinical practice
- a suggested approach to presenting direct from test to health outcomes evidence for an investigative health technology.

TG 10.2 Direct from test to health outcomes evidence

Direct from test to health outcomes evidence refers to evidence from trials or studies specifically designed to measure the effect of a test on a health outcome. For simplicity, direct from test to health outcomes evidence may be called clinical utility evidence. Clinical utility evidence is characterised by the measurement of key patient-relevant health outcomes in a study in which patients receive a test that informs treatment decisions. A study that provides information on the categorisation of patients and their subsequent health outcomes is considered to provide 'direct

from test to health outcomes evidence' only if 1) the treatment (and therefore health outcomes) was influenced by the results of the test, or 2) the study provides sufficient test result and treatment arms to make it clear whether following results of the test affected health outcomes.

This is distinct from a linked evidence approach in which the effect of test results on change in management, and the effect of change in management on health outcomes are not captured in a single study (Figure 12).

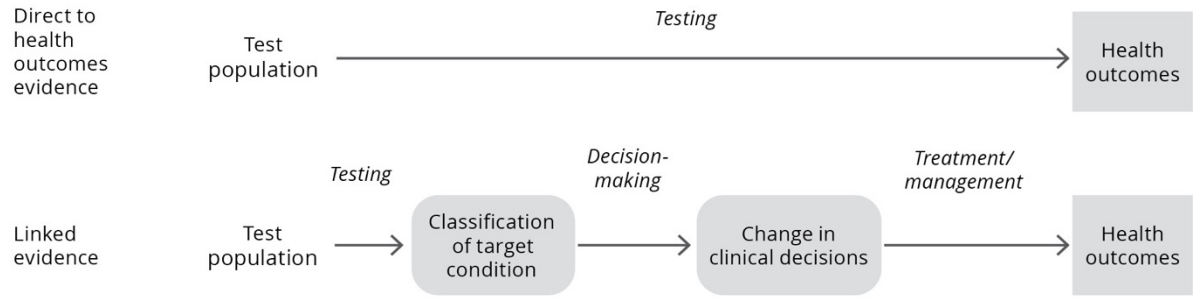


Figure 12 **Single study of test population in direct from test to health outcomes evidence compared with multiple studies informing each step of a linked evidence approach**

TG 10.3 Direct from test to health outcomes study designs

Direct from test to health outcomes evidence can encompass a variety of study designs, both comparative and noncomparative. However, as MSAC is informed by a comparison of the proposed test with current practice (to establish the incremental benefits and harms), noncomparative studies of the proposed test will require additional evidence of the main comparator so the relevant comparison can be performed, as well as supportive evidence to establish the transitivity of the studies.

Several study designs may provide clinical utility evidence; however, comparative studies that randomised patients to the proposed test versus comparator test are preferable (Figure 13).

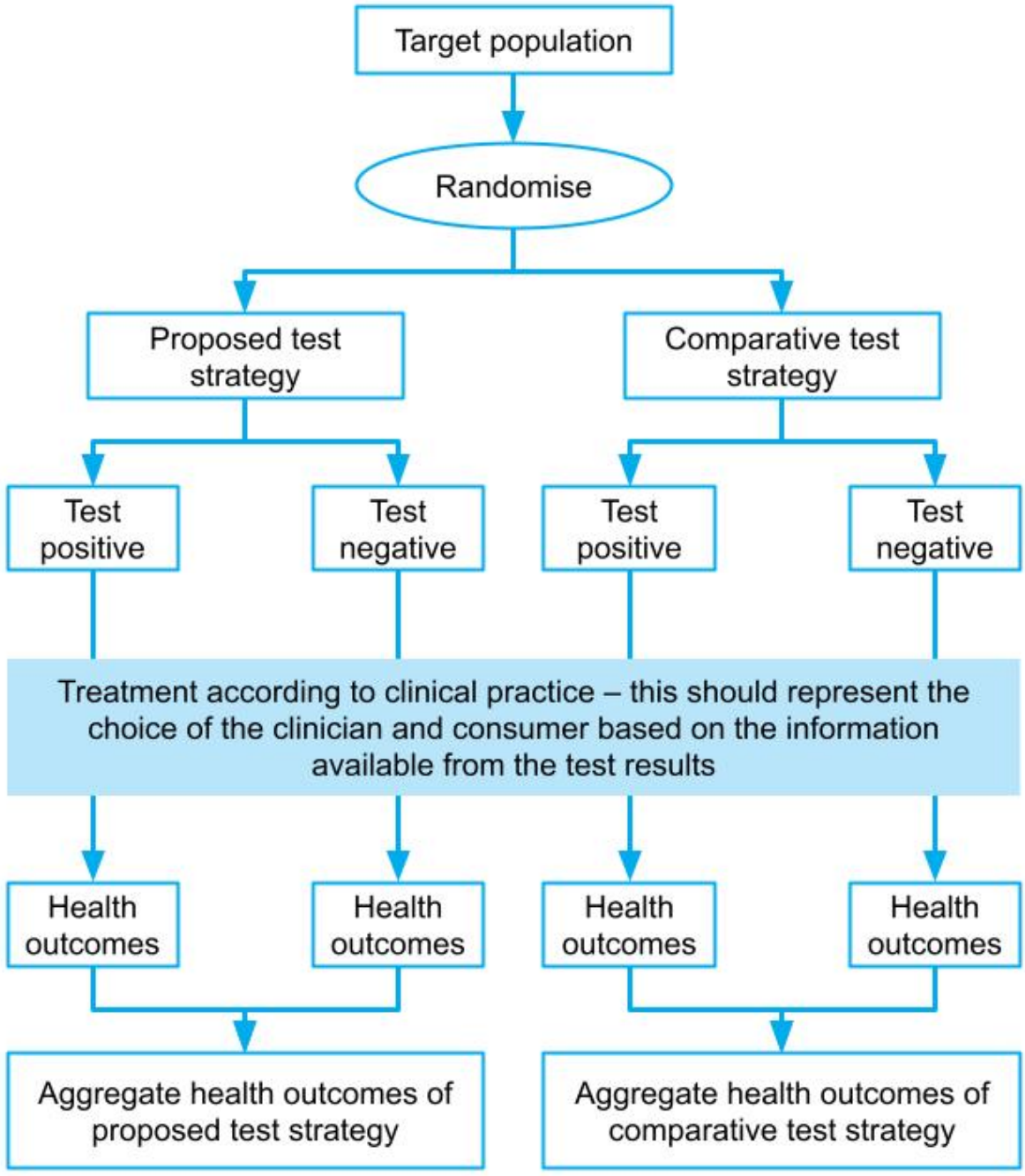


Figure 13 **Single-randomised direct from test to health outcomes trial evidence comparing health outcomes from new test strategy and existing test strategy, where test information is used to derive treatment condition (preferred option for assessment of any claim regarding the clinical utility of a test)**

A study that randomises patients/participants to receiving the proposed test versus the comparator test (or standard practice) will provide evidence of the overall comparative health benefit of the test. Many study designs will provide part of the information derived by this design. Studies that provide incomplete estimates of the comparative impact of a test on health outcomes should be used with caution (see TG 10.4).

TG 10.4 Considerations relevant to a direct from test to health outcomes evidence approach

Direct trials of tests may not provide results that are transferable or generalisable to the target population, and are often underpowered to detect a difference in health outcomes.[21] In these circumstances, it may be beneficial to assess additional components of linked evidence.

Observational studies

Observational (nonrandomised) comparisons of patients who receive a test versus those who do not may represent patient self-selection based on adherence or other factors, which may introduce confounding. For example, those who adhere to testing may differ systematically to those who refuse or do not seek the test.[22] This confounding may have a large impact on the subsequent health outcomes, and study results should therefore be treated with caution.

Incomplete direct from test to health outcomes evidence

Ideally, the health outcomes following the use of the proposed test are compared with the health outcomes following the use of the comparative test strategy (standard practice) in the same study. Some studies will provide only health outcomes associated with the use of the proposed test and not the comparative test strategy (Figure 14).

If more than one study is required to compare the outcomes associated with the proposed test with those of the comparative test strategy, the transitivity of these studies must be adequately described. Indirect comparisons of test to health outcome studies have a high risk of bias. Unlike indirect comparisons of treatments, where often the largest uncertainty relates to the baseline characteristics of the populations, an indirect comparison of tests may have transitivity issues related to the populations as well as prevalence of a biomarker, the thresholds of the tests, the clinical practice decisions and the treatments. As such, indirect comparisons of test to health outcome studies should be avoided, where possible. If they are required, the transitivity of each of the components of the studies must be rigorously assessed (see Appendix 6 for sources of heterogeneity).

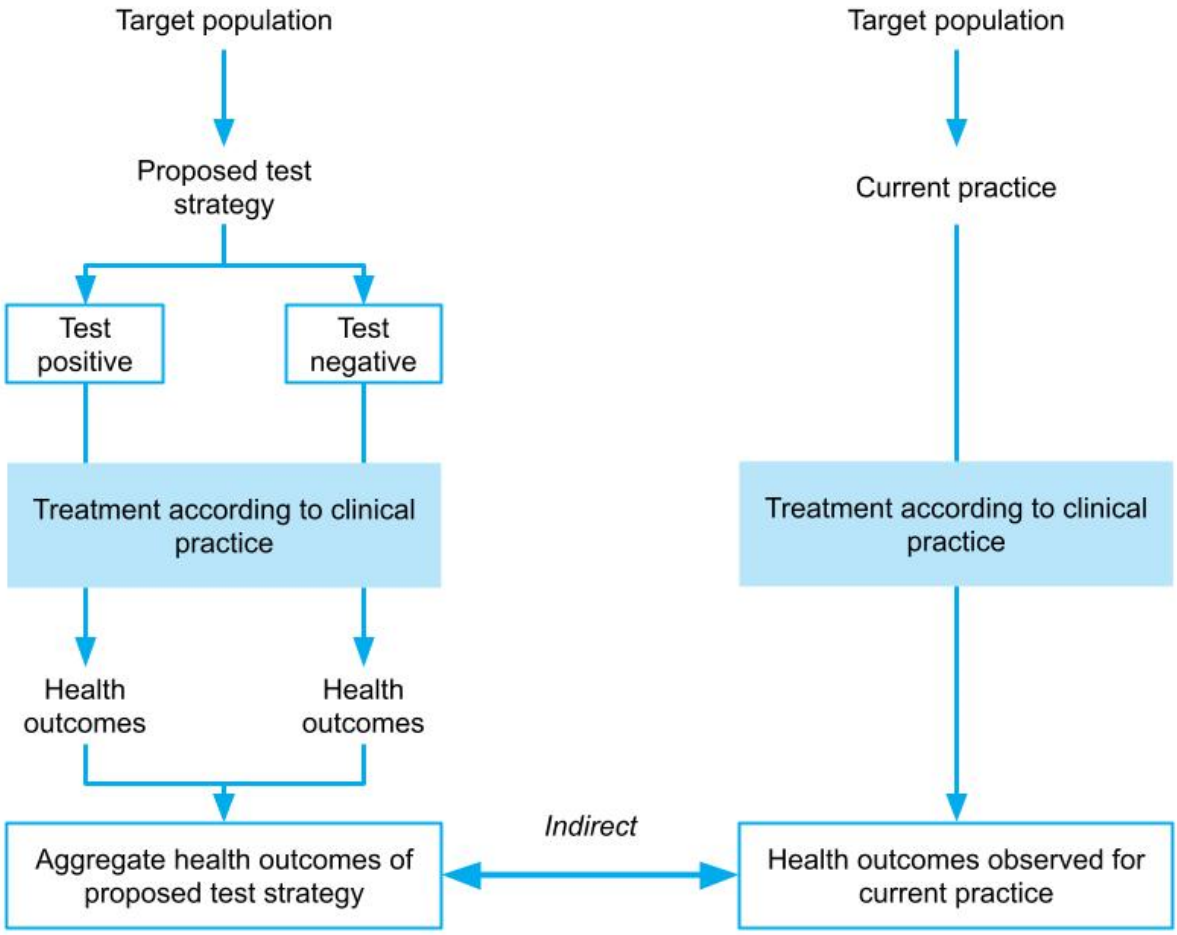


Figure 14 **Single-arm study of the test reporting on health outcomes, with an indirect comparison of current health outcomes in the absence of the test (or using an alternative test)**

TG 10.5 Assessment of the applicability of direct from test to health outcomes evidence

Direct from test to health outcomes evidence is characterised by a study that measures the impact on health outcomes following the use of a test. While this type of evidence provides more robust internal validity than a linked evidence approach, the applicability of the evidence must be assessed to ensure that the results from the study will be replicated in clinical practice.

Concerns relating to applicability arise when the population or interventions (and/or circumstances of use) in a study differ from the target setting. For therapeutic technologies, an assessment of applicability primarily consists of comparing the populations and the interventions across the study and target setting, and identifying differences that may have an impact on the expected outcomes.

The applicability of direct from test to health outcomes evidence depends on the applicability of not only the population and the test, but also the actions or subsequent steps taken in the study. Where there are concerns with the applicability of any component of the direct from test to health outcomes evidence, additional supportive evidence may be required.

An assessment of applicability includes a comparison of the study and the Australian setting (Figure 15) for the characteristics shown in Table 8.

Table 8 Applicability assessment of direct from test to health outcomes evidence

Applicability domain	Sources of applicability concerns	Additional supportive evidence that may be required
Applicability of the study population	Baseline patient characteristics *Prevalence of the biomarker/disease*[a]	Comparison of the prevalence of the biomarker in the trial participants with that of the target population
Applicability of the study test	Is the test used in the study (the clinical utility standard) the same as the proposed test?	Comparison of patient classification through use of the clinical utility standard with that of the tests available in Australia
Applicability of the clinical decision-making	Will clinical practice in Australia reflect the clinical practice in the study? Was the choice of treatments at clinician discretion in the study, or was it protocol driven?	Additional evidence that the test will impact clinical decision-making. Evidence that the treatment options are the same in Australia as in the study.
Applicability of the treatment/management	Was the treatment delivered in the study as it would be delivered in Australia (in terms of clinical setting, dose, duration, concomitant therapies etc.)?	Evidence that treatments are delivered similarly in the Australian setting as in the study

a The prevalence of the biomarker is fundamental to assessing the applicability of direct from test to health outcomes evidence. Prevalence of the biomarker in the study may be influenced by methods for identifying the test population or may vary by race (particularly if the biomarker is a genetic marker). Eligibility criteria (such as requiring particular symptoms or including only high-risk populations) may be used to enrich the biomarker positive population. If the biomarker prevalence in the study differs from the prevalence in the Australian target population, the aggregate health outcomes from the study will not be valid.

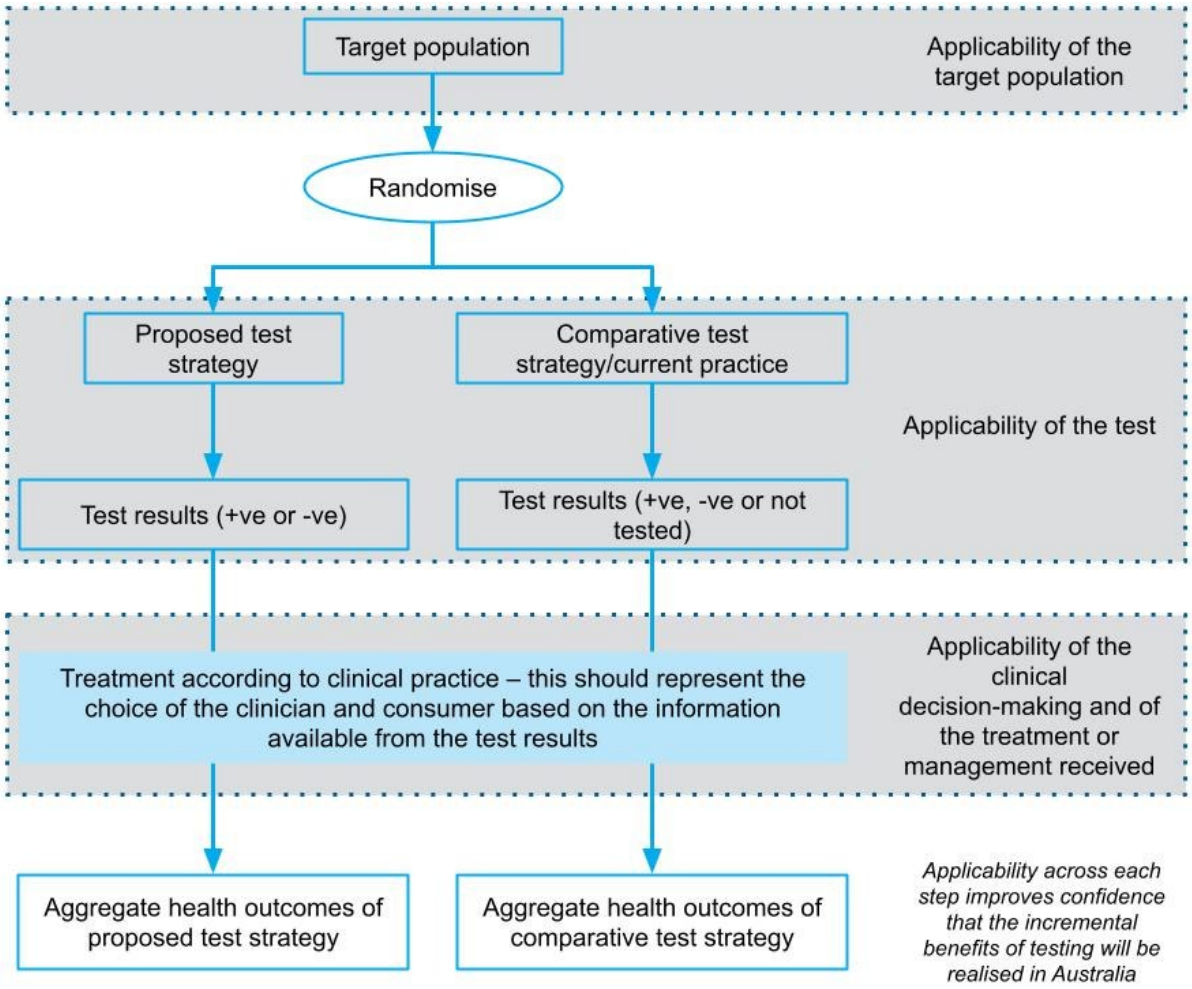


Figure 15 Assessment of applicability of direct from test to health outcomes evidence

Applicability of the study population

Where the proportion of patients with the biomarker (prevalence) in the study population is different to that in the target population, the aggregate health outcomes associated with the test may also differ. Additional evidence comparing the target (testing) population and the study population is required. The population in the study may differ from the target population in several ways; however, a key concern is if the study population has been enriched. Enrichment can occur by using eligibility criteria that narrow the study population to patients with a higher risk of having the biomarker; this would result in a difference in the proportions of patients who receive a particular test result (i.e. prevalence of the biomarker) in the trial compared with the target population. If prevalence differs, the aggregated impact on health outcomes may not be applicable and may require adjustment.

Applicability of the study test (clinical utility standard)

The health outcomes observed in a direct from test to health outcomes study are applicable to the characteristics of the test used in that study (the clinical utility standard).

If test characteristics of the proposed test differ from the clinical utility standard, or there are multiple tests that may be available in clinical practice (that may be eligible for the same funding arrangement), health outcomes may differ from those observed in the study. This direct from test to

health outcomes evidence must be augmented with evidence that compares the test characteristics (sensitivity, specificity and/or concordance) of the tests that will be available in clinical practice with the clinical utility standard (see Technical Guidance 11).

Applicability of the clinical decision-making

For the health outcomes associated with the proposed test or comparative test strategy to be valid, the change in management observed in the study must mimic the change in clinical practice following the availability of the proposed test.

Where the management decisions associated with test results (such as treatment with a particular medicine or a surgical procedure) are written into the protocol of the study, the study cannot provide evidence of change in management. Evidence with protocol-enforced management must be augmented with evidence for change in management (see Technical Guidance 12).

Applicability of the treatment/management

The final step for assessing applicability relates to the treatment or management that is provided in the study. It is important to compare the treatments in terms of dose, duration and frequency, and concomitant treatments. It is also important to compare the setting and other circumstances of use.

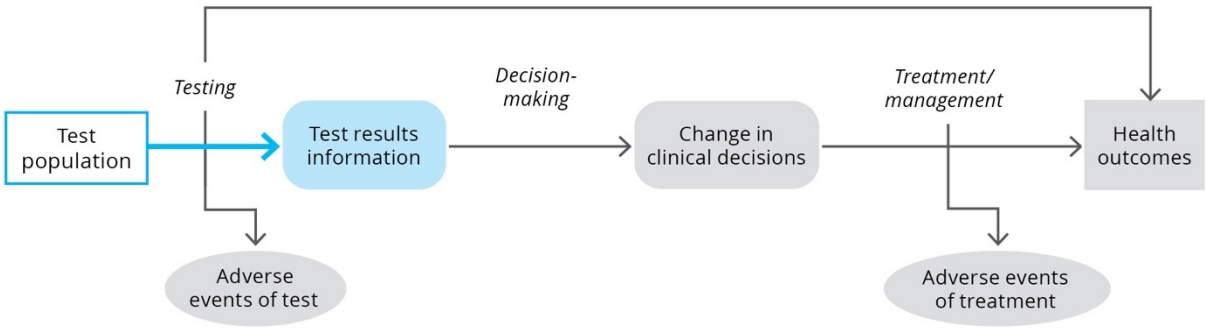
TG 10.6 Presentation of direct from test to health outcomes evidence

The principles for presenting direct clinical trial evidence of the effect of an investigative health technology on patient health outcomes are similar to those for presenting clinical trial evidence for a therapeutic technology. That is, MSAC's preference for study designs is as per the hierarchy in the NHMRC levels of evidence for interventions, and risk of bias checklists relevant to interventions should be used (rather than checklists designed for studies assessing accuracy).[10]

However, the following additional components are required to present this information clearly and comprehensively:

1. Describe how the direct from test to health outcomes evidence has been constructed. If multiple sources of information have been used, describe why they are necessary.
2. Present the direct from test to health outcomes evidence in the same way as presenting a therapeutic health technology (i.e. describe the literature search, risk of bias and trial characteristics, present the results, and meta-analyse, if appropriate).
3. Describe applicability concerns of the direct from test to health outcomes evidence. Explain how additional evidence has been used to address issues with the applicability.
4. Present evidence of the harms (adverse events) experienced by patients who receive the proposed test versus the comparative test (standard practice). These harms should include direct test-related harms and harms that are associated with subsequent management decisions (see Technical Guidance 14).

Technical Guidance 11 Linked evidence – test accuracy



KEY CONSIDERATIONS

For investigative technologies with cross-sectional evidence

- **Present the sensitivity and specificity compared to the reference standard and/or the clinical utility standard, for individual studies, in a table or forest plot (TG 11.4).**
- **If appropriate, present a meta-analysis and/or subgroup analysis of the sensitivity and specificity values from appropriate studies, as a forest plot and/or HSROC curve (TG 11.4).**
- **Determine the positive and negative predictive values of the test using the prevalence rate (determined as indicated in TG 11.9), and the pooled estimate of sensitivity and specificity (TG 11.4).**
- **In the absence of a reference standard, the concordance between the test and a clinical utility standard should be presented as overall, positive and negative percent agreement and/or as a Cohen's kappa coefficient measure of agreement (TG 11.6).**
- **Report any issues regarding the reliability of the test with respect to inter-rater and inter-laboratory variability/agreement, the proportion of failed and equivocal test results, and/or any other reliability measures that may apply to specific tests (TG 11.8).**

For investigative technologies with longitudinal evidence

- **When presenting evidence for a test to establish a predisposition for a disease or the prognosis of the clinical course of disease, the reference standard, or clinical endpoint of interest, must be clearly defined (TG 11.5).**
- **When data are presented in a time-to-event format, Kaplan–Meier curves and hazard ratios should be presented (TG 11.5).**

TG 11.1 Purpose of the guidance

In the absence of high quality direct from test to health outcomes evidence, an assessment will take a linked evidence approach. One key uncertainty with this approach is whether patients are appropriately categorised by the test (i.e. test accuracy). This information is needed so that the flow-on effects of test categorisation on subsequent evidence linkages can be determined (i.e. how the proposed test would change patient management and its likely impact on patient health outcomes). The impact of the categorisation on clinical management and health outcomes is addressed in Technical Guidance 12 and Technical Guidance 13, respectively.

This TG section will discuss the methods required to determine the proportion of patients who will be appropriately classified by the test and the proportion who will not. It will provide guidance on:

- estimating the sensitivity and specificity of the proposed test and the comparator

- comparing the results of different tests available in Australia
- estimating the prevalence of the disease and/or biomarker in the target population.

A crucial concept for the assessment of test accuracy is how the proposed test performs against other tests. The other tests of interest are:

- the reference standard
- the clinical utility standard
- the comparator (this may be a current test, or a clinical diagnosis/prediction).

The assessment report should present how the proposed test categorises patients into test positive and test negative, compared with another test. This is most simply done using a 2-by-2 table, where the positives and negatives for each test are reported and compared; the number and nature of the discordant test results can then be identified and discussed. When compared against a reference standard, the test results should be categorised as true positive, true negative, false positive and false negative.

As well as reporting on the nature of the discordant cases, it is important to explain the discordance. A key uncertainty in the linked evidence approach arises when the proposed test categorises patients differently to the comparator (e.g. results in more or fewer diagnoses), as it is unclear whether these differences represent a change in the spectrum of the disease being identified, or some other systematic reason for differences between the tests. Where there is a change in the *types* of patients categorised as test positive and test negative, there is less certainty that the subsequent steps in a linked evidence approach (change in management and health outcomes) would be applicable if these steps were informed by a standard that categorises patients differently.

TG 11.2 Key concepts

Terminology used in the current guidelines

The current guidelines use the term 'test accuracy' to encompass multiple types of test performance. The terms 'analytical validity' and 'clinical validity' were used in previous guidelines. Analytical validity was intended to represent the sensitivity and specificity of the proposed test measured against a nonclinical reference standard, as well as its reliability and reproducibility. The results would report how well the proposed test identified the presence of a biomarker or analyte compared with the reference standard test. Clinical validity was intended to represent how well the proposed test could identify the target condition, and would include estimates of test sensitivity and test specificity.

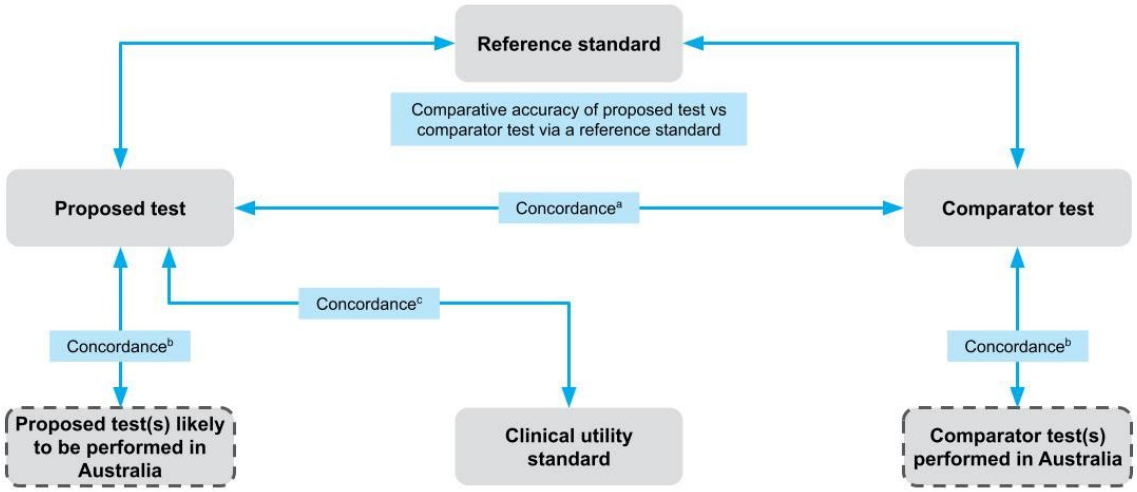
The method of assessment is not driven by the purpose (analytical or clinical validity) but rather by the type of evidence (cross-sectional versus longitudinal).

The term test accuracy has been used to encompass both concepts, and also to reduce confusion that may arise with the use of analytical validity, which may have different meanings across settings.

Comparisons between proposed tests, comparator tests and reference standards

Figure 16 describes the comparisons that may be relevant in an assessment of test accuracy, depending on whether or not a reference and/or clinical utility standard have been identified. The various reference standards that may be available are discussed below.

It is important to provide an explanation for why each of the comparisons presented is necessary for interpreting subsequent steps in the linked evidence approach. Providing comparisons against all possible tests without justifying the need for the comparison, or interpreting the result, is unhelpful.



a Primarily used if the proposed test is claimed to be noninferior to the comparator test. When reporting concordance, the nature of discordance (ie those that are positive with test A but negative with test B, and vice versa) is described and discussed.

b Where there is variation in the proposed test or comparator test (in terms of test characteristics, thresholds, interpretation) across Australia, concordance of the test providing evidence and the Australian tests is required. If the proposed test is the clinical utility standard, see note c.

c If direct from test to health outcomes evidence is available, and the proposed test is not the clinical utility standard, the most important comparison is of the proposed test vs the clinical utility standard. The comparison reports the nature of discordance and the implications.

Figure 16 Comparisons of interest to determine the accuracy of the proposed test compared with other available tests

The assessment of the accuracy of the proposed test may include one or more of the following comparisons:

- If a reference standard is available, the accuracy of the proposed test(s), the clinical utility standard and any comparator tests that are used in Australia should be measured against the reference standard.
- Accuracy measures could include sensitivity and specificity, positive and negative predictive values, etc. The health outcomes for false-positive and false-negative patients should be discussed.

If a clinical utility standard is available, the accuracy of the proposed test(s) should be compared to the clinical utility standard. The results of a comparison with the clinical utility standard may, in some cases, be more relevant than a comparison against a reference standard. Equivalence of the proposed test and the clinical utility standard provide confidence that the proposed test will result in the same health outcomes as observed in the study that used the clinical utility standard. Present the concordance (positive and negative percent agreement) between the proposed test and the clinical utility standard, and discuss the discordant results (i.e. results being considered as false positive or false negative compared to the clinical utility standard).

In the absence of both a reference and a clinical utility standard, the test concordance between the proposed and comparator tests is required. In this scenario, patients with discordant outcomes cannot be identified as having either true or false test results with respect to either test. However, the nature of the discordant results is important to discuss.

In many cases, the comparator test will be a currently used test, or it may be no testing. There will also be circumstances where the comparator test could be the reference standard, an imperfect reference standard or the clinical utility standard.

Reference standard

The term 'reference' or 'gold' standard refers to a benchmark that is the best available test for determining the presence or absence of the condition of interest. It is not likely to be a perfect test, but merely the best test currently in use. The identification of the reference standard is discussed in TG 2.4.

The availability of an appropriate reference standard creates more certainty around the evidence presented. This allows quantitative assessment of sensitivity and specificity, and informs whether the proposed test is superior or noninferior to the main comparator in terms of accuracy and reliability.

The reference standard may be either nonclinical (comparing the ability to detect a biomarker) or clinical (comparing the ability to detect a disease or symptom of disease). If a clinical reference standard is available, the accuracy of the proposed test against this clinical reference standard would be preferred over the nonclinical reference standard in most cases.

A clinical reference standard may often be a composite standard, involving multiple tests and clinical assessments to diagnose the disease, and may, at times, include the results from a comparator test or even the proposed test itself. Note that if the reference standard incorporates information from the proposed test, the results will be subject to incorporation bias.[23]

In longitudinal accuracy studies, the reference standard is more likely to be clinical outcomes at a later time point; this is relevant to prognostic tests (health outcomes) or predictive tests (response to treatment).

Clinical utility standard

A clinical utility standard is the test (and methods of testing) that was used to generate direct clinical outcomes in patients with and without a biomarker. If the clinical outcome is response to a targeted treatment, the clinical utility standard may be registered with the Therapeutic Goods Administration as a 'companion diagnostic'.

Any inaccuracies in the classification of patients by the clinical utility standard are accounted for in the clinical outcomes of the key study involving the clinical utility standard. A comparison of the concordance between the proposed test and the clinical utility standard will identify whether more or fewer patients would receive targeted treatment if the proposed test is used. The implications of the discordant test results are important to discuss. It may be appropriate to assume that discordant results derived by the proposed test are false positives and false negatives unless there is robust evidence to support an alternative conclusion (such as the availability of comparative evidence against a reference standard).

Imperfect reference standard

Reference standard tests may be imperfect, and incorrectly identify a proportion of the population as test positive or test negative. Often the imperfect reference standard will be well established in diagnostic laboratories for routine diagnosis of the biomarker or condition.[24] This may be due to a more accurate test being unavailable, costly or unnecessarily invasive.

When comparing the accuracy of the proposed test to an imperfect reference standard, care should be taken when interpreting the false-positive and false-negative rate. If the proposed test is more accurate, these 'false' test results may actually be true positives and/or negatives that are misclassified by the imperfect reference standard. In cases where the imperfect reference standard is clearly inferior in terms of accuracy (both sensitivity and specificity) to the proposed test, and it is not used to direct treatment, a comparison against an imperfect reference standard is of limited value, and an alternative (such as a clinical reference standard) may be more relevant.

Trikalinos and Balion[25] indicate that test accuracy measured against an imperfect reference standard can be assessed in four different ways:

- Assess the proposed test compared with a clinical reference standard instead of the imperfect reference standard.
- Assess the concordance or agreement between the two tests.
- Calculate 'naive' estimates of the index test's sensitivity and specificity compared with the imperfect reference standard, but qualify study findings and discuss in which direction they are biased.

Adjust the 'naive' estimates of sensitivity and specificity of the index test to account for the imperfect reference standard; the 'adjusted' approach generally requires patient-level data to be available.

Even if the chosen reference standard is imperfect, if it has errors independent of both tests, it may still be able to establish which of the intervention and comparator is superior.[24] For example, level of exposure can be used as a 'fair umpire' test for an infectious disease. Ewer et al. reported on a comparison of two types of test for tuberculosis: the tuberculin skin test (TST) and the enzyme-linked immunospot (ELI) assay. These were examined in a school-based cohort where there was a tuberculosis outbreak, with differing levels of exposure (same class as index case; share some classes with index case; same year level but did not regularly share class with index case; different year levels). The better test was the one with the stronger correlation to level of exposure. The TST reported 90% positive in the closest exposure group and 20% in the least exposure group, whereas the ELI assay showed a stronger gradient, with 100% in the closest exposure group and 17% in the least exposure group. Therefore, although level of exposure alone would be insufficient to suggest who had latent tuberculosis, it was able to assist with distinguishing which testing strategy was superior.[24]

When a comparison against an imperfect reference standard is required, the approach taken should be justified. In many cases, additional supplementary evidence may be available to support the conclusion of improved sensitivity and/or specificity of the proposed test compared with the imperfect reference standard.

Partial verification and differential verification

In some instances, test accuracy studies may use the reference standard only as a confirmatory test. In these studies, only those samples/patients with a positive test result are tested with the reference

standard. These studies should only be included if no studies comparing both tests for all samples/patients are available.

It should be noted that if not everyone receives the reference standard (e.g. if only positive screening tests have further testing), the results will be subject to partial verification bias.[23]

In some cases, a reference standard may not be able to be applied. For example, in screening mammography, observed lesions may be biopsied to determine the presence of cancer with histopathology (reference standard). Mammograms without an observed lesion cannot be biopsied. To determine whether a negative mammogram was accurate, the individual will have to be followed up to see if cancer is detected later (differential verification).

Use of a clinical versus a nonclinical reference standard when evaluating test accuracy

In some circumstances (such as biochemical, cytogenetic and molecular genetic testing), it is important to distinguish between how accurate the test is in detecting a biomarker and how accurate it is in detecting the clinical disorder or outcome of interest.

If good quality and applicable estimates of accuracy against a valid clinical reference standard are available, test accuracy against a nonclinical reference standard may not needed.

Examples of clinical and nonclinical reference standards and a clinical utility standard for some example tests are shown in Table 9.

Table 9 Clinical and nonclinical reference standards and/or clinical utility standards for some example tests

Test	Purpose of test	Nonclinical reference standard	Clinical reference standard	Clinical utility standard
NAAT for tuberculosis	Diagnosis of TB	Ability of the test to accurately detect *Mycobacterium tuberculosis* (RS: microbial culture of suitable specimens) Accuracy measures: sensitivity, specificity	Ability of the test to detect a case of tuberculosis (RS: composite reference standard including clinical findings, microscopy, histology, cytology, chest radiographic findings, site-specific CT scan/MRI results, culture results and response to anti-TB drug therapy) Accuracy measures: sensitivity, specificity, PPV, NPV, likelihood ratios	N/A

Test	Purpose of test	Nonclinical reference standard	Clinical reference standard	Clinical utility standard
BRCA1/2 variant test	Determine presence of a *BRCA1/2* class 4 or 5 variant as a surrogate measure of likely response to a PARP inhibitor	Ability of the test to accurately detect *BRCA1/2* class 4 or 5 variants (RS: Sanger sequencing ± MLPA to detect *BRCA1/2* variants) Accuracy measures: sensitivity, specificity, likelihood ratios, PPV, NPV, post-test probability of having the biomarker	N/A due to heterogeneity of the tumour genomes (pathogenic variants in other genes may influence response to PARP inhibitors)	Ability of the test to predict response to treatment in biomarker-positive patients (CUS: *BRCA1/2* variant test used in RCT with direct health outcomes)
CFTR variant carrier testing	Determine *CFTR* carrier status of family members of someone with cystic fibrosis	Ability of the test to accurately detect the familial *CFTR* variant (RS: Sanger sequencing) Accuracy measures: diagnostic yield	N/A as no clinical RS for family members	N/A
Digital breast tomosynthesis	Diagnose breast cancer	N/A	Ability to detect architectural distortions, focal asymmetries and micro-calcifications in benign versus malignant cancers (RS: histological examination of biopsy samples) Accuracy measures: sensitivity, specificity, PPV, NPV, likelihood ratios	N/A

BRCA1/2 = breast cancer gene 1 and 2; *CFTR* = cystic fibrosis transmembrane conductance regulator gene; CT = computed tomography; CUS = clinical utility standard; MLPA = multiplex ligation-dependent probe amplification; MRI = magnetic resonance imaging; NAAT = nucleic acid amplification testing; N/A = not applicable; NPV = negative predictive value; PARP = poly ADP ribose polymerase; PPV = positive predictive value; RCT = randomised controlled trial; RS = reference standard; TB = tuberculosis

In the absence of a clinical reference standard, the clinical accuracy of a test depends on both the ability of the test to detect the biomarker compared to the nonclinical reference standard as well as the strength of the biological plausibility linking the surrogate measure (test results) with the clinical condition of interest. For example, the link between HbA1c (glycated haemoglobin A1c) levels and fasting blood glucose levels following an oral glucose tolerance test in diabetes has been well established, and thus the surrogate measure provides a solid basis (or strong biological plausibility) for the HbA1c test being able to identify people with diabetes. However, the link between *BRCA1/2* pathogenic variants and response to PARP inhibitors is not absolute (as other genes with pathogenic variants also influence the likelihood of response) and provides a weaker basis (or biological plausibility) for determining clinical test accuracy, and subsequently the clinical utility, of the test.

Information to support the comparisons

The comparison of tests, particularly comparisons involving imperfect reference standards, or incomplete use of reference standards, may benefit from supplementary information. The additional

information should seek to improve the understanding of the derived sensitivity/specificity or concordance. Additional information may be needed to explain why the proposed test results in more or fewer positive and/or negative test results. This information may be critical to determining that a reduction in sensitivity or specificity against an imperfect reference standard may be due to the proposed test having greater accuracy.

The following questions may be relevant to explore:

- How do the compared tests differ in terms of test parameters? For example, lower limits of detection or resolution.
- Is there a difference in method of classification of test results across tests? For example, this may occur if the tests use different thresholds for positivity, or access different databases for variant calling.
- Are there differences in what the tests can detect? For example, is the test designed to detect copy number variants in addition to single nucleotide variants, or non-FDG-avid tumours as well as those that are FDG avid?

If no additional information is provided to clearly indicate why the proposed test results differ from a reference standard, clinical utility standard or the comparator, the discordant test results should be regarded as false test results.

If an improvement in accuracy is expected, and can be supported by additional information, it is important to discuss whether newly positive patients (i.e. the increase in true positives due to the proposed test) have the same spectrum of disease as the positives previously detected, and if they will receive the same benefit from treatment. Discuss the possibility of overdiagnosis; that is, a diagnosis in an individual where there is no evidence that this state leads to reduced health outcomes, or a lack of evidence that management decisions will benefit the individual. Note that the best evidence for determining whether a test is more accurate than a comparator test is direct evidence of test effectiveness; that is, showing the impact of the proposed test on patient health outcomes.

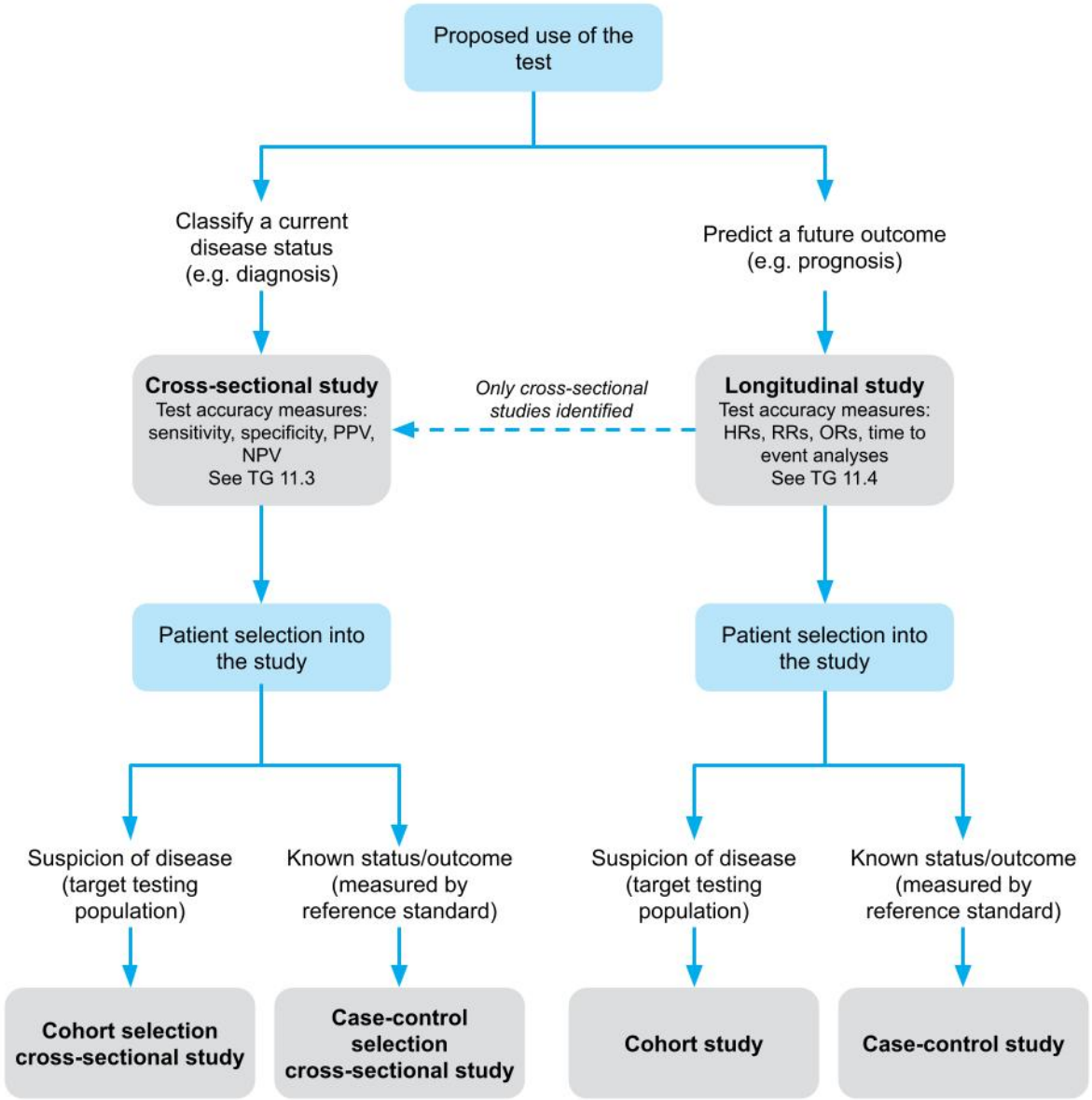
TG 11.3 Test accuracy for differing test purposes and study designs

Tests can be performed for a variety of purposes including:

- diagnosis of disease in symptomatic patients
- determining suitability for a targeted treatment in patients with disease
- monitoring of disease status/progression in affected patients
- treatment outcome assessment of affected patients
- prognosis or prediction of future disease outcomes in symptomatic and/or high-risk patients
- risk assessment in asymptomatic individuals at increased risk
  - monitoring of disease occurrence/recurrence
  - cascade testing of relatives at risk of having a heritable condition
- screening or carrier testing of the general population.

The approach taken to assess test accuracy depends on whether the test determines a current health state (e.g. diagnostic) or a future health state (e.g. prognostic or predictive) (Figure 17). These two categories of tests are often accompanied by different evidence:

- If the test determines a current health outcome, cross-sectional studies will provide the evidence base for test accuracy measures. The accuracy of the proposed test should be assessed as outlined in TG 11.4. This would include tests conducted for diagnostic purposes, as well as those used for triaging, monitoring, screening and staging.
- If the test determines a future health outcome, test accuracy should be assessed using longitudinal data as outlined in TG 11.5. Should the evidence base for a prognostic and/or predictive test include only cross-sectional studies, the test should be assessed in the same way as outlined in TG 11.4.



HR = hazard ratio; NPV = negative predictive value; OR = odds ratio; PPV = positive predictive value; RR = relative risk

Figure 17 Flowchart for determining the type of evidence and the approach to the assessment of test accuracy (adapted from Mathes et al., 2019)[26]

TG 11.4 Cross-sectional accuracy

Study designs

Cross-sectional cohort studies with consecutive or nonconsecutive patients who meet the test population defined by the PICO, and which compare both the proposed test and any comparative test against the reference standard, provide a higher level of evidence than studies with a case-control study design.[10, 27] If the evidence base is large, there may be grounds for not including case-control studies in the analysis.

Individual study results

Present individual study results. Test statistics that should be presented include sensitivity, specificity, positive and negative likelihood ratios, and the diagnostic odds ratio. These statistics can be derived from 2-by-2 tables populated during data extraction (Appendix 7).

Positive and negative predictive values of a test can be derived from a study with a 2-by-2 table; however, the estimates will only be accurate for the prevalence of the condition or biomarker in the study (which may be dissimilar to the target population). It is preferable to derive positive and negative predictive values from the pooled sensitivity and specificity values (if pooling is appropriate), and an applicable estimate (or range of estimates) for prevalence.

Briefly tabulate the sensitivity and specificity, as well as the sample prevalence, for each included study. The methods for calculating these test statistics (as well as other test accuracy measures) are provided in Appendix 7.

Assessing the transitivity and applicability of the included studies

Studies included in a meta-analysis of test accuracy should be transitive, although it is recognised that evidence on tests is often heterogeneous. The included studies should be assessed to determine whether there are any key differences between them that may affect test accuracy. These differences may include assay characteristics, sample handling, differences in interpreting the results, thresholds for determining positive results, and biological characteristics of the test population. Issues pertaining to the transitivity of studies should be considered in any meta-analyses of test accuracy data.

The same characteristics should be assessed for their applicability to the Australian diagnostic setting. If certain population subgroups are not relevant to the target population, these subgroups should be omitted from the analysis. Any concerns relating to applicability should be discussed during the interpretation of the results. Studies that are clearly not applicable should be identified as such.

Meta-analysis of test accuracy studies

While most measures of test accuracy can be pooled, the preferred approach for an assessment is to pool only sensitivity and specificity, and to derive the other measures of test accuracy from the pooled estimates of sensitivity and specificity. This is for two key reasons:

Test accuracy measures that vary across increasing/decreasing prevalence rates should not be pooled.

There are established bivariate meta-analysis methods for correcting for the correlation between sensitivity and specificity. Further guidance on bivariate meta-analyses is provided in Appendix 7.

Heterogeneity between the studies included in a meta-analysis should be explored as described in Appendix 6. The results between studies would be expected to differ due to chance alone, as a consequence of differences in the included samples taken from the entire theoretical population. Statistical heterogeneity is identified when there is more variability than expected, and is a frequent occurrence with test accuracy studies, partly due to the impact of the test thresholds. Thus, it is useful to determine the proportion of the variability that could be attributed to the threshold effect and to chance.

If differences in characteristics across studies are expected to affect test accuracy, present separate meta-analyses or subgroups within meta-analyses. Where the effect of study differences on test accuracy is uncertain, or a threshold effect is predicted, a subgroup analysis should be undertaken.

Following a meta-analysis, it is common to present an assessment of publication bias. Publication bias occurs when the outcome of an experiment or research study influences the decision whether to publish. An assessment of publication bias is described in Appendix 4.

For a more detailed discussion of the methods used for meta-analysis of test accuracy studies, see Appendix 7.

Presentation of test accuracy evidence

The key test accuracy results presented in the assessment report will depend on the types of comparisons that are required to best describe the true positives, true negatives, false positives and false negatives from the proposed test. The assessment of test accuracy evidence should include the following:

- Describe the literature search for test accuracy studies. Assess studies for a risk of bias and extract study characteristics.
- Present the results for individual studies in a table. Provide relevant test statistics, including sensitivity, specificity and the prevalence of the disease or biomarker in the study (if estimable).
- Discuss the transitivity of the included studies, and justify the separate presentation of any studies based on transitivity concerns.
- Describe the approach for meta-analysis (or narrative synthesis if meta-analysis is not possible), and discuss possible reasons for heterogeneity of results. Present results in a table or forest plot (see Appendix 7).
- Indicate whether there is a specific test threshold that should be used to determine test positivity/negativity (if applicable); that is, where the test sensitivity and/or specificity is highest for achieving the purpose of the test. Consider presenting a hierarchical summary receiver operating characteristic (HSROC) curve (see Appendix 7).
- Apply the prevalence rate of the disease in the PICO population (as determined in TG 11.9) to the pooled estimate of sensitivity and specificity to generate other test statistics (e.g. positive and negative predictive values) (see Appendix 7).
- Interpret the results. The interpretation should include a description of the comparison, and an explanation of why the comparison is important for interpreting subsequent steps in a linked evidence approach. If the proposed test is identified to have a different accuracy to that of the comparator, reference standard or clinical utility standard, discuss whether this represents an

improvement in accuracy or a loss of accuracy, and provide supplementary data (e.g. test characteristics such as scoring algorithms, thresholds, read depth) to justify the judgement.

- Explain the implications of changes in test accuracy for the management of patients (change in management evidence), and the likely impact on patient outcomes.
- Present a summary of the quality of the body of evidence (for example, by using GRADE; see Appendix 4).

Repeat this approach for all necessary comparisons (proposed test versus reference standard, proposed test versus clinical utility standard, proposed test versus comparator).

In a separate section, describe the search for sources of prevalence estimates, and present a range of estimates. Nominate the most applicable estimate of prevalence and provide justification.

Applicability of results to subsequent linked evidence

A new test that detects additional cases of apparent disease can create uncertainty about whether these additional cases should be classified and treated in the same way as current practice.[24] For example, a new test for a suspected disease may widen the spectrum of patients considered to have the disease compared with the reference standard or clinical utility standard, and the correlation between findings of the test and the eventual clinical course may be poor. This may indicate that the additional diagnosed cases are either at lower risk, with the treatment having a smaller beneficial effect, or that some patients have been incorrectly diagnosed (false positive, or overdiagnosis) and may have received unnecessary treatment. Therefore, care must be taken when assessing the health outcomes for these newly diagnosed patients.

Discuss the implications of a change in the spectrum of disease or in the accuracy of the test (compared with the comparator) when presenting evidence of change in management (Technical Guidance 12) and health outcomes (Technical Guidance 13).

TG 11.5 Longitudinal accuracy

Longitudinal accuracy is when a test is performed for the purposes of determining a future health state. The accuracy of this prediction is measured against a reference standard, which is the health outcome of interest at a later time point (e.g. length of survival, or response to treatment).

Particular attention is required for the nomination of an appropriate comparator. In many circumstances, tests that predict a future event will not have a reimbursed test as a comparator. However, clinical practice may have many validated risk scores that guide management decisions. To determine the benefit of the proposed test, the incremental benefit of the test over and above the information gained from existing risk stratification should be considered.[28]

This section describes how to approach the assessment of the longitudinal accuracy of a test.

Examples of tests that may include a longitudinal accuracy study design are:

- tests that establish a predisposition to a disease
- tests that estimate a prognosis to predict a patient's clinical course
- tests that predict a response to treatment to identify suitability for that treatment
- tests that measure an early effect on a surrogate outcome to predict a later effect on more clinically relevant outcomes.

A discussion of the types of tests that are likely to require longitudinal accuracy data is provided in Technical Guidance 15.

Study designs

The hierarchy of informative study types (in the absence of direct from test to health outcomes evidence) is shown in the NHMRC levels of evidence for prognosis.[10] MSAC will be most influenced by the results of prospective data.

Individual study results

In some cases, key measures of effect generated from longitudinal accuracy studies may be similar to cross-sectional accuracy (i.e. if the outcome of interest is a dichotomous variable). If this is the case, an approach similar to that for cross-sectional accuracy can be taken (TG 11.4).

As the purpose of a predictive or prognostic test is to provide an estimate of a future event, dichotomous variables as outcomes are less commonly used. On occasion, time-to-event data is converted to a dichotomous outcome, such as a landmark analysis. Care should be taken in the interpretation of these data to ensure that the chosen time point adequately captures the incremental differences between tests.

Longitudinal accuracy data are presented as relative risks, odds ratios, etiologic risk (population-attributable risk), logistic regression measures or interaction terms, or they may incorporate data over time through the use of Kaplan–Meier curves and hazard ratios. Estimates presented should include the *incremental* value of the proposed prognostic/predictive test (i.e. what additional value is derived from the proposed test, over and above the information that would be derived in the absence of the test).

The reference standard, or clinical endpoint of interest, must be clearly defined, including the time period of follow-up. This is a key issue as the time interval and intervening variables, such as treatments, can impact the accuracy of the predictions. Clarify what thresholds are used to determine risk classifications, and whether they reflect the thresholds that would be used in Australian clinical practice.

Considerations relevant to the assessment of longitudinal accuracy

If more than one study is required to establish the longitudinal accuracy of a test, discuss the transitivity of studies. Multiple studies are likely to be required if the key study only includes test-positive patients, or where the key study is noncomparative.

Discuss the applicability of the studies to the Australian setting. For studies that include a clinical utility standard, an important applicability issue is whether the proposed test is equivalent to the clinical utility standard.

See TG 11.4 for further details on assessing transitivity and applicability of the included studies.

Surrogate outcomes

Longitudinal accuracy studies involve time-to-event outcomes. It is not uncommon for studies to use surrogate outcomes presumed to reflect downstream outcomes. If the longitudinal accuracy study includes the use of surrogate outcomes, they must be validated. There needs to be a clear association between the two outcomes, and a biological rationale for how the two relate.[29]

For further details on the use of surrogate outcomes, see Appendix 12.

Assessing the risk of bias

It is particularly important to consider the risk of bias in the findings of included studies on predictive testing. Publication bias is likely to be a major concern for prognostic or predictive studies, as these studies are often performed using retrospective analyses of existing clinical databases, or as post-hoc analyses of trials. As such, there will be no indication that the study has been performed until it has been published.[30] Selective publication of prognostic studies is likely to result in publication of smaller studies showing larger effects, and a greater chance that positive studies are published while negative studies are not.[31]

Studies should be assessed to determine whether selective outcome reporting has occurred. For example, trials will often assess the principle outcomes of time to death (overall survival) and time to recurrence (disease-free survival); however, articles reporting on the studies will often only present one of these outcomes.[30] Trial registries should be checked to see whether additional results are available. Another area of concern is if the studies only report unadjusted results, as these are generally larger in magnitude than adjusted results and confounded by covariates. There are also concerns regarding the risk of selective reporting in some subgroups, so it should be made clear whether the subgroups were preplanned or not.[30] Likewise, the thresholds used in the validation studies must be consistent with those likely to be used in the target Australian setting, as studies which retrospectively select the 'optimal cut-off point' introduce considerable bias.[31]

The time-lag between the prognostic testing and clinically important events should be assessed to determine whether it is long enough, or whether participants in the studies are beyond the age when clinical expression is likely.[28]

TG 11.6 Concordance

Concordance analyses are useful when comparing the proposed test (to be used in Australia) with tests used in the key studies.

When the proposed test is evaluated by comparison to a clinical utility standard, the terms sensitivity and specificity may not be appropriate if the clinical utility standard is not the reference standard. However, a comparison should still present a 2-by-2 table showing how the positive and negative test results for the proposed test and the clinical utility standard compare. The positive percent agreement and negative percent agreement should be reported. These values are calculated using the same numerical calculations used to estimate sensitivity and specificity (see Appendix 7). When comparing the proposed test to a clinical utility standard, include a discussion of the discordant results as false positive or false negative compared to the clinical utility standard. The exception would be if there was compelling evidence that the proposed test is more accurate than the clinical utility standard, and the discordant results from the proposed test are more likely to be true (although this would be difficult to justify without direct test to health outcomes evidence for the proposed test).

Companion testing

Concordance is likely to be important for determining equivalence between two or more tests that could be used interchangeably in the Australian clinical or diagnostic setting to assess the same disease outcome or biomarker.

One example would be commercially available PD-L1 tests used as companion diagnostics for determining eligibility for targeted immunotherapy. In this case, the tests were originally designed to measure the biomarker under different circumstances. They were initially optimised to measure PD-L1 expression on different cell types (tumour-infiltrating immune cells versus tumour cells), using different scoring algorithms (total proportional score versus combined positivity score versus

inflammatory score) and different cell locations (cytoplasmic versus cell membrane). Thus, the level of concordance between these tests for a specific application, and the downstream health outcomes from treatment with the targeted therapy, would be important for determining the clinical utility of these companion tests.

TG 11.7 Cascade testing for heritable diseases

Cascade tests are usually a modified use of broader genetic tests that are used to identify the pathogenic variant causing disease in the index case. Thus, they are used to identify one specific genetic variant in high-risk first- and second-degree relatives. The assessment of test accuracy for cascade tests is not usually required. Further guidance on cascade testing is provided in TG 15.1.

TG 11.8 Test reliability

Determining the reliability of the test is important when determining test accuracy.

The term reliability refers to the agreement between different operators or instruments applying the same test, or the same operators applying the same test at different time points. Reliability is sometimes referred to as reproducibility or repeatability.

The reliability of a test result depends on the variability of the same person or instrument making the same test score on two different occasions (intra-observer or intra-instrument variability/agreement), and the variability between different observers or instruments (inter-observer or inter-instrument variability/agreement). Reliability might be further affected by factors such as the number of observers, or tissue storage and processing. An investigative technology that has poor reliability cannot have good test accuracy. On the other hand, good reliability does not assure good test accuracy.

Inter-laboratory variability/agreement should also be considered. However, any variability between laboratories should be mitigated (or controlled) by an appropriate National Association of Testing Authority–approved quality assurance program.

Other reliability measures include the rate or proportion of failed and equivocal test results across the included studies. In addition, other reliability measures may apply to specific tests. For example, next generation sequencing (NGS) tests should include:

- minimum read depth of included genes/regions and how this affects identification of variants
- test limits of detection for different types of sequence variants
- identification of sequence regions or variant types that the test cannot detect with the intended accuracy and precision.

TG 11.9 Prevalence of the disease or biomarker in the PICO population

Prevalence refers to the proportion of individuals in a population who have a disease or condition, and includes both new and old cases. Thus, prevalence is the product of incidence of new disease and the duration of the disease.[32]

The approach for identifying an appropriate estimate of prevalence depends on the requirements of the decision problem. The prevalence of the biomarker is relevant if test accuracy is measured as the ability of the test to identify the biomarker (using a nonclinical reference standard); the prevalence of the disease is relevant if the test accuracy is measured as the ability of the test to identify the disease (using a clinical reference standard). The prevalence of the biomarker or disease in the target population will always be required for estimating the extent of use of the investigative

technology, and in the economic analysis. Thus, even if the sensitivity and specificity of the test cannot be determined, the diagnostic yield, which is equivalent to the prevalence, is an important input in the economic analysis.

The most applicable sources for prevalence estimates of disease in the Australian setting may be administrative databases, registries or surveillance data. Some examples include the Australian Institute of Health and Welfare (AIHW), and specific disease registries.

However, it may not be possible to convert these data to the prevalence of the disease in the target population specified in the PICO. This population may consist of high-risk symptomatic patients for whom the likelihood of having the disease would be much higher than the general population.

Carefully assess the inclusion and exclusion criteria to ensure that the study population has not been enriched (i.e. narrowed to patients with a higher risk of having the biomarker or disease), and that enrolment into the study reflects the proposed use of the test in the target population. Studies that do not report how patients were selected, or that adopt a design that is not suited to estimating prevalence, should be excluded.

The primary source for the estimate of the prevalence of the biomarker or disease would most likely be studies that provide test accuracy data. The key considerations in identifying which of these studies is appropriate to include in estimating the prevalence relate to the applicability of the source to the target population. The selection of the studies should consider whether the study population is derived from:

- an Australian population or non-Australian population
- a population with similar disease characteristics
- patients at the same point in the clinical management algorithm (i.e. the same prior tests or assessments have been performed). If triaging of patients prior to the study test differs from clinical practice, the prevalence of the biomarker/disease may differ from the target populations.
- a population with similar baseline characteristics. In the case of genetic testing, gender and ethnicity may be particularly relevant.
- a population with similar risk factors for some somatic pathogenic variants or disease biomarkers. For example, different oncogenes may be more prevalent in populations with different levels of cigarette consumption.

In addition:

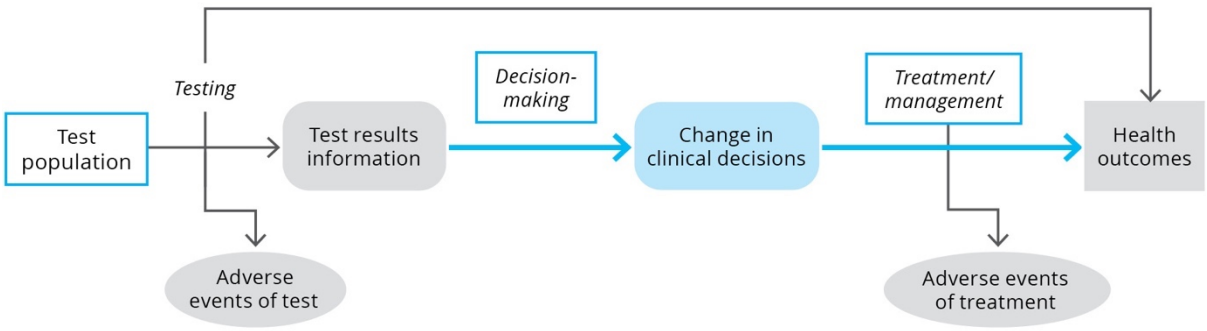
- The study population must be large enough to provide a reasonably robust estimate of prevalence.
- Prevalence of the biomarker or the disease should be measured using the reference standard or another appropriate test that is in common use in clinical practice.
- Tests have become more sensitive over time, and definitions of disease or disease stage have changed over time. Therefore, more recent studies are generally preferred.

Due to the inherent heterogeneity or variability between test accuracy studies, determining the median prevalence rate and the range may be more appropriate than determining the mean and standard deviation. If pooling of prevalence studies is required, meta-analysis methods using binomial distributions and/or transformations to approximate normal distributions (e.g. logit) would be appropriate.

In the absence of studies that completely agree with the target population described in the PICO, the most closely aligned studies should be selected for the prevalence rate estimate. It may be necessary to identify any concerns due to the inclusion or exclusion of certain patient subgroups in these studies. This is particularly relevant if the prevalence of the biomarker or disease in the study population is expected to vary from that expected in the target population.

Generally, there is uncertainty around the most appropriate prevalence rate estimate to use. A range of estimates is therefore usually applied in sensitivity analyses of the estimates of positive and negative predictive values, and in the economic model.

Technical Guidance 12 Linked evidence – change in management



KEY CONSIDERATIONS

For investigative technologies

- **Present the evidence for change in management in the same way as for a therapeutic health technology (TG 12.6).**
- **Discuss the applicability of the change in management evidence to the Australian setting (TG 12.5), as well as any confounding or other reasons for variation in clinical management in patients with similar test results (TG 12.6).**

TG 12.1 Purpose of the guidance

An impact of a test on health outcomes can only be achieved if the interpretation of the test results leads to a change in the management for a patient. This guidance describes the assessment of change in management following the use of a test. Although the guidance section is labelled 'linked evidence of change in management', evidence for change in management may also be relevant to augment direct from test to health outcomes evidence, particularly when the change in management in the direct from test to health outcomes study did not reflect current clinical practice. This TG section discusses:

- a definition of change in management evidence
- how to assess change in management evidence
- considerations relating to the assessment of change in management evidence
- how to assess the applicability of change in management evidence
- a suggested approach to presenting change in management evidence for an investigative technology.

TG 12.2 Change in management evidence

A biomarker may be used to diagnose a condition, measure disease severity, measure response to treatment, monitor patients over time or predict the prognosis of a patient.[21] The impact that the biomarker has on the clinical utility of a test (the net benefit or harm in regard to health outcomes across test result arms) depends on the series of actions and reactions that happen as a consequence of the test result. The variety of uses of biomarkers means that the method of assessing the indirect impact on patient-relevant health outcomes needs to be flexible. The ability

for a test to change the clinical management of a patient depends on many factors, such as (but not limited to):

- whether treatments are available for a disease identified, or whether treatments differ for differential diagnoses
- whether the test provides a different result from the comparator (change in diagnosis/prognosis etc., or whether it changes the spectrum of patients treated)
- whether clinicians trust the test result sufficiently to incorporate it into their 'diagnostic thinking' and treatment recommendations
- whether patients are willing to receive the treatment recommended
- whether the treatment influences patient's adherence to recommendations.

Change in management evidence may be required to determine what subsequent interventions are received following a test, or to satisfy that there should be no change in management (if the proposed test is claimed to be noninferior to the comparator test). Evidence to support change in management must incorporate the management decisions for each test result (e.g. positive or negative; or high, moderate or low risk). Change in management may also include a time component. For example, the availability of a new test may result in the same management decisions for patients, but at an earlier time point. In this circumstance, the comparative management strategies would be early versus late intervention.

The nature of the change in management may differ depending on test purpose. For example:

- A diagnostic test may result in the decision to use a treatment for a disease, rather than an alternative treatment or no intervention.
- A diagnostic test may also result in earlier treatment compared with waiting for a clinical diagnosis.
- A staging test may determine whether radiotherapy is required in addition to surgery.
- A prognostic test may determine whether a patient is likely to have disease recurrence or not (e.g. whether adjuvant chemotherapy or more intensive monitoring should be considered).
- A predictive test may determine whether a treatment is likely to be beneficial for the patient and should be initiated.
- A predisposition test may influence the rate of lifestyle modifications or adherence to surveillance.

Change in management involves several sequential steps. Evidence may represent how a test result is interpreted (diagnostic thinking), what recommendations are made, and what is adopted by patients (i.e. the actual change in management). It is important to discuss the limitations of the evidence based on earlier steps of change in management, as changes in diagnostic thinking or recommended treatment might differ from the actual treatment received (Figure 18).

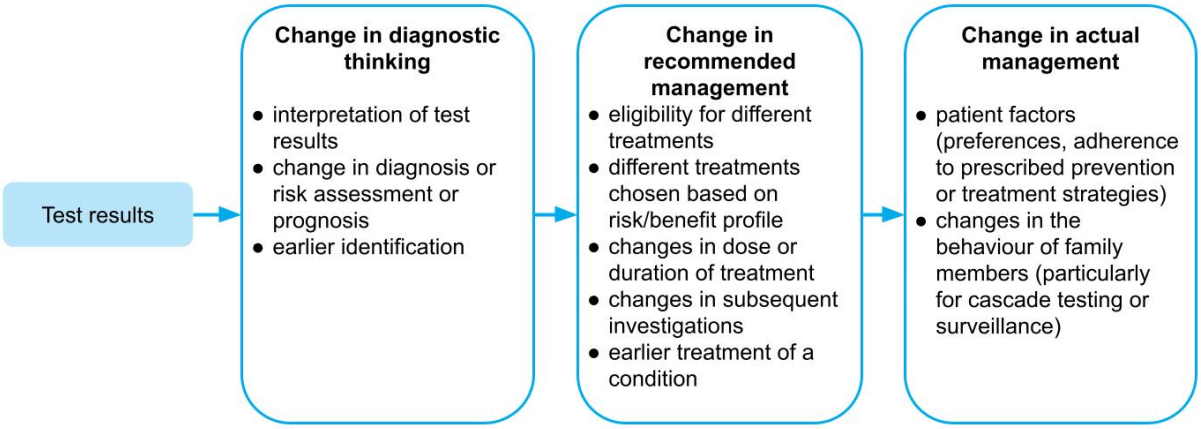


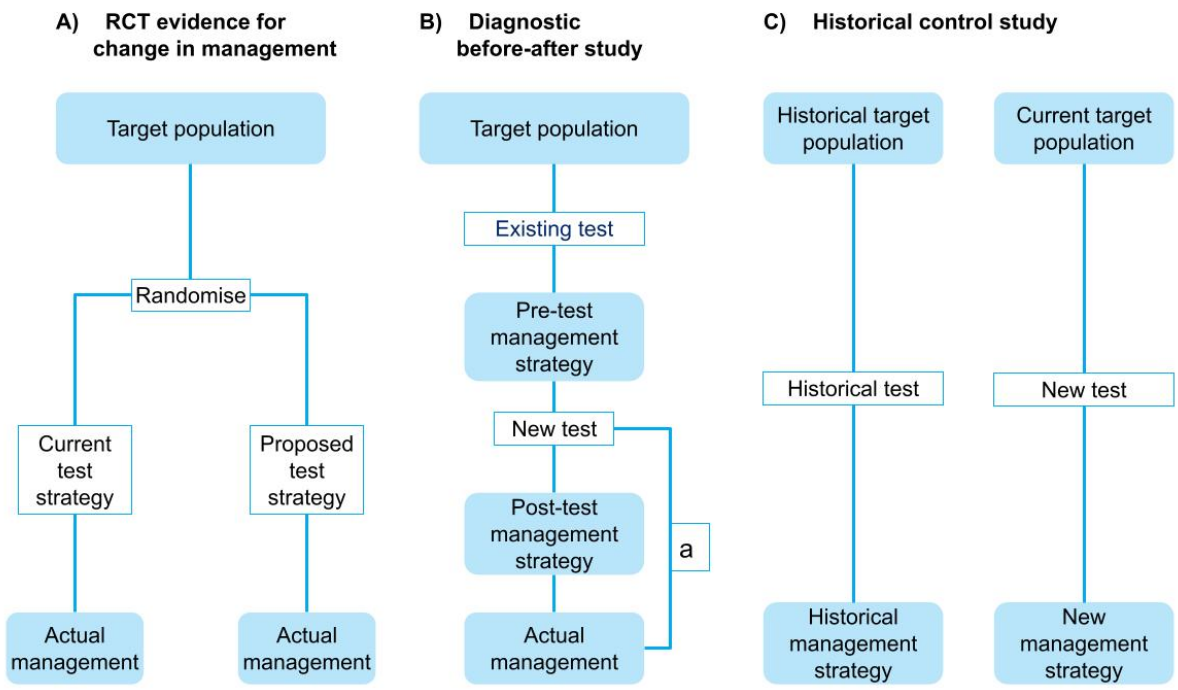
Figure 18 **Change in management components from test results to actual management undertaken**

There may be some instances where change in management evidence is unlikely to be reported. One situation would be when a test is clearly performed in order to be eligible for a treatment (e.g. a test–drug codependent technology). Another example is where a large gene panel is used to detect a wide variety of conditions, and the impact that a test has is so disparate that authors do not attempt to collate the recommendations and management that follow. In these cases, change in management may need to be assumed, based on the diagnostic yield or change in diagnostic thinking.

TG 12.3 Change in management study designs

Change in management studies may be either experimental or observational. Studies that report actual management provide more directly relevant information than those that report hypothetical planned management. Therefore, a randomised trial, which reports *actual* management following the use of the proposed test strategy versus the comparative test strategy (Figure 19 – A), contains less risk of bias than a within-patient comparison of a pre-test management plan and post-test management received. Randomised trials are suitable for all test types (e.g. replacement tests, add-on tests, triage tests). If there is no longer clinical uncertainty that the test is beneficial (i.e. if the test has become part of 'best practice'), other outcomes such as patient acceptability and safety of the tests can be reported.[33]

Despite the advantages of randomised trials, the most common design for change in management studies is the observational 'diagnostic before-after' study (Figure 19 – B). This study design is useful when the test is an add-on to an existing test strategy, that is, the existing test strategy matches the 'before' component, and the proposed test strategy matches the 'after' component with the addition of the new test. Because this cannot be the case if the proposed test is to be added *before* the existing test strategy, the before-after study design is not suitable for replacement or triage tests. Diagnostic before-after studies are subject to bias, as the clinician may not use the same amount of caution in determining the pre-test management plan if they know they will subsequently receive the result from the proposed test.[33] It should be made explicit whether the management plan in the study was made prospectively by the clinician, or retrospectively based on case notes. Measuring the concordance between the post-test management plan and the actual management received may indicate how planned management is put into practice.[33]



RCT = randomised controlled trial

a Diagnostic before-after studies may also capture actual management to validate the post-test management strategy. In some cases, the post-test management strategy may not be included, with the study reporting on only the pre-test management strategy and the actual management following the test.

Figure 19 Change in management study designs (adapted from Staub et al., 2012)[33]

Another study design that may be informative for change in management outcomes is the historical control study, which reports practice prior to and after the introduction of the new test (Figure 19 – C). Historical control studies are at risk of bias due to factors *other* than the test also changing at the same time (e.g. more effective treatments becoming available, or recall bias). In a similar manner, cohort studies with a concurrent control are also likely to be biased. If the cohort study compares 2 settings, which use or do not use the proposed test, there is the risk that the settings will not be similar enough to be able to attribute the changes in management to the test, instead of confounding factors. If an observational study compares individual patients in whom the test is used with those in whom the test is not used, there is the risk that there will be strong selection bias that could influence the results.

If evidence is not available to demonstrate that changes in management occur (i.e. an absence of evidence), expert opinion from a relevant provider of the service will be required to supplement the evidence review, to justify any assumptions regarding the impact a test may have on the behaviour of clinicians, patients, family members etc.

Change in management studies are assessed in the same way as therapeutic studies, using the hierarchy of ideal study designs in the NHMRC levels of evidence for interventions,[10] and risk of bias tools relevant for therapeutics.

TG 12.4 Considerations relevant to change in management

Risk assessment

If a test is used to classify patients into a high- or low-risk group, with different treatment indications, the consequences of being reclassified will differ depending on the upwards or

downwards reclassification. In some circumstances, there will only be a change in management if the reclassification happens in one direction (e.g. current practice is to treat all patients, but the proposed test identifies very low risk patients who may not need treatment). Presenting the results for the different subgroups is therefore helpful, as the overall impact on the whole study sample will poorly reflect the results of the individual subgroups.[22]

Addressing change in management as the first step

In general, when performing a linked evidence approach that requires the complete assessment through to health outcomes, the assessment should focus on evidence to support a change in management in the first instance. A lack of evidence to support change in management for claims of superior health outcomes will require additional justification (e.g. clinical expertise) and a thorough discussion of assumptions.

Need for empirical evidence

While it is always ideal to support a linked evidence approach with good quality evidence for change in management, there are 2 examples where change in management data should be measured and robustly assessed:

- *Tests used for monitoring*: A monitoring test is the observation of a parameter over time. In the absence of direct from test to health outcomes evidence, clear evidence of the impact of monitoring on change in management is required. This may include observations that patients start or stop treatment, change the dose or duration of treatment, or take some other action. Compared with tests that commonly have a clearly defined purpose and threshold for action, monitoring tests may not necessarily result in changes to treatment, or may trigger further investigation(s) that ultimately do not lead to changes in treatment.
- *Tests used for outcome monitoring*: Outcome monitoring describes a test used to determine response to an intervention. An example may be a CT scan used to determine whether a medicine is having the desired effect on a tumour. While monitoring for response to treatment may commonly be used as part of stopping rules in clinical studies, it is not guaranteed that such stopping rules will be applied in clinical practice. Clinicians may be reluctant to withdraw a treatment if the viable alternatives are limited (e.g. in the case of later line therapies), or if the treatment is perceived to provide a prophylactic mechanism (e.g. continuing glucocorticoids following resolution of a COPD [chronic obstructive pulmonary disease] exacerbation). As such, the results from clinical studies that employ clear stopping rules cannot be used to inform change in management in clinical practice, and empirical evidence of change in management is required.

In some circumstances, empirical evidence from studies reporting change in management may be unnecessary, for example, where earlier diagnosis of a serious disease is highly likely to result in earlier treatment for that disease, or identification of a biomarker by a codependent test leads to the use of a targeted medicine.

Impact of the change in management on the health system

If the proposed test results in changes in management of patients, this may also have an impact on health care providers for the intervention or the comparator, or downstream investigations/treatments etc. For example, if a triage test reduces the number of patients being referred to a specialist, this may have an impact on the specialist workforce, waiting lists etc. These flow-on impacts should be discussed under ‘Other relevant considerations’ rather than ‘Change in management’.

TG 12.5 Assessment of the applicability of change in management evidence

Investigative technologies depend on the downstream consequences in order for health outcomes to change. These downstream consequences can vary greatly in different settings, as organisational factors may affect their implementation and uptake. It is therefore important to consider how applicable the evidence is to the target Australian population and setting. If the change in management evidence is derived from a different setting to where it is proposed to be used, the evidence may not be applicable. For example, if a test is proposed to be used in the general practice setting, but most of the evidence is derived from a specialist setting, this needs to be raised as an uncertainty and the applicability explored with subgroup analysis or supplementary evidence.

Variations in management decisions occur across countries and within Australia.[34] Causes of variations in management decisions may be related to medical opinion,[35] be clinically driven or be influenced by nonclinical factors.[36] It may be useful to consider 4 key categories of factors that, should they differ across settings, may influence management decisions and therefore the applicability of change in management evidence from other settings (Table 10).

Table 10 Applicability issues relating to evidence for change in management

Causes of variation in management	Description of factors resulting in variation in management
Health system factors	Differences in referral systems, payments, remuneration or incentives may influence clinical practice. Geographical barriers or access (e.g. highly specialised care facilities vs rural facilities) or distribution of clinicians may also influence decision-making.
Supply-related factors	Differences in the availability of technologies across the settings, both in terms of regulatory (market access) availability, as well as whether technologies are subsidised differently across settings. Clinicians in Australia will be more inclined to prescribe technologies that are available in their public hospital, or that are subsidised by the Commonwealth government, such as those technologies available on the Medicare Benefits Schedule or Pharmaceutical Benefits Scheme. Prescribers in other countries may likewise be compelled to recommend technologies that are subsidised by government or insurance.
Demand-related factors	Differences in culture, ethnicity, personal beliefs and values, and patient expectations may influence clinical management decisions, or the adherence to management decisions. Education of patients and medical advertising can influence patient expectations.
Need-related factors	Differences in population health, indicators of which may include age or demographics, socioeconomic status or environmental factors

Source: Derived from Australian Commission on Safety and Quality in Health Care[34]

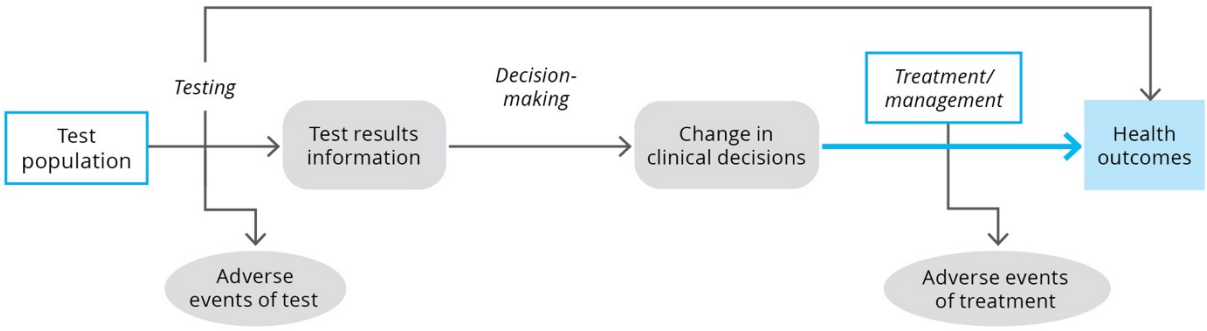
TG 12.6 Presentation of change in management evidence

The principles for presenting change in management evidence are similar to those for presenting clinical study evidence relating to a therapeutic technology. The key result that is sought by the assessment is the extent to which management changes (and the nature of the change) in a circumstance where the proposed test is available compared with when it is not available (and a comparator test or usual practice is applied). In addition to the main results for change in

management, the assessment of change in management evidence should include these additional points:

1. Present the evidence for change in management in the same way as presenting evidence for a therapeutic technology (i.e. describe the literature search, risk of bias and trial characteristics, present the results, and meta-analyse, if appropriate).
2. Discuss reasons for variation in clinical management in patients with similar test results. Discuss whether the change in management may be confounded by other patient factors rather than the test results.
3. Discuss the applicability of the change in management evidence to the Australian setting. Where the evidence for change in management is partially applicable to the Australian setting, explore, where possible, the variation in management across subgroups, or present supplementary evidence to support the generalisability of the study results across settings.
4. Present a summary of the quality of the body of evidence (for example, by using GRADE; see Appendix 4).

Technical Guidance 13 Linked evidence – health outcomes



KEY CONSIDERATIONS

For investigative technologies

- **The principles for presenting health outcome gains evidence are similar to those for presenting clinical study evidence for a therapeutic health technology (TG 13.6).**
- **Discuss the applicability of the health outcome gains relating to the change in management in the target Australian setting (TG 13.5). Include the impact of applicability issues, as well as any confounding or other reasons for variation in clinical management in patients with similar test results (TG 13.6).**

TG 13.1 Purpose of the guidance

Demonstrating that a test affects health outcomes provides the most confidence for MSAC to support the utility of the test. If direct from test to health outcomes evidence is available demonstrating that a test improves clinical outcomes compared to the comparator, guidance for presenting that evidence is provided in Technical Guidance 10. More commonly, evidence of health outcomes is demonstrated through a linked evidence approach, showing that a test changes clinical thinking, management recommendations and treatments received. The last step of the linked evidence approach (called therapeutic effectiveness evidence) is to establish the impacts of the change in management on health outcomes.

This TG section will discuss:

- a definition of therapeutic effectiveness evidence
- therapeutic effectiveness study designs
- considerations relating to therapeutic effectiveness evidence
- how to assess the applicability of therapeutic effectiveness evidence
- a suggested approach to presenting the therapeutic effectiveness as the final step in a linked evidence approach for a test.

TG 13.2 Therapeutic effectiveness evidence

Therapeutic effectiveness evidence, as the final step of the linked evidence approach, includes an estimate of the impact of all the management decisions made as a consequence of using the proposed test in the place of standard practice.

In general, therapeutic effectiveness evidence should attempt to derive the highest quality evidence for the incremental difference in outcomes associated with treatment decisions informed by the

proposed test versus treatment decisions informed by the comparator test. If therapeutic effectiveness evidence achieved this goal without concessions, it would resemble direct from test to health outcomes evidence.

In practice, the assessment of therapeutic effectiveness rarely achieves the certainty of direct from test to health outcomes evidence, and relies on assumptions relating to the generalisability of the evidence across differently selected populations.

Therapeutic effectiveness evidence may provide an estimate of the impact on health outcomes for individual test populations, but may not provide an estimate of the magnitude of the impact on health outcomes of the proposed test compared with standard practice. This is because there may be multiple sources of evidence for different populations, and aggregating the overall health outcomes cannot be easily performed without a decision analytic model that links together test accuracy, change in management and treatment effect.

Therapeutic effectiveness evidence has several parts as shown in Table 11.

Table 11 Assessent of therapeutic effectiveness evidence

Assessment question	Description
Is there a treatment available?	Identify whether there are management strategies or treatments available for each of the test populations (this is informed by the change in management link).
Is there evidence that treatments are effective?	Identify evidence that the treatments are effective for the appropriate indication.
What are the implications for false positives and false negatives?	Discuss the implications of misclassification (false positives and false negatives) on the health outcomes.
Is there evidence or a risk of a change in the spectrum of disease?	Consider whether the evidence for treatment effectiveness can be generalised from an unselected or differently selected population to the new test categories (including a discussion of whether the spectrum of disease following testing has changed).

Potential areas of concern in therapeutic effectiveness evidence relate to the applicability of the available evidence to each of the populations identified by test strategies, the transitivity across the evidence, and the subsequent impact on the validity of the economic analysis.

The key clinical uncertainty associated with the final step in a linked evidence approach is that the treatment outcomes are not commonly derived from patients with a known test status. In many cases, treatment outcomes are sourced from unselected populations or differently selected populations, and it may remain unknown whether the results are generalisable to the test-positive or test-negative populations for the proposed test.

The identification of suitable health outcomes evidence is an iterative process involving:

- identifying whether there is a treatment or management strategy for the target condition
- identifying whether there is evidence that the treatment works in the target condition
- identifying whether there is evidence that the treatment works for the test subgroups

- assessment of the uncertainty or gaps in the evidence (applicability to the target population, generalisability of the evidence from an unselected to a selected population)
- assessment of the impact of the uncertainty (direction of the effect of applying the identified evidence)
- identifying supplementary data to support or reject the use of the identified evidence.

The process is iterative because it may not be apparent that the identified evidence is suitable until an assessment of the evidence for applicability and generalisability has been performed. Following this assessment, should the evidence be rejected, further searches may be required. It is not expected that these searches are performed systematically, rather, that targeted searches are performed to try and identify the highest level, or best, evidence that addresses the impact of the change in management. For more information on literature search methods, see Appendix 2.

TG 13.3 Therapeutic effectiveness study designs

Study designs required to provide therapeutic effectiveness data vary depending on the results from the change in management evidence. In general, the types of included study designs are guided by availability of the evidence rather than what might be ideal. Studies may include randomised controlled trials, systematic reviews of randomised controlled trials, observational studies or registry data.

The following general guidelines may assist in determining the types of studies that may be useful:

- Comparative studies are useful to explore the impact of changing from one management strategy to another.
- A relative treatment effect *is not* useful to describe the differences between treatments that are prescribed for different test populations (e.g. positive and negative biomarker status).
- Observational studies are useful for determining the natural history of the disease.
- Studies comparing the outcomes of the same treatment by biomarker status may also be useful in identifying whether a patient's biomarker status has a prognostic effect. Understanding whether the biomarker is prognostic or not may inform whether evidence from unselected populations receiving treatment A can be generalised to test-selected populations using treatment A.
- Studies reporting on subgroup analyses defined by population characteristics or biomarker status may be useful for determining the applicability of the evidence to the target population.

TG 13.4 Considerations relevant to linked evidence of health outcomes

Generalisability of the evidence

Unless a new test is substantially safer (and avoids adverse events), for a new test to have an impact on health outcomes, it must result in a change in management and alter the allocation of patients across treatments. If the test is relatively new, there is unlikely to be evidence for the outcomes of patients allocated to treatments according to results of the proposed test. Therefore, health outcomes evidence for the treatments identified in the change in management section may not be generalisable to the population receiving that treatment following the use of the proposed test in practice. Figure 20 is a simple diagram showing the discordance.

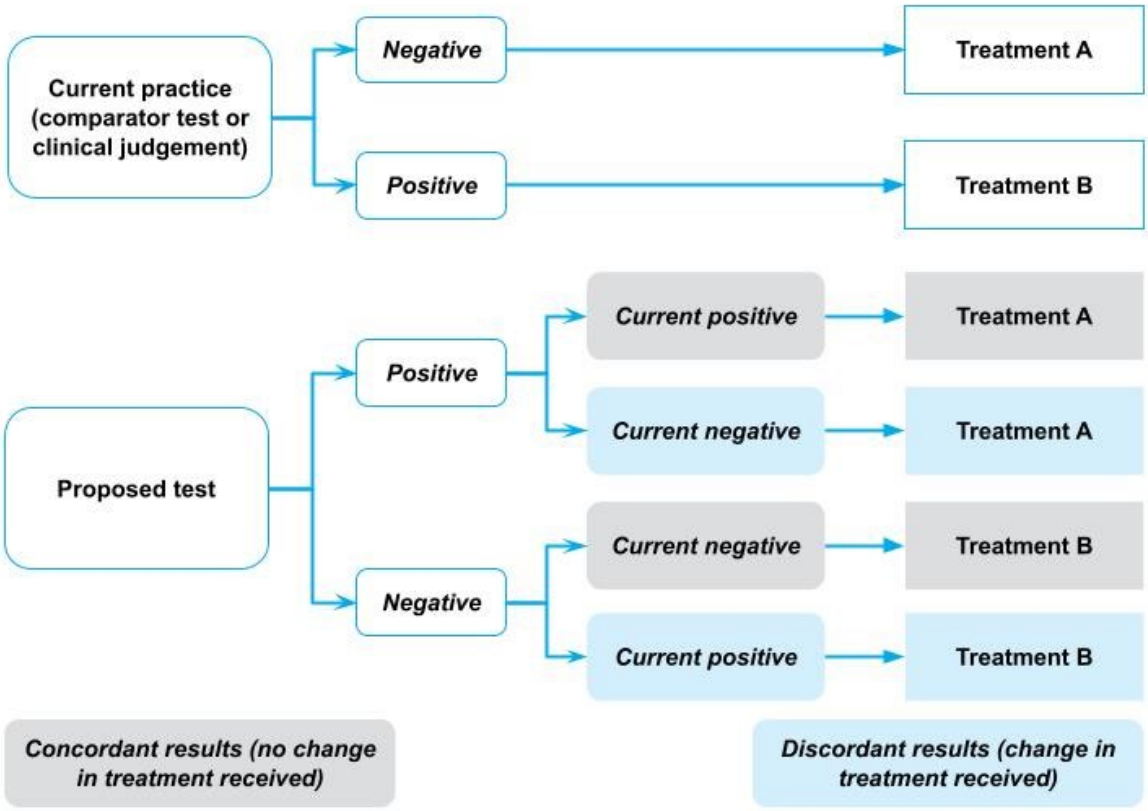


Figure 20 **Diagram showing the change in the treatment of patients categorised using the proposed test compared with current practice**

In Figure 20, current practice would allocate patients to established treatments A or B. Current practice may reflect current testing or clinical assessment alone. The treatments provided in the diagram above may be any type of treatment decision (e.g. the decision to give 2 different treatments, adjuvant therapy, an add-on treatment, or to withhold treatment).

The figure shows that, using the new test, some patients who previously received treatment B will now receive treatment A, and some patients who previously received treatment A will now receive treatment B.

In general, evidence will be available for the treatment outcomes for patients receiving Treatment A and for patients receiving Treatment B. A key concern is whether the evidence, which contains some patients who would be allocated differently by the proposed test compared with the comparator test, can be used to approximate the health outcomes for the proposed test. Not only may the treatment effect for patients differ depending on their test result, but the prognosis of patients with different test results may also differ. An assessment of the generalisability of the evidence to different populations should therefore include the following:

- a description of the patients who change management – Do these patients differ from those who did not change management? For example, are the patients who test positive with the proposed test similar to those who test positive with the comparator test? Do any differences indicate a change in the spectrum of disease being identified?
- evidence for whether the test status is related to prognosis.

Spectrum of disease

A change in the spectrum of disease being identified may occur when a new test is introduced, such that either less severe (more common) disease or more severe disease is identified. For example, a test that is more sensitive may detect disease earlier, or detect disease that is less severe or even inconsequential (i.e. overdiagnosis).

When considering the generalisability of the evidence, it is important to consider whether the proposed test has resulted in a change in the spectrum of disease being identified. A change in the spectrum of disease may result in:

- increased lead time (earlier detection of a parameter)
- change in the efficacy of treatments (either in relative or absolute terms).

If there is a risk that the patients treated following the proposed test differ in the spectrum of their disease compared with the patients treated in the identified therapeutic effectiveness evidence, the evidence may not be generalisable.[20] An assessment of the possible impact of a change in the spectrum of disease identified should seek to first define how the spectrum of disease has changed.

If the proposed test is more sensitive, more patients will be identified as positive. In terms of test accuracy, this reflects the current categorisation of false-negative patients moving to a true positive state. Evidence of the treatment effect in these additional patients, who may have less severe or earlier stage disease, may be required. A comparison of early versus late treatment may be informative.

If the proposed test is more specific, fewer patients will be identified as positive. In terms of test accuracy, this reflects the current categorisation of false-positive patients moving to a true negative state. Evidence of the harms of inappropriate treatment of negative patients may be required.

Although not exclusively related to a change in the spectrum of disease, detecting earlier or less severe disease leads to a risk of overdiagnosis. Overdiagnosis may occur if identification of an abnormal disease state leads to an individual being erroneously classified as having a disease. There may be no evidence that this state leads to a poor health outcome or that further investigations/treatments will be of benefit; in contrast, there may be evidence of potential harm.[37] Five indicators may be used to identify potential overdiagnosis:[38]

- Is there potential for increased diagnosis?
- Has diagnosis actually increased?
- Are additional cases subclinical or low risk?
- Have some additional cases been treated?
- Might harms outweigh benefits?

Overdiagnosis may result in unnecessary health care and harm. Therefore, it is important that evidence is presented on the likely future benefits or harms of identifying a condition as an abnormal disease state.[39]

Risk of bias and transitivity

The assessment of risk of bias is important for determining the internal validity of studies identified for establishing the health outcome gains resulting from management changes. Care should be

taken when assessing risk of bias, particularly if the use of the identified study for the purposes of the assessment report differs from the original assessment question of the study. This may arise if:

- only part of the identified study is used, such as a single arm of a randomised controlled trial
- subgroups are used to address the health outcome gains of the management strategy (this may occur if biomarker subgroups are identified in the trial and are used to determine the treatment outcomes).

The assessment of risk of bias should reflect how the study was applied in the assessment report, rather than the original intent of the study.

A second key concern will arise if more than one source of evidence is required to assess the health outcome gains of all the management options (which is highly likely). Under these circumstances, it is important that the differences in health outcomes across studies is a consequence of the different treatments, and that the populations differ only if the evidence is intended to represent different populations (i.e. biomarker negative or biomarker positive). Within the economic analysis that will be informed by the health outcomes and treatment effects derived from the health outcome evidence, the different sources of evidence may be applied independently, with outcomes aggregated in the test and comparator arms. This is similar to a naive indirect comparison, and therefore the transitivity of the evidence presented for the health outcome gains link is important.

Regression to the mean

It may be necessary to source single-arm evidence to inform the therapeutic effectiveness link. Single-arm evidence can be subject to regression to the mean. If patients are selected for treatment based on the severity of the condition, there is the chance that patients will improve due to regression to the mean.[40] Assessing the effectiveness of selecting patients who will benefit from treatment therefore needs to consider whether the patients could have improved anyway without treatment (this is only likely if the condition is episodic or relapsing/remitting). If randomised evidence is not available, preference should be given to studies where the baseline measure of the outcome variable is separate from the measurement used to select patients.[40]

TG 13.5 Assessment of the applicability of health outcome gains evidence

Therapeutic effectiveness evidence can differ from the target setting in multiple ways, and the applicability of the evidence is markedly influenced by the applicability of prior steps in the linked evidence approach.

The following applicability concerns may be considered:

- Test-related applicability – Are testing components in the health outcome gains evidence (proposed, comparator or standard practice) the same as the current or proposed clinical practice? (see Technical Guidance 11)
- Change in management applicability – Is the interpretation of the test results, and change in management in the health outcomes evidence the same as the current/proposed clinical practice? (see Technical Guidance 12)
- Health outcome gains applicability – Assess the applicability of the health outcome gains evidence in the same way as assessing applicability for a therapeutic intervention (comparison of the evidence with the Australian setting for patient characteristics [demographics, disease] and intervention characteristics [dose, duration, setting]) (see Appendix 6).

TG 13.6 Presentation of health outcome gains evidence

The principles for presenting health outcome gains evidence are similar to those for presenting clinical study evidence relating to a therapeutic technology. The key results presented in this section will depend on how the evidence for comparing the treatment outcomes across the proposed test and comparator test strategies is constructed. The assessment of health outcome gains evidence should include the following points:

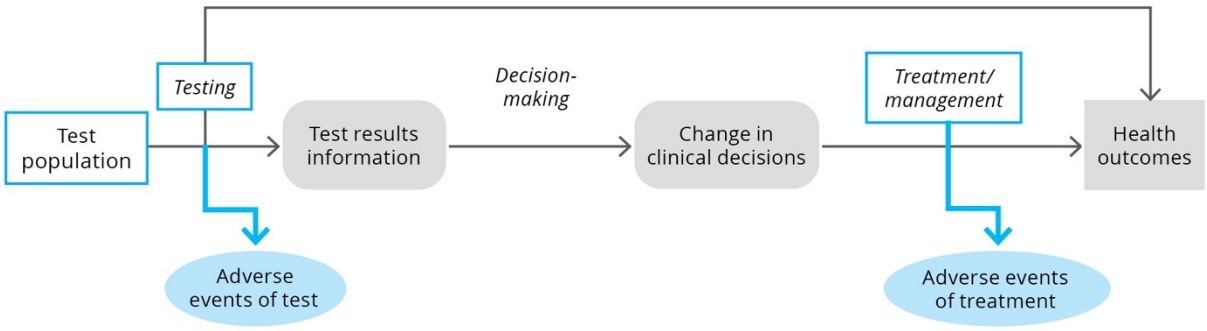
1. Explain how evidence for each of the management pathways has been constructed. Clearly identify when outcomes for one group are based on a relative treatment effect compared to another group, and when the evidence is derived from different sources.
2. Present the evidence for health outcome gains (for each different management pathway) in the same way as presenting evidence for a therapeutic technology (i.e. describe the literature search and any subsequent searches for supplementary evidence to explore uncertainties in the evidence, risk of bias and trial characteristics, present the results, and meta-analyse, if appropriate).
3. When presenting the results, provide an assessment of the outcomes relating to the change in management (e.g. if a test results in 20% of patients receiving Treatment A instead of Treatment B, a comparison of Treatment A versus Treatment B is appropriate).
4. Clearly describe the assumptions required to generalise evidence across groups (e.g. the treatment effect is assumed to be the same for patients who test positive using both the comparator test and the proposed test).
5. Present evidence to support the generalisability of the evidence across different populations. This may include evidence to address the risk of a change in the spectrum of disease identified, or a change in the prognosis associated with the biomarker. If there is a change in accuracy that may alter the spectrum of disease, clearly discuss the implications of changing treatment on test positives (particularly the new true positives) and test negatives (particularly the new true negatives). Where appropriate, include a discussion of prognosis and overdiagnosis.
6. Discuss the applicability of the health outcome gains evidence to the target Australian setting. Include the impact of applicability issues identified at the test accuracy step or the change in management step. Explore the impact of applicability in subgroup analyses.
7. Present a summary of the quality of the body of evidence (for example, by using GRADE; see Appendix 4).

If the outcomes are considered surrogates or intermediate outcomes, rather than critical outcomes of value to patients, the link between these outcomes and patient-relevant outcomes should be assessed (see Appendix 12 on surrogate outcomes).

Preference will be given to evidence of health outcomes where the treatments provided are available to patients in Australia. The health benefit of a test that is only demonstrated through the subsequent use of a treatment only available in a trial setting should not be considered.

The safety of any downstream effects of the test should be discussed as part of the health outcome gains. For example, if the test is a triage test to rule out invasive testing in those who do not need it, the safety of the further testing should be discussed. If the test results in different proportions of patients receiving treatments than would have been treated had they received the comparative test strategy, any adverse effects of treatments received should be assessed.

Technical Guidance 14 Safety of investigative technologies



KEY CONSIDERATIONS

For investigative technologies

- **If no quantitative safety data can be presented, discuss why the direct safety of the proposed test versus the comparator test is considered the same (or identical) (TG 14.1).**
- **Discuss how the change in management will alter the safety outcomes in patients being exposed to different treatments (TG 14.2).**
- **Summarise the safety issues for patients with a positive test result and a negative test result, and discuss any implications for patients with false positive or false negative test results (TG 14.2).**
- **If the included evidence is not sufficient to capture long-term or rare adverse events, or adverse events in patients with comorbidities or receiving concomitant treatments, an extended assessment of the safety of the test is required (TG 14.3).**

An assessment of the impact from use of a health technology includes an assessment of its relative safety versus the main comparator. The assessment of safety has 3 key parts for investigative technologies:

- assessment of the direct and more immediate impacts (adverse events) of the use of the health technology (often captured to a varying degree in the included clinical studies)
- assessment of downstream implications related to the management decisions following a test
- assessment of longer term or rarer safety events unlikely to be captured in clinical studies.

The objective of an assessment of safety is to identify the relative safety of performing the proposed test versus the main comparator, which may be an alternative test or no test. The assessment of the safety of a test (or the comparator) involves assessing both the direct (often immediate) harms associated with the test itself, and harms associated with downstream consequences of testing. It is important to present the direct harms of testing and harms of downstream management decisions separately. However, a narrative synthesis that discusses the entire safety profile may be informative, particularly in cases where there is a trade-off (e.g. the proposed test has an inferior safety profile compared with the comparator, but using the test results in improved safety outcomes for treatment).

Comparative safety data may be available from direct from test to health outcomes evidence, or from test accuracy studies. However, additional searches for test safety may be required.

TG 14.1 Test-related adverse events

The direct harm of testing is the immediate or delayed safety consequences of physically performing the test. For most tests (particularly in vitro tests), the direct safety of a test is related to the method of performing the test or retrieving samples. In many cases, the mode of testing for an investigative technology is already established (e.g. imaging using a CT scanner, a biopsy, or a blood test).

Assessment of the direct safety of testing may not be necessary if:

- the proposed test uses the same modality as the comparator test it is intended to replace
- the proposed test will be used in the same proportion of patients.

A discussion of why the direct safety of the proposed test is considered the same as (or identical to) the comparator test is required if no quantitative safety data are presented.

Assessment of the direct safety of testing will be necessary if:

- the proposed test is used in addition to current testing
- the proposed test will be used in a greater number of patients (due to acceptability of the test, or the need for additional samples/repeat biopsy)
- the proposed test is performed using a different modality to current testing (e.g. a change in biopsy techniques, a change in how invasive the sample retrieval is, a change in the extent of radiation delivered for scans, or a change in patient factors due to tests being performed at different time points).

Where comparative data are available, these should be presented as dichotomous outcomes.

Report adverse event data as both the number of patients reporting an adverse event in each category and the absolute number of adverse events in each category. The absolute number of events in each category may be a more appropriate estimate for costing adverse events in an economic or financial analysis because the number of patients who experience an adverse event will not capture patients who experience 2 events in the same category.

For each important adverse event, present the results as for dichotomous data, and include relative risks and risk differences with their 95% confidence intervals across the groups. Present each study separately. Where appropriate, meta-analyse the results using a random effects model and provide an interpretation.

Analyse the relative adverse event rates (events per period at risk), if the average period at risk per participant varies substantially between treatment groups (e.g. using a straight Poisson regression or a negative binomial approach). Present the assumptions associated with statistical analyses and how they were tested.

If the evidence for comparative safety of a proposed test is insufficient, the assessment should provide some context for the likely safety profile of the proposed test and the comparator by searching for good quality evidence of the safety of tests that use the same method. For example, studies of a CT scan of a similar region of the body in a similar aged patient may provide some indication of the safety of a proposed CT test, or studies of a biopsy of the same organ in patients with similar performance status or tumour stage may provide some indication of the safety of a proposed pathology test.

Supplementary evidence is only required if the identified safety evidence for the proposed or comparator test is insufficient or uncertain. The search for this evidence does not need to be systematic, but should aim to identify high-quality studies that are applicable to the Australian setting. If supplementary evidence is included, discuss the applicability of the evidence to the assessment of the proposed test.

TG 14.2 Downstream safety consequences

The use of the proposed test may result in patients receiving different treatments compared with the use of the comparator. This change in management will result in patients being exposed to different safety profiles associated with treatments. The assessment of safety outcomes for treatments informed by testing is similar to the assessment of efficacy outcomes for a therapeutic technology (see TG 7.1). Considerations relating to direct from test to health outcomes evidence (Technical Guidance 10) or to linked evidence of the health outcomes (Technical Guidance 13) are relevant for the assessment of safety outcomes.

Summarise the safety issues for patients with a positive test result and a negative test result separately. If there is a reason that the type, severity or number of safety events would differ for true and false positives, or true and false negatives, explain why and describe.

The overall summary of evidence (described in Technical Guidance 16) should clearly describe the safety issues for false positives and false negatives alongside the impact of misclassification on treatment outcomes.

TG 14.3 Test safety unlikely to be captured in clinical studies

Assessment of longer term or rarer safety events that may occur beyond clinical studies may be relevant if the change in management involves a therapeutic technology, particularly if the therapeutic technology is novel or being used in a new indication.

An extended assessment of the direct safety of a test may also be useful if a novel mode of testing is used, or there is uncertainty about the longer term or rare effects of testing.

Ideally, the estimate of the relative safety of a health technology is derived from high-quality comparative studies. Clinical trials are often inadequate for providing data on comparative harms for several reasons:

- Trials tend to enrol patients who are healthier, or have fewer comorbidities or concomitant medications, and they have more stringent monitoring than would occur in the target population.
- Trials are usually underpowered and of insufficient duration to detect long-term or rare adverse events.
- Adverse events in clinical trials designed to emphasise efficacy results are often underreported.[17]

Discuss whether the included evidence base is adequate for identifying:

- less common adverse events or safety concerns
- adverse events that may occur in the longer term
- harms that may occur due to differences between the target population and the more selected population that may be enrolled in a clinical trial.

If the included evidence is not sufficient to capture long-term or rare adverse events, or adverse events in patients with comorbidities or receiving concomitant treatments, present additional evidence. Describe the search strategy for identifying nonrandomised studies of the proposed health technology, or registry data. If appropriate, include evidence for safety of the proposed health technology in other indications. Where the proposed health technology is delivered in combination with an implantable device, provide an assessment of the safety of that device. Sources of safety information may include device registries, regulatory databases, complaints registries and post-market surveillance studies.

Technical Guidance 15 Special cases

KEY CONSIDERATIONS

For investigative technologies

- **Screening can encompass broad or narrow populations. Considerations relevant to the assessment of universal screening programs, targeted screening, predisposition testing and cascade testing will differ (TG 15.1).**
- **Tests used to determine prognosis may have value relating to improved health outcomes if they inform change in management, or only have benefits that fall within the concept of value of knowing. An assessment of a prognostic test reports how well a test can differentiate future health events compared with current prognostic (often clinical) tests, as well as how this will impact clinical management (TG 15.2).**
- **Predictive tests are used to determine how well a patient would respond to a treatment. Typically, these tests identify a biomarker that inform the eligibility for a treatment, and is most commonly applied in a co-dependent context. A critical concept within predictive testing is the clinical utility standard, and an assessment includes a robust comparison of the proposed test against the clinical utility standard (TG 15.3).**
- **The assessment of a test used for monitoring is similar to the assessment of tests for other purposes, using either a direct or linked evidence framework. However, additional aspects of the test, such as its responsiveness to changes in response to an intervention (signal to noise ratio), the detectability of long-term change, and the ease of use and interpretation of the test, should be discussed (TG 15.4).**
- **An assessment of a multifactorial algorithm involves the presentation of the biological plausibility, the characteristics of the training and validation sets (for establishing applicability), and additional results to ensure that algorithms (particularly dynamic algorithms) are not subject to bias when used in the Australian population (TG 15.5).**
- **A codependent submission is required when the Minister for Health requires advice from 2 different expert advisory committees because listing of the codependent technologies involve 2 separate reimbursement schemes (TG 15.6).**

TG 15.1 Screening

Screening is a form of investigative technology used (in isolation or combination) for early detection of a target condition that may benefit from early intervention. There are different types of screening, with different aims. Screening tests should not be confused with diagnostic tests, which are investigative technologies (in isolation or in combination) that tend to be applied to individuals to elucidate information that explains and/or assists in managing their current clinical presentation.

Screening is similar to predisposition testing (see below), but aims to detect preclinical signs of disease (e.g. mammograms for detecting breast cancer or faecal occult blood tests for detecting colorectal cancer).

Universal or population screening

Universal screening involves testing all people from the population who meet certain criteria (e.g. through programs such as newborn bloodspot screening, BreastScreen Australia, the National Bowel Cancer Screening Program and the National Cervical Screening Program). Delivery of preventive services (such as screening) is predominantly under the remit of the states and territories, although the Commonwealth and states/territories may sometimes share responsibility.

MSAC may occasionally be requested to assess universal screening tests (examples of prior assessments are for neonatal hearing screening and digital mammography).

Universal screening programs are considered to be associated with a high financial risk. MSAC therefore has a clear preference for direct from test to health outcomes evidence (Technical Guidance 10), and for claims based on health outcomes rather than intermediate outcomes or value of knowing that may be uncertain or difficult to quantify.

Guidance for assessments of population-based screening programs is similar to that for other investigative technologies. However, the following key considerations carry greater weight for population screening:

- The test population must be clearly defined (it is not often the whole population). The test population should include both those eligible for the test, and those who are likely to accept the test.
- Estimates of the prevalence of the disease or biomarker in the whole population and in the test population are important. Prevalence is used to estimate the likely yield, and numbers of false positives and false negatives.
- Where the prevalence is very low, and the test population is large, false-positive results often outnumber true-positive results. Confirmatory diagnostic tests are used to increase the accuracy of screening. It is important to accurately report the effects on the individual of receiving a false-positive test result, and the additional resources required to validate results.
- Health outcomes related to screening should account for both lead-time bias and overdiagnosis.[41]

Targeted screening

Targeted screening is testing of asymptomatic people who are at high risk of a given clinical condition or disease. Screening may be targeted so that the harms associated with the screening test (e.g. radiation exposure and overdiagnosis) are outweighed by the benefits of earlier detection of the condition or disease. The people screened may be considered high risk due to personal characteristics (e.g. age, gender, history of known medical risk factors), family history and/or specific exposures (e.g. workers in lead battery factories). For targeted screening, the health effect of the condition or disease prevented may be more minor than considered for universal screening (e.g. nausea or vomiting). Targeted screening may be legally required (e.g. miners who work with lead or chromium) and used in follow-up to environmental health incidents.

Targeted screening of genetic conditions is usually restricted to circumstances where there is a ≥10% pre-test risk of having the disease or the biomarker.

Targeted screening may be assessed using either the linked evidence approach or direct from test to health outcomes evidence.

Predisposition testing

Predisposition testing provides information on the likelihood of an asymptomatic person developing disease in the future. Currently, predisposition testing is most commonly associated with heritable diseases. An example of predisposition testing is the testing of women with characteristics (including family history) that render them at high risk of carrying a breast cancer–causing gene. A special example of predisposition testing is cascade testing (discussed below) of family members of someone with a known pathogenic variant. Cascade testing is regarded as a special case of

predisposition testing, as the pathogenic variant has been identified in the index case and therefore the cascade testing is based on a known variant.

At the time of writing, most predisposition tests assessed by MSAC have been for specific conditions. However, advances in genetic testing, in particular the use of gene panels, have meant it is possible to test people for a number of conditions simultaneously. Assessment of these panel tests requires use of an exemplar/facilitated assessment (see TG 5.2).

If the data are presented without a time-to-event element (i.e. a comparison of test results and clinical outcome with time not specified, or at one time point), present the positive predictive value and negative predictive value of the test (as described in Appendix 7).

When evaluating a predisposition test, it is important to make sure that the evidence is derived from a population with the appropriate prevalence of disease, or that the applicability of evidence from another population is considered. The positive predictive value of the test (the probability that a person with a positive test result will develop disease) depends on the prevalence of the disease in the population and the penetrance of the biomarker. For example, data derived from universal screening will not be applicable to targeted screening, as the proportion of false positives to true positives may vary widely. For a more in-depth discussion on the influence of prevalence on the positive and negative predictive values, see Appendix 7.

If data on the accuracy of the test for determining the biomarker are identified, without consideration of how well it predicts disease, a separate step of considering the penetrance of the biomarker/pathogenic variant will be required.

When data are presented in a time-to-event format, Kaplan–Meier curves and hazard ratios should be presented.

For many conditions, the appropriate comparator to the predictive test will be existing risk assessment approaches. For example, if an application for a new predisposition test for cardiovascular disease is proposed (such as testing for a polymorphism on chromosome 9p21), this would need to be compared against the information derived from existing risk-stratification approaches based on known risk factors for cardiovascular disease such as age, sex, smoking, hypertension, diabetes and lipid levels.

Cascade testing

When a patient is found to have a heritable genetic variant (often called the index patient or proband), related family members are at risk of having the same variant. Cascade testing is usually used to identify one specific genetic variant in high-risk first-degree relatives, or first- and second-degree relatives.

The assessment of a genetic test for identifying heritable variants must include at least 2 PICO sets. The first will describe the target at-risk population, and the second will describe the cascade testing population. Additional considerations and information requests are required for the assessment of a cascade test.

PICO: Present a PICO set (Technical Guidance 2) that describes the characteristics of the cascade population and the proposed cascade test. Describe the clinical management with and without the cascade test. Explain the usual relationship between the cascade population and the proband (e.g. are they more likely parents, siblings or children of the proband), and explain whether cascade testing would occur in only first-degree relatives, or also in second-degree relatives.

Defining the eligible cascade population: The yield of the genetic test in affected individuals helps define the size of the eligible cascade population. Further considerations that define the size of the cascade population are the age of the proband and the average number of relatives at risk of carrying the genetic variant. Describe the size of the population when presenting the PICO. Estimates of the size of the population and likely yields are relevant to the economic analysis and financial impact of a test.

Estimating uptake: Not all family members at risk of carrying a genetic variant will be offered or accept cascade testing. Provide an estimate of the uptake rate derived from relevant data sources. If the uptake rate of cascade testing is unknown for the target disease, use evidence from a similar disease and discuss applicability. The uptake rate of cascade testing may be reported (and justified) alongside evidence of change in management (Technical Guidance 12). Uptake rates are applied in the economic evaluation (Section 3) and/or in estimations of use in practice (Section 4). Barriers to the uptake of cascade testing may be informative to present alongside other relevant considerations (Section 5).

Test accuracy: As the genetic variant is known at the time of cascade testing, the accuracy of the cascade test is high, and accuracy studies are uncommon for cascade testing. Present the accuracy of cascade testing, or discuss the basis for assuming a particular accuracy if there is insufficient evidence.

Yield: The proportion of family members who are identified in cascade testing as carrying the genetic variant will typically follow Mendelian inheritance patterns. However, cascade testing yields may vary depending on whether uptake is selective (e.g. if family members at higher or lower risk are more likely to accept testing). Estimates of yield are commonly available for cascade testing and should be reported, if identified.

Disease inheritance and penetrance: The likelihood of family members carrying the genetic variant who then experience disease (and the time taken before the disease would reach a clinical diagnostic threshold) is influenced by disease inheritance patterns and penetrance. Autosomal dominant, autosomal recessive and X-linked genetic variants will affect the likelihood that disease will develop in family members. In addition, not all carriers of an autosomal dominant variant or all carriers of homozygous recessive variants will develop disease if the development of disease is influenced by additional genetic or environmental factors. The penetrance of the disease in family members with the genetic diagnosis is a measure of the longitudinal test accuracy of the cascade test. Provide estimates of penetrance when discussing longitudinal test accuracy, and use these estimates to inform the downstream implications of cascade testing in the economic evaluation.

Additional value of cascade testing: There may be other impacts of cascade testing that can be quantified or discussed. Cascade testing may release individuals who are at risk of developing disease from ongoing clinical monitoring if they are found to not be carriers of the pathological variant identified in the proband. It may be less acceptable to release family members from monitoring if the evidence for the clinical validity of the variant is poor (and therefore the variant may not be the cause of the disease), or if the disease in the proband is a result of both the pathological variant and multiple unidentified pathological variants. Discuss the impact of cascade testing on the monitoring requirements of the cascade population when presenting change in management (Technical Guidance 12). In addition, the identification of a pathological variant in a family member may affect planning for the future, result in changes to individual behaviour, help people connect with support networks and influence family planning decisions. These may be discussed alongside other relevant considerations (Technical Guidance 29). If a

specific claim relating to additional value (value of knowing) has been made in response to TG 1.4, quantify the additional value according to the requests in Technical Guidance 28.

Partner testing: Diseases caused by recessive variants require that both parents carry the recessive gene. If genetic testing or cascade testing identifies a family member as a carrier, there may be benefit associated with testing a partner to assist with family planning. The value of partner testing will be related to the prevalence of the pathogenic variant and the severity of the disease. Present an estimate of the yield of partner testing (which should be identical to the yield in the general population). Presenting health outcomes associated with partner testing is difficult and should be discussed with care (refer to Ethical analysis in TG 29.2).

Economic evaluation: An estimate of the cost-effectiveness of testing for a heritable variant would provide an incremental cost-effectiveness ratio (ICER) for testing the affected individual population, and a second ICER that incorporates the cascade testing population. Uncertainty in penetrance, uptake rates and effects on management of the cascade testing population should be explored in sensitivity analyses. It is also informative to present scenario analyses exploring additional cascade testing options (such as expanding the cascade population to include second-degree relatives, or partner testing).

TG 15.2 Testing to determine prognosis

A prognostic test provides information about a patient's prognosis, without specific consideration of downstream therapies chosen (i.e. it characterises the natural history of the disease). A prognostic test that does not result in a change in management does not have health outcomes benefits, but may have value of knowing.

Some prognostic tests may be used to inform a change in management. An example of this might be a test that predicts the likelihood of recurrence without adjuvant therapy. The test may then influence the benefit/risk assessment of using adjuvant therapy.

Many patient characteristics may provide useful information for determining their prognosis. Prognostic information may be considered of inherent value for the sake of the knowledge itself (see Technical Guidance 28 on value of knowing), as well as for its influence on the downstream health care that people receive. Prognostic tests are developed to assist (not replace) clinical judgement regarding the likely future health outcomes of the patient, and enhance patient decision-making.[42]

Many prognostic tests combine multiple variables to predict the risk of experiencing a specific endpoint within a specific time period. This formal combination of multiple factors is called a prognostic algorithm. Some prognostic algorithms may be considered 'black-box algorithms' (see Multifactorial algorithms; TG 15.5).

In order to show the relationship between the proposed test result and the endpoint, present the univariate analyses with the estimated effect (e.g. hazard ratio and survival probability) to demonstrate the prognostic strength. The same information should be provided for the comparator. For the effect of a marker on time-to-event outcomes, a Kaplan–Meier plot is recommended (Figure 21), showing the curve for each category.[30]

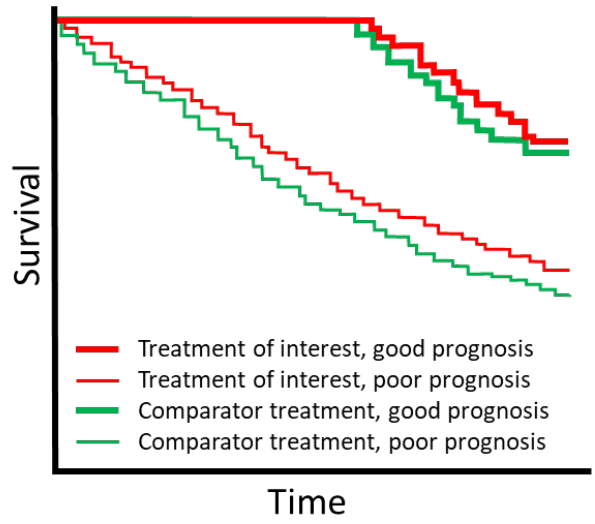


Figure 21 **Kaplan–Meier curve showing time-to-event outcomes for patients with good versus poor prognosis for the treatment of interest and the comparator**

To estimate the incremental value of the prognostic marker, multivariate analyses are required; these analyses demonstrate the additional information gained by the proposed test over and above existing risk stratification methods. It may be reasonable to report the categorisation of the same patients by the proposed test and the current stratification methods so that differences in categorisation can be determined. A comparison of the time-to-event data for the discordant cases would be informative.

If using measures such as odds ratios or relative risks, consider whether they are adequate for discriminating between people who are likely to develop the outcome of interest, and those who do not.[43] For example, if the absolute risk of developing a disease of interest is only 3 in 1,000 people, a relative risk of 3.0 (considered large in epidemiological research) would only mean that people with the particular marker had a 9 in 1,000 risk of developing the disease. Information on the absolute risk of the disease should be provided. Relative measures should be used with caution as a means of classifying individuals by risk.[43]

Where the results of the prognostic test are likely to affect clinical management, the benefit of the test may include health outcomes. The key comparison of interest is the proportion of patients who would change management following the test, and the incremental health benefits associated with this change in management. No health changes are considered for those patients who do not change management; however, there may be value of knowing considerations associated with the knowledge of prognosis.

TG 15.3 Testing to predict treatment effect

A predictive test provides information on the expected effect of a therapeutic technology (e.g. a test for the *HER2* gene to predict response to breast cancer treatment). This may result in 'personalised medicine', allowing the therapeutic technology to be restricted to those who are most likely to benefit, and avoiding the harms associated with the intervention in those unlikely to benefit. If the predictive test is codependent with a drug being submitted to the Pharmaceutical Benefits Advisory Committee (PBAC), see Appendix 8 on using a codependent technology assessment approach.

As with any investigative technology, the utility of a predictive test is best proven through the use of direct from test to health outcomes evidence, comparing health outcomes of patients whose management is guided by the predictive test with outcomes for those whose management is guided

by the comparative test strategy (which could include treatment without a test) (see Technical Guidance 10). However, the evidence on predictive tests is rarely generated in this manner.

Assessing a predictive test

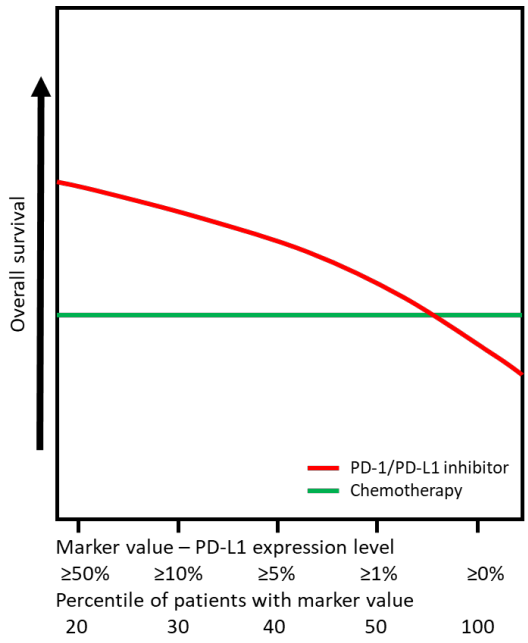
Describe the study design. Include whether testing was prospective or retrospective, and whether both biomarker-positive and biomarker-negative patients received both the targeted therapeutic intervention and the control intervention.

Explain how the study information is used to establish that the test is predictive. Discuss whether the results of the study could be affected by differences in prognosis by biomarker status. This is of particular concern when only biomarker-positive patients are included in the study.

If outcomes for test-negative patients are not captured in the study, explain how they are included in the assessment (or what assumptions have been made).

A critical concept within predictive testing is the 'clinical utility standard'. The clinical utility standard is the test, and the testing circumstances (methods of performing and interpreting the tests), that were used in a study reporting the health outcomes following the use of a treatment that has been targeted by a test. A comparison of the proposed test (to be used in Australia) with the clinical utility standard is required. This comparison would be performed similarly to a cross-sectional accuracy comparison, with a clear discussion of which test results are discordant and the implications of the discordance on the downstream treatment effects (see TG 11.4). It is important to present a 2-by-2 table (not simply a pooled measure of test agreement).

For continuous markers (for which statistical methods are limited), it is suggested that marker-by-treatment curves are presented, if available, to illustrate how particular test results correspond to different health outcomes, depending on which treatment is chosen. See Janes et al. (2011) for details.[44] This approach may be informative for determining a clinically meaningful threshold for test positivity (Figure 22).



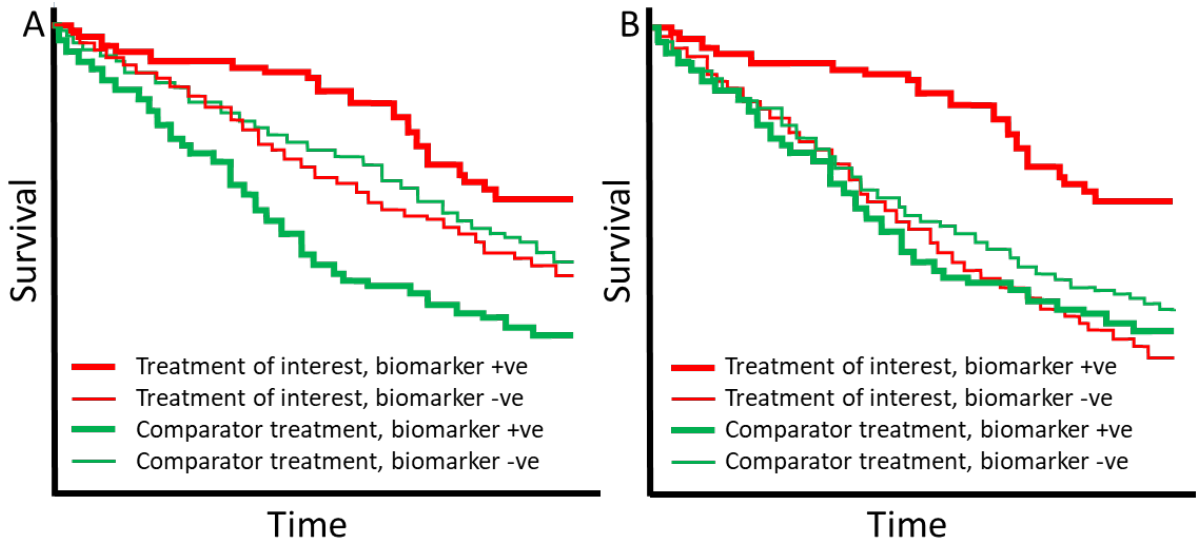
PD-L1 = programmed cell death-ligand 1, PD-1 = programmed cell death protein

Patients with non-small cell lung cancer (NSCLC) who have higher levels of PD-L1 expression are likely to have a greater benefit from treatment with PD-1/PD-L1 inhibitors than from chemotherapy.

Figure 22 A marker-by-treatment curve showing response to second-line PD-1/PD-L1 inhibitors and chemotherapy in patients with NSCLC according to PD-L1 expression level

To determine whether there is a treatment effect modifier, it must be determined whether response to treatment (versus control) varies by the test result (i.e. will the test select a subgroup who respond to treatment differently to those who are test negative).

Distinguishing whether a biomarker determines prognosis or is a treatment effect modifier may not be possible unless a study provides health outcomes for those who are biomarker positive and negative, and those who have the treatment versus control (i.e. all four arms) (Figure 23). If a study does not contain the required comparisons, additional evidence may be required to differentiate between prognostic and predictive effects of the test.



(A) The biomarker is both predictive of response to the treatment of interest and prognostic. (B) The biomarker is not prognostic but is predictive of response to the treatment of interest.

Figure 23 Kaplan–Meier curve showing time-to-event outcomes for patients with and without a biomarker

When determining whether there is an interaction between treatment effect and subgroup, results are often presented separately for subgroups, with the erroneous conclusion that a significant treatment effect in one subgroup, and a non-significant treatment effect in the other, indicates that the effect differs by subgroup.[30] Ideally, a test of interaction should be performed to rigorously assess whether the effects are different by subgroup, or whether the difference in significance level is due to one subgroup being underpowered (too small to allow detection of a significant difference). The magnitude of the interaction does not describe how useful a marker is for patients.[44] Therefore, if the test of interaction is significant, further evaluation may be required to determine the nature of the interaction, that is, whether the effects are in the opposite direction, or in the same direction but of a different magnitude.[30]

For continuous variables, categorisation is a common approach for assessing interactions, but it is highly dependent on the thresholds used for the different categories. Altman et al. (2012) suggest that the multivariable fractional polynomial interaction approach, which avoids specifying the thresholds, is preferred.[30] A checklist allows interaction between a binary and a continuous variable to be investigated, with or without adjustments for other variables.[30] Alternatively, the subpopulation treatment effect pattern plot, as described by Bonetti et al.[45] may be used.[30]

The interpretation of the evidence depends largely on whether the subgroup analyses are performed on subgroups determined a priori, or whether they are conducted post hoc.[30] MSAC's strong preference is for subgroups determined a priori based on prespecified classifications, to reduce the likelihood that the finding occurred by chance. If the analyses are performed post hoc, they are considered to be hypothesis generating, and therefore require validation using an independent sample.

TG 15.4 Monitoring

Some investigative technologies are intended to be used as part of a disease-monitoring strategy. Monitoring can be summarised as consisting of 5 phases:[46, 47]

1. pre-treatment monitoring (surveillance): to screen individuals for the need to start treatment

2. initial response monitoring: to determine whether the individual's response to treatment is as expected from the mean response observed in trials
3. on-treatment long-term monitoring: to assess whether treatment remains adequate over the long term
4. after a significant change in the disease process or treatment has occurred
5. determining whether it is possible to stop treatment.

The assessment of a test used for monitoring is similar to tests for other uses. This includes the need to investigate the clinical utility of the test (direct from test to health outcomes evidence), how the monitoring test influences management, the impact of the change in management on patient health outcomes, and test accuracy. Increased monitoring may lead to increased anxiety for the individual being tested, or could conversely increase their feelings of empowerment. These outcomes should be addressed either under direct from test to health outcomes evidence, or in a discussion on the value of knowing.

Additional aspects of test accuracy (in addition to diagnostic or predictive accuracy) that are relevant to address and which should be discussed for monitoring tests are:[46]

1. responsiveness: how much the test result changes in response to an intervention/treatment change relative to background random variation (signal-to-noise ratio)
2. detectability of long-term change: the size of the change in the test result over the long term relative to background random variation, thereby informing the logical frequency of monitoring.

Further assessment should also be performed on:

3. practicality: the ease of use and interpretation of the test, cost, and level of invasiveness.

The 'responsiveness' criterion is especially important for the initial response phase of monitoring soon after a new treatment has been started. Although less obvious, this criterion is also important for both pre-treatment and long-term monitoring. For all monitoring phases, ideally the test result should be responsive to treatments that alter the patient's risk of the clinical outcome. Such interventions may be lifestyle changes in the pre-treatment phase, pharmacologic treatments in the initial response phase, or measures to improve adherence in the long-term monitoring phase.

Related to the concept of responsiveness is the speed of change in response to an intervention. Preferably a test should be able to detect a rapid response to treatment. This is obviously a necessity when the change in outcome in response to the intervention is also rapid, for example, risk of hypoglycaemia for glucose-lowering medicines (monitor glucose) or bleeding risk for patients on warfarin (monitor international normalised ratio). In other situations in which the change in outcome is much slower, it is still preferred that the test response can be quickly judged as to whether treatment is working as expected, for example, risk of a cardiovascular event (monitor cholesterol and blood pressure). Not all responsive tests detect rapid changes in response to treatment; in fact, some parameters take months or years to change, for example, changes to HbA1c or bone mineral density. Because changes in the results of the tests reflect average treatment effects over a longer period of time, these services may be preferred for judging effects over the medium to long term.

If initial response monitoring is considered to be of value, the frequency of monitoring should be considered and justified.

Once a patient is stable on treatment, the frequency of monitoring can be decreased. The frequency of monitoring would depend on the within-person variability of the monitoring test, and the range of rates of long-term change (e.g. the rate of progression or regression of the disease).

The concepts of 'signal' and 'noise' are relevant to both response monitoring and long-term monitoring. For response monitoring, signal includes both the population mean change and between-person variation in response. If the between-person variation component of the signal is small, the signal for an individual can be estimated using the population mean change without needing to monitor. If the between-person variation is not small, it is difficult to estimate signal on the basis of population mean change alone. The individual's true deviation from the mean change also needs to be estimated and this is best done where there is a favourable signal-to-noise ratio. Noise is a result of background random variation within individuals because of measurement error and biological fluctuations. The amount of noise in investigative technologies used for monitoring may not be appreciated by clinicians, and variations due to noise may be wrongly attributed to real change.

A study by Bell et al. (2008) addresses the assessment of initial response monitoring, and when it is worth monitoring initial response to treatment.[48] Treatment for patients with chronic conditions is often monitored by using surrogate outcomes (such as blood pressure or cholesterol). A surrogate outcome should only be considered for monitoring if it is known to predict the treatment effect on risk of the clinical outcome. Monitoring initial response to treatment should be avoided unless it is expected to be useful in informing clinical decision-making. Monitoring is unlikely to be of value when there is no evidence of variation in the response to treatment between patients/subgroups, or when it is highly likely that therapeutic targets will be met. Variability in treatment effects between individuals can be estimated from placebo controlled randomised trials.

'Detectability of long-term change' describes the ability of the test to discern true long-term changes in the patient's condition (signal) from short-term measurement variability (noise). The signal for long-term change monitoring is the true long-term trend in level within an individual over time. This is a combination of the population mean change and the between-person variability or individual deviations from the mean change. As for response monitoring, noise is the short-term random variation in level within an individual. Unlike response monitoring, where the often small between-person variation component of the signal renders monitoring unnecessary, in long-term monitoring, there is usually substantial between-person variation in the long-term trends. This means that it is difficult to estimate a signal on the basis of population mean change alone, and there is a need to also estimate the individual's true deviation from the mean change under conditions of a favourable signal-to-noise ratio. The frequency of long-term monitoring should be determined based on the rate of 'drift', and the closeness to the target or threshold value.[47, 49] The closer the measurement is to the threshold, the sooner a repeat measurement is needed.

Finally, the 'practicality' of the test as a monitoring tool describes its ease of use, level of invasiveness and cost. Although every assessment should consider the practicality of the proposed technology, monitoring tests are more likely to be repeated than other forms of test, and may involve a component performed by the patient or family member. Ease of use and invasiveness are therefore likely to have a higher impact on compliance than for a once-off test such as used for diagnosis.

TG 15.5 Multifactorial algorithms

A multifactorial or multicomponent algorithm is part of an investigative technology, and can be diagnostic, prognostic or predictive. Algorithms can be developed from rule-based prediction models or using adaptive self-learning approaches. Algorithms may be fixed or static (they are trained to the

point at which they are validated, and then cease to develop further), or they may be dynamic and continue to evolve as they are exposed to additional information. The assessment of an investigative technology containing a multifactorial algorithm follows a similar approach to the assessment of a typical investigative technology (see Section 2B), but has some additional considerations that need to be addressed:

- Discuss the biological plausibility of the algorithm. Include information to support the link between the variables used in the algorithm (clinical or genetic characteristics) and the purpose of the algorithm (diagnosis of a disease, prognosis of a disease, or prediction of response to treatment).
- Describe the output of the algorithm. Algorithms tend to report continuous or discrete scores, but their interpretation may require a threshold. Explain how the score is used, and define the threshold for different actions.
- Describe the precedent steps for using the algorithm (to generate the raw data for the algorithm). These may include a gene panel or a radiograph. Explain whether these steps are standardised across all providers of the service and the risk to the accuracy of the algorithm should the precedent platforms differ. Describe the quality assurance approaches to ensure standardisation of the data generation.

Fixed algorithms

Describe the process of developing the algorithm. Key information includes the number of variables initially examined and the number of variables ultimately included in the algorithm (e.g. 250 genes were initially tested, and 30 genes were selected for the final panel). Explain the criteria for inclusion of variables (the threshold for inclusion). If available, provide the weight assigned to each of the variables in the algorithm, and the way that the variables are combined.

Static or fixed algorithms are developed in 2 steps: a training or discovery step (where the algorithm is developed) and a validation step (where the performance of the algorithm is confirmed). Present details of any iterations of the algorithm. Confirm that the training dataset is exclusive of the validation dataset. The performance of the algorithm cannot be quantified if the algorithm proposed for reimbursement or the algorithm assessed in the validation study differ from the algorithm defined by the training dataset. Confirm that the proposed algorithm is identical to the algorithm used in the validation study and developed in the training study. Changes to the algorithm will require additional evidence from subsequent validation studies.

The generalisability of the algorithm[50, 51] is a key concern that should be addressed in the assessment report. Present the characteristics of both the training and validation studies and compare these with the characteristics of the target population. The characteristics may include the variables used to train the algorithm, and other important predictors of disease, prognosis or treatment response. Identify characteristics that differ in the target population. Clearly identify any groups in the target population that are underrepresented in the training and validation cohorts. Where there are differences across populations, present evidence to support the generalisability of the performance of the algorithm.

Perform an assessment of the investigative algorithm using a standard approach. Follow the guidance in Section 2B to produce an assessment report based on direct from test to health outcomes evidence, or using a linked evidence approach.

Dynamic algorithms

Adaptive self-learning algorithms change as they are challenged with new data, and their predictions will also change.[52] The algorithms use machine learning or neural networks to optimise performance

over time. Like fixed algorithms, the performance of dynamic algorithms is tested in validation studies. Key concerns with dynamic algorithms relate to how the algorithm changes over time and the impacts this has on effectiveness and safety.

The evaluation of a dynamic algorithm follows the same approach as for a fixed algorithm. Present results from the training dataset(s) and validation set(s). Confirm that the validation dataset(s) are independent from the training dataset(s). In addition, identify changes in the algorithm predictions over time by presenting results for the same datasets across different 'iterations' of the algorithm. Describe the likely clinical implications of this change in prediction.

Present the characteristics of the datasets and compare these with the characteristics of the target population.

Explain the mechanism of self-learning. Include the source of the new data, how they are incorporated into the algorithm and how frequently the algorithm is revised.

Algorithms are used to 'predict' a diagnosis, prognosis or treatment effect; however, the goal of the self-learning may be to optimise the sensitivity or specificity (or a combined metric). Explain what the algorithm is programmed to optimise (and how it achieves this optimisation). Provide the results of the performance of the algorithm at different time points to demonstrate the nature of the optimisation trend, which should be trending towards the optimising asymptote over time.

Describe the quality assurance activities required to ensure that the performance of the algorithm does not decline as new data are introduced, to prevent adversarial attacks, and to prevent biases associated with under- or overrepresentation of subpopulations.

If action taken as a result of the algorithm prevents the predicted outcome, this outcome would not be observed in the post-implementation period and could bias results in ways that are difficult to ascertain.[53] Describe quality assurance activities that may address this concern.

Quality appraisal

Fazel and Wolf[54] developed a simple 10-point checklist covering key factors to consider when determining the quality of risk assessment algorithms. Some of these factors may relate more closely to static or fixed algorithms (such as knowledge of the included variables and the weighting of the variables). When applying the checklist to dynamic or self-learning algorithms, amend the included factors as appropriate. The key factors to consider are:

- What study type was used (external validation study in a new sample, or a derivation and internal validation study)?
- Was the study based on either a protocol or prespecified analyses?
- Did the algorithm have a set objective and clearly defined outcomes?
- Did the study report the discrimination and calibration measures used in the algorithm?
- Were the variables included in the algorithm the same for both the validation and derivation studies?
- Was the weighting of the variables the same for both the validation and derivation studies?
- Was the study population the same as the target population?
- Was the sample size sufficient?
- Did the study report predefined output categories or thresholds?

- Was the study published in a peer-reviewed journal? Note that this checkpoint item should not conflict with the preference for the best available evidence. In general, all evidence of the proposed health technology should be provided. The evidence would preferably include complete clinical study reports and protocols, and any additional supportive evidence that may be required.

TG 15.6 Codependent technologies

Health technologies are codependent when the patient health outcomes related to the use of a therapeutic health technology (e.g. a medicine) are improved by the use of another health technology (e.g. an investigative technology). The combined use of these technologies leads to their intended clinical effect, and therefore the benefits of both technologies should be assessed together (rather than assessing each technology in isolation).

Investigative technologies could be involved in codependent technologies for the purpose of:

- establishing a predisposition or estimating a prognosis
- identifying a patient as suitable for a therapeutic technology
- measuring an early treatment effect on a surrogate outcome as the basis for predicting more patient-relevant health outcomes
- monitoring a patient over time after an initial investigation to guide subsequent treatment decisions.

A codependent submission is required when the Minister for Health requires advice from 2 different expert advisory committees because listing of the codependent technologies involves 2 separate reimbursement schemes. For example, codependent technologies that require new listings or amendments to both the PBS and the MBS (such as a genetic test to determine eligibility for a medicine) would need advice from both PBAC and MSAC. Codependent submissions can be integrated (one combined submission for the 2 technologies that is considered jointly by both MSAC and PBAC) or streamlined (separate submissions for each of the technologies). Appendix 8 provides detailed guidance on the preparation of a submission that involves codependent technologies between MSAC and the PBAC.

The most common codependent technologies are a genetic test and a medicine. Other codependent pairs include an MBS service (assessed by MSAC) in combination with an implant (which would be listed on the Prostheses List); an imaging or blood test to determine eligibility for a therapeutic service or medicine; a medicine delivery system in combination with a medicine; or a monitoring test to determine specific therapeutic management or medicine dose.

When an application for a codependent technology is purely for a therapeutic purpose (e.g. the combination of an MBS service with an implant), the 2 interventions would most likely be assessed together as one therapeutic intervention.

Technical Guidance 16 Interpretation of the investigative evidence

KEY CONSIDERATIONS

For investigative technologies

- **Summarising the evidence base for a test must account for the quality and strength of the evidence for each individual component, as well as an overall assessment of the effect of the proposed test on health (or other) outcomes (TG 16.1).**
- **The conclusion regarding the clinical utility of the investigative technology should be a simple and unequivocal statement about the superiority, inferiority or noninferiority of the test to the comparator and/or reference standard that is supported by evidence provided in the submission (TG 16.2).**

The objective of summarising the overall evidence base in the assessment report is to describe the results as they apply to the clinical claim in the specific context of the Australian setting. Follow the guidance provided in TG 8.1. In addition, the interpretation of investigative evidence requires an overview of how the evidence has been constructed, the effects on subgroups defined by testing, and an identification of areas of uncertainty, particularly when evidence has been linked. This is described in TG 16.1.

TG 16.1 Investigative evidence interpretation

Summarising the evidence base for a test must account for the quality and strength of the evidence for each individual component, as well as an overall assessment of the effect of the proposed test on health (or other) outcomes.

Consider the following structure for the summary of the evidence base:

- a description of the evidence approach that was taken. Describe the direct or linked approach, and any supplementary evidence that was required
- a summary of each evidentiary step to complete the link between test and outcomes and health
- an overall interpretation of the comparative impact of the proposed test on health outcomes (and value of knowing outcomes, if required), with reference to the key uncertainties identified for each evidentiary step
  - This interpretation should discuss the implications of evidence and uncertainty at each step on the results of subsequent steps in a linked approach; for example, the impact of a test with poor specificity on treatment outcomes, or the impact of uncertainty in change in management on treatment outcomes.
  - When a linked evidence approach to evaluating a test has been used, it is critical for the assessment to link the evidence together, to reach a conclusion regarding the impact of the test on outcomes compared with the comparator. One way to summarise the information is to start with a hypothetical population who undergoes testing, and follow the population through to the range of different outcomes, incorporating change in management, the proportion treated appropriately versus inappropriately (based on accuracy of test results), and the patient-relevant health outcomes. This creates a comparative effectiveness model, which can then be expanded on to form the basis of a cost-effectiveness model in the economic evaluation. The model, or narrative synthesis, of clinical utility should capture the trade-off inherent in testing and subsequent decisions. It should also identify crucial areas of uncertainty in the existing data where more primary data collection is required.[22]

- a summary of the health impacts on patients by test status (positive, negative and inconclusive) including those who are misclassified (false positives and false negatives)
- a table of the key uncertainties for each evidentiary step, and an overall assessment of the quality and certainty of the evidence (see Table 12). While this table is a compilation of the summaries described above, it is reasonable to present the table in the assessment report before the longer summaries.

Table 12 Summary table of key uncertainties in investigative evidence

Evidence component of the assessment	Interpretation and key uncertainties
Test accuracy	
Change in management	
Health outcomes	
Safety of the test	
Safety of the treatment	
Overall assessment of the evidence	

Provide a summary of the overall evidence base (without repeating evidence from other sections). Consider:

- the level of the evidence, taking account of the directness of the comparison
- the quality of the evidence
- the clinical importance and patient relevance of the effectiveness and safety outcomes
- the statistical precision of the evidence
- the size of the effect
- the consistency of the results across the clinical studies and across subgroups
- the strength or certainty of the evidence
- the applicability of the evidence to the Australian setting
- any other uncertainties in the evidence, including missing outcomes or populations
- other relevant considerations that may influence decision-making, particularly implementation and ethical factors.

TG 16.2 Conclusion of clinical utility

The interpretation of the clinical data presented in Section 2 is crucial in determining the success of the submission. It is important to classify the health outcomes of the proposed health technology in relation to its main comparator, that is, whether it is superior, inferior or noninferior to the comparator.

The conclusion of the clinical utility of the investigative technology should be a simple and unequivocal statement that is supported by evidence in the submission.

Example:

The use of [proposed health technology] results in superior/noninferior/inferior effectiveness compared with [comparator].

The use of [proposed health technology] results in superior/noninferior/inferior safety compared with [comparator].

Section 3 Economic evaluation

Introduction

An economic evaluation of substituting the proposed health technology for the main comparator in the context of the requested listing should be presented in Section 3 of the assessment report. This should include a full and transparent description of the economic evaluation (the base-case analysis), with sensitivity analyses to characterise the uncertainty around the results. The base-case analysis represents the most plausible or preferred set of inputs and assumptions, and is guided by the MSAC reference case.

Flowchart 5 shows the structure of the guidance provided in this section for the economic evaluation.

Separate guidance is provided for a cost-effectiveness analysis (CEA) (including cost-utility analysis, CUA) (Section 3A), and for a cost-minimisation approach (Section 3B).

A cost-minimisation approach should only be used when the proposed service has been demonstrated to be noninferior to its main comparator(s) in terms of both effectiveness and safety. In this case, the difference between the service and the appropriate comparator can be reduced to a comparison of costs. However, because there may be uncertainty around such a conclusion, MSAC may subsequently request a cost-consequences analysis (CCA), CUA and/or CEA to be presented. Further, in some circumstances where a clinical claim of noninferiority is made in Section 2, other supportive factors may justify an increase in costs to the health system (see Technical Guidance 28 and Technical Guidance 29). Under such circumstances, a CEA and/or CUA could be presented to support the proposed increase in costs.

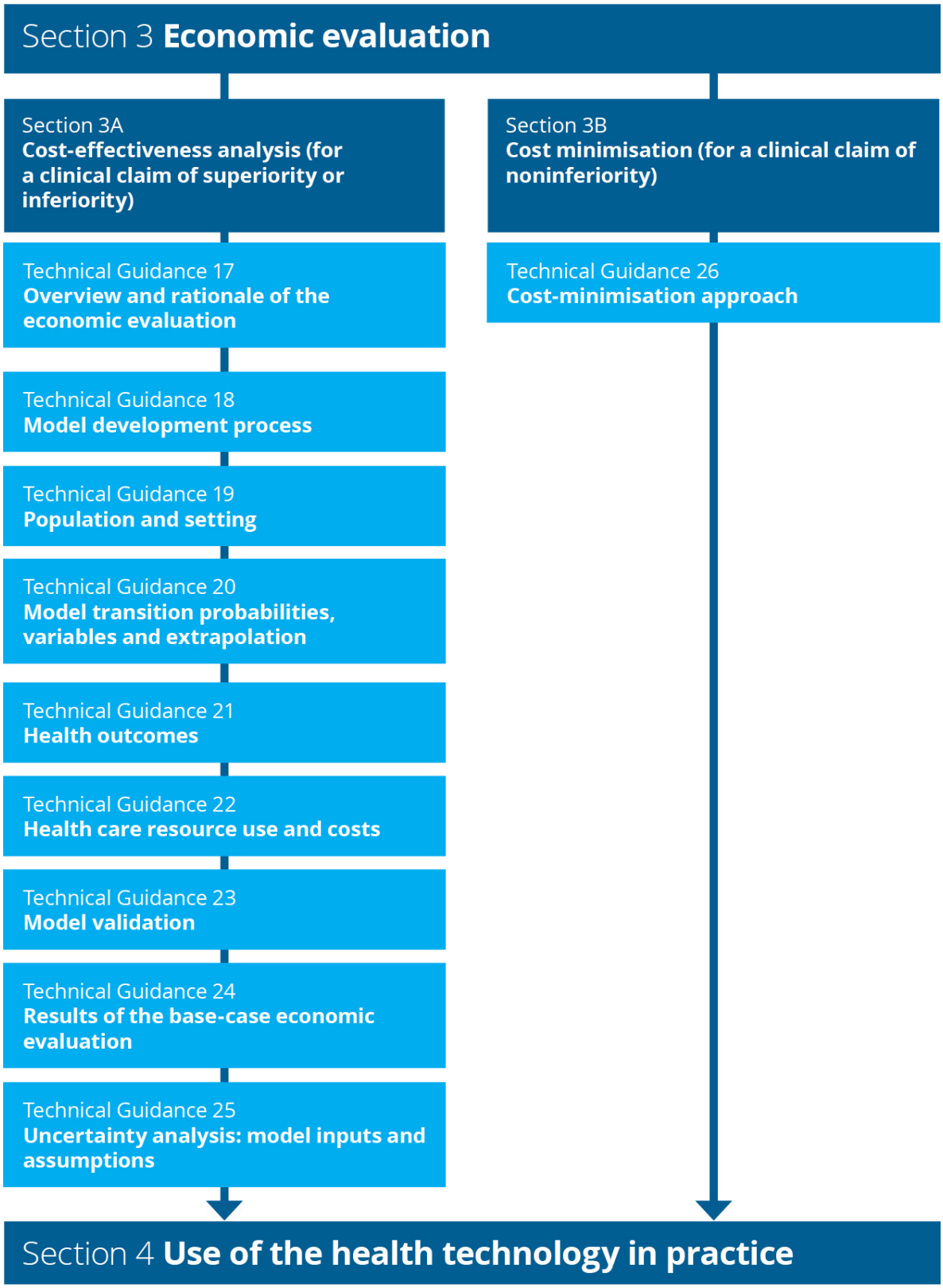
In circumstances where a high degree of uncertainty in the clinical assessment means that an economic evaluation would also be associated with a high degree of uncertainty (and therefore have limited usefulness to MSAC), this should be raised with the Department as soon as possible during the development of the assessment report. In these cases, the value of information generated by a model-based economic evaluation may diminish due to the quality of underlying data and greater uncertainties introduced through the process of modelling. Progression through modelling steps should continue until the results become uninformative for decision making. However, extending model steps when data are uncertain may provide MSAC with insight to the key evidentiary gaps and areas of uncertainty, and therefore may remain informative for decision making. Justify the truncation of analyses to modelling steps earlier than cost per quality-adjusted life years (QALYs) or other CEA outcome.

In the unusual circumstances where the proposed health technology is indisputably demonstrated to be therapeutically inferior, an economic evaluation may not be required as MSAC is unlikely to recommend government subsidy of the service. However, other supportive factors (see Technical Guidance 28 and Technical Guidance 29) may be present, and an economic evaluation may be useful if a therapeutically inferior service can be funded at an overall lower cost to the health care system.

The most useful presentation of results from the economic analysis might vary with the level of evidence available. For example, in some circumstances, the evidence base might be weak (e.g. if a claim of a service's safety and 'promising' effectiveness is based on low-level evidence, such that the claim cannot yet be considered proven). In such cases, a threshold analysis that examines incremental cost-effectiveness over a range of possible benefits might be more informative than the

incremental cost-effectiveness ratio (ICER) based on a single point estimate of incremental effectiveness.

The objective of the CEA should be to provide an unbiased, plausible estimate of the incremental cost-effectiveness of the proposed health technology. Where there is considerable uncertainty around an assumption or the value of a parameter, a relatively conservative approach should be used in the base-case analysis.

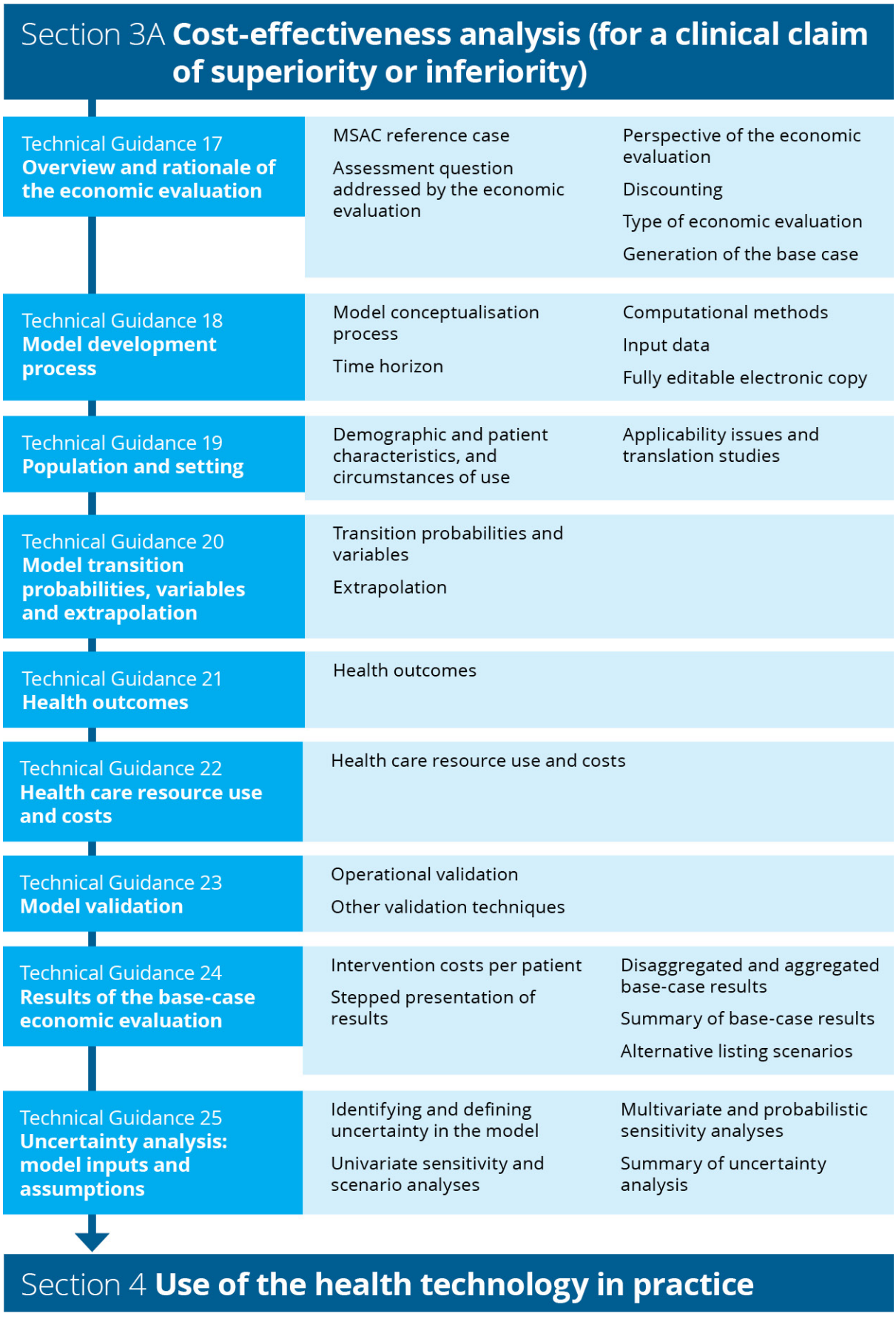


Flowchart 5 **Structure of the guidance for the economic evaluation**

Section 3A Cost-effectiveness analysis

Section 3A provides guidance for preparing a CEA (including CUA). Flowchart 6 shows the structure of the guidance in this section.

MSAC prefers that the economic evaluation is based on results from direct randomised trials, with adjustments or additions to the trial data as required to account for differences in the population and setting, timeframe of analysis or outcomes of interest. Adjustments should be presented transparently in a stepped manner. For economic evaluations that rely on results from indirect comparisons of randomised trials, comparisons based on nonrandomised studies or linked analyses, an adaptation of the stepped approach is recommended.



Flowchart 6 **Structure of the guidance for cost-effectiveness analysis**

Technical Guidance 17 Overview and rationale of the economic evaluation

KEY CONSIDERATIONS

- **A consistent approach to the economic evaluation assists in decision making. MSAC's preferred methods are included in the MSAC reference case (TG 17.1).**
- **Describe the assessment question addressed by the economic evaluation (TG 17.2).**
- **Report the perspective (TG 17.3) and the discount rate (TG 17.4) used in the economic evaluation, and the type of economic evaluation (TG 17.5).**
- **Explain the method for generating the base case (whether the analysis is modelled, and identify the steps required to generate the results) (TG 17.6).**

TG 17.1 MSAC reference case

A reference case (Table 13) has been defined that specifies the preferred methods for economic evaluations to be presented to MSAC. These have been specified to promote consistency across economic evaluations of different technologies and disease areas.

Table 13 MSAC reference case for economic evaluations

Component	Description	Relevant guidance
The assessment question	As defined in the PICO confirmation	Technical Guidance 17
Comparator	As defined in the PICO confirmation (the currently available service that is most likely to be replaced by the new service)	Technical Guidance 17
Perspective on outcomes	Personal health of person receiving the intervention	Technical Guidance 17
Perspective on costs	Health care system (health care costs incurred by the public or private health care provider; includes costs incurred by the patient)	Technical Guidance 17
Type(s) of analysis	Cost-utility analysis, or a cost-effectiveness analysis where a cost-utility analysis is not feasible	Technical Guidance 17
Time horizon	Sufficient to capture all important differences in costs and outcomes between the intervention and the comparator	Technical Guidance 18
Source of effectiveness inputs	Derived from the systematic review conducted in Section 2, translated as necessary	Technical Guidance 19 and Technical Guidance 20
Measuring and valuing health effects	QALYs. However, where transformation to QALYs is not feasible, the outcome measure should be that which most closely and validly estimates the final health outcome from a patient perspective.	Technical Guidance 21
Evidence on resource use and costs	Where available, use the source of costs recommended in the PBAC Manual of resource items and their associated unit costs.[a] However, for MBS-funded services, patient out-of-pocket costs, including average charges above the schedule fee, should be used where possible.	Technical Guidance 22

Component	Description	Relevant guidance
Discount rate	Annual rate of 5% for both costs and outcomes	Technical Guidance 17
Sensitivity analyses	Parameter uncertainty should be explored using deterministic (univariate and multivariate) analyses. Use scenario analyses to address translational and structural uncertainty.	Technical Guidance 25

MBS = Medicare Benefits Schedule; PBAC = Pharmaceutical Benefits Advisory Committee; PICO = population, intervention, comparator and outcomes; QALY = quality-adjusted life year

a www.pbs.gov.au/info/industry/useful-resources/manual

Where non-reference-case methods or analyses are relevant, MSAC prefers these to be presented as supplementary analyses. If non-reference-case methods are used in the base-case analysis, these should be clearly specified and justified.

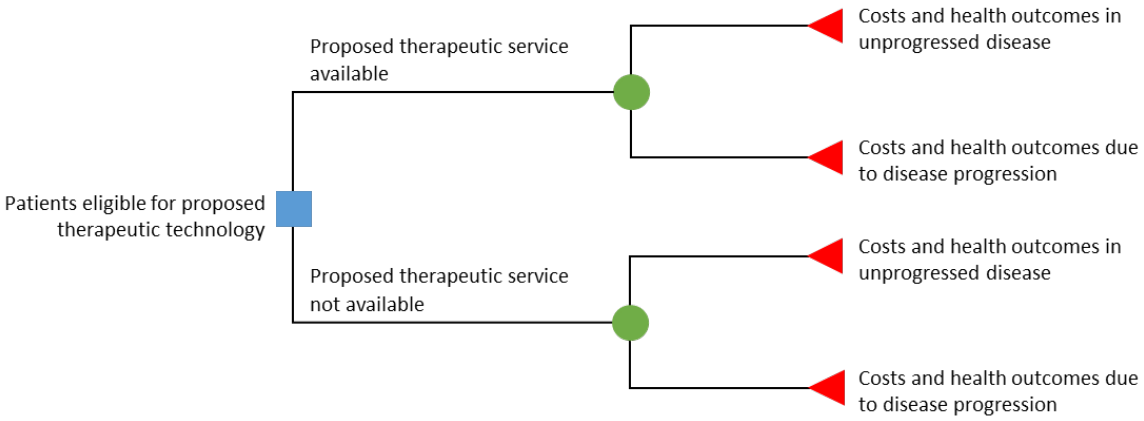
TG 17.2 The assessment question addressed by the economic evaluation

Present a clear statement of the assessment question that the economic evaluation aims to address. The statement should define the interventions being compared and the relevant patient group(s), and should be consistent with the population, intervention, comparator and outcomes (PICO) criteria in the PICO confirmation. Any differences from the PICO confirmation must be clearly presented and justified.

A decision-tree diagram may be presented which characterises the primary decision that the economic evaluation addresses, based on the information provided in response to Technical Guidance 2. Use this diagram to provide a conceptual overview rather than the complete computational structure of the economic model. After the decision point of the tree, define alternative choices, uncertain events and outcomes. For investigative technologies, include the diagnostic decisions and outcomes, where relevant. Where the model is particularly complex, collapse and summarise branches, and clearly indicate where this has been done. Detail collapsed branches or a more suitable complete diagram of the model structure (e.g. a health state transition diagram) in Section 3A.1.2 of the assessment report.

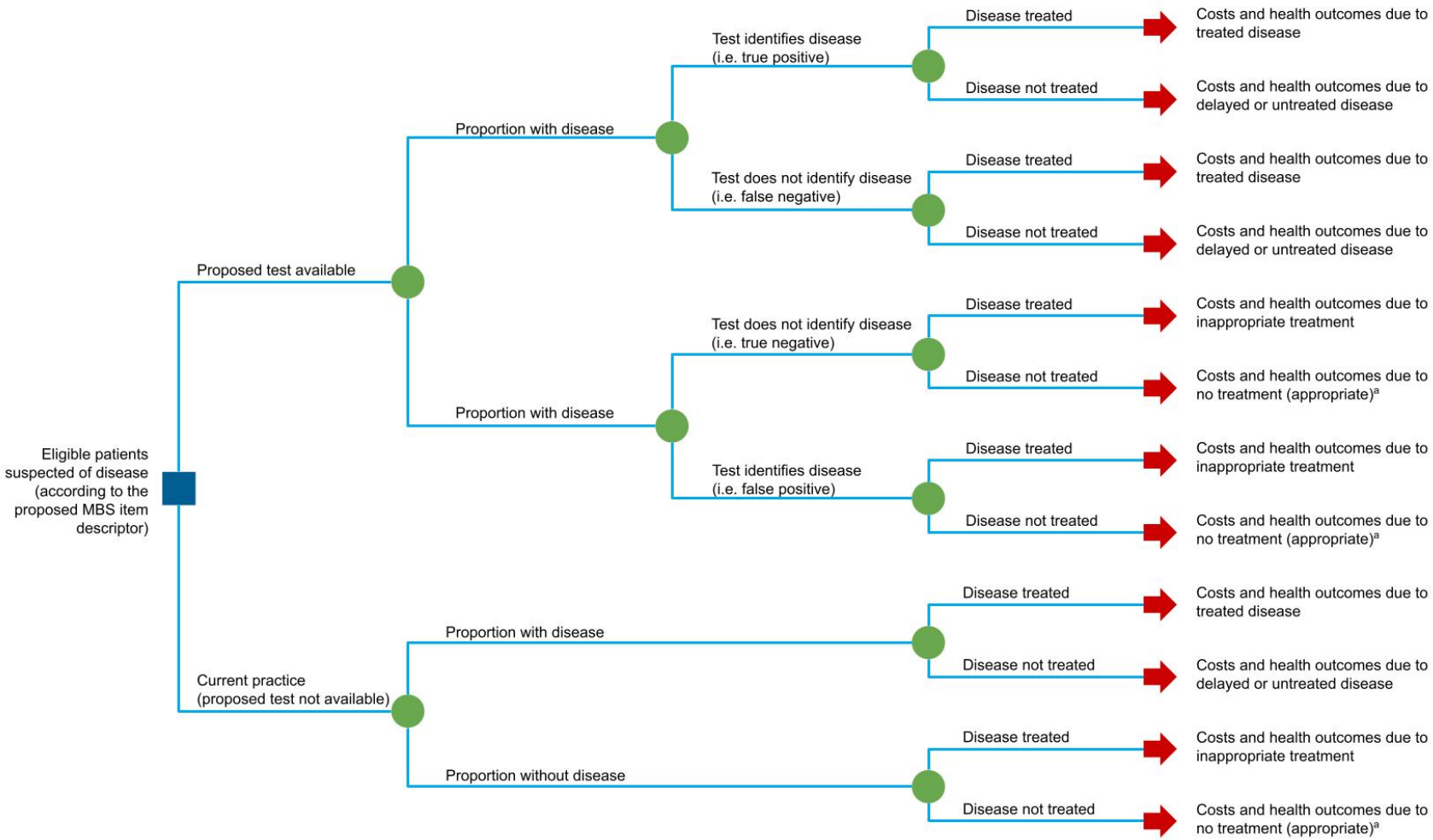
Ensure that the pathways depicted in the decision tree are consistent with the existing and proposed clinical management algorithms presented in Section 1 of the assessment report. Cross-reference to the clinical management algorithms if they sufficiently represent the decision analytic of the economic model.

Examples of decision-tree diagrams are presented in Figure 24 and Figure 25.



Note: While the conceptualised structure is the same across the arms of the economic model, a reduction in disease progression is expected with the proposed health technology, based on the results of the clinical assessment.

Figure 24 Decision-tree diagram conceptualising the assessment question of a therapeutic technology



a Where no treatment is appropriate, further investigations to diagnose underlying cause of disease could be included, but this might not necessarily be quantified.

Note: While the conceptualised structure of the assessment question is similar across the arms of the economic model, an increase in the proportion of appropriate treatment decisions is expected with the proposed health technology, based on the results of the clinical assessment.

Figure 25 Decision-tree diagram conceptualising the assessment question of an investigative technology

TG 17.3 Perspective of the economic evaluation

MSAC prefers a health care system perspective, which includes health and health-related resource use (costs and cost offsets), and health-related outcomes. Health care costs include those incurred by the patient, and the public or private health care provider; health outcomes are those associated with the patient. For investigative technologies, this includes both the benefits and harms directly related to the service (e.g. an adverse event due to exposure to an imaging contrast agent) and those indirectly related to the service (such as those that arise from subsequent changes in treatment). Do not include costs and outcomes that are not specifically related to 'health and/or provision of health care' in the base case (see Technical Guidance 21 and Technical Guidance 22).

Where a broader societal perspective is relevant, quantify and incorporate considerations beyond the patient and the health care system in a supplementary analysis. A well-justified and well-supported analysis will form a more compelling case (see Technical Guidance 21 and Technical Guidance 22 for guidance on identifying, measuring and valuing nonhealth outcomes and costs).

Supplementary analyses may be appropriate where the proposed intervention has important societal implications extending beyond the health outcomes of the patient receiving the proposed health technology, and beyond the health care system. Examples include costs/savings or socially relevant outcomes in domains such as education, housing or justice, or economic productivity impacts (see Appendix 10). Also, where the beneficiaries of health or other relevant outcomes are broader than the treated patient population (e.g. community, carers, dependants), these generally should be included as supplementary analyses. However, where important and relevant, the omission of these costs and outcomes from the base case should be drawn to the attention of MSAC.

TG 17.4 Discounting

The values of costs and benefits incurred or received in the future are generally discounted to reflect the present value. Discount both costs and outcomes at a uniform, annual (compounding) rate of 5% per year for all costs and health outcomes that occur or extend beyond one year in the base case.

Present sensitivity analyses using fixed discount rates of 3.5% and 0% per year (applied to both costs and outcomes). If relevant, present supplementary analyses using other discounting methodologies (e.g. a different uniform rate, differential rates, time-varying rates) and justify the alternative approach.

TG 17.5 Type of economic evaluation

State whether a CUA and/or CEA will be used. Identify the incremental health costs and incremental health outcomes (QALYs for a CUA, or as nominated for the CEA). If no single outcome measure can be presented that appropriately captures the overall health of the patient, or when the evaluation of the wider benefits of a technology is more useful, it may be reasonable to present a CCA in the base case.

The various types of economic evaluations are not mutually exclusive, and more than one analysis can be presented to make a stronger case for cost-effectiveness (e.g. both a CUA and a CEA, or CUA and CCA). (See the HTA advisory committee glossary[a] for definitions.)

A cost-benefit analysis should not be presented in the base-case analysis.

a www.pbs.gov.au/info/industry/useful-resources/glossary

Cost-utility analysis

A CUA presents the health outcomes in terms of QALYs, which represent society's preferences for the health outcome experiences relative to full health.

A CUA is preferred over a CEA, particularly where:

- there is a claim of incremental life years gained in the economic evaluation (to assess the impact of quality on that survival gain)
- there is an improvement in quality, but not quantity, of life
- relevant direct randomised trials report results using a multiattribute utility instrument (MAUI).

Where transformations or external data sources are required to estimate QALYs, present a stepped transformation from a CEA to a CUA, to transparently indicate the implications of the transformation and/or use of external data.

Other relevant factors, including prognosis, severity, age, distributional effect, context (e.g. emergency or prevention), and other equity and ethical issues that are ignored in measurements using a MAUI, should be considered alongside – but not within – a CUA. Where this is important and relevant, the assessment report should draw these issues to MSAC's attention.

Cost-effectiveness analysis

A CEA measures the incremental cost per extra unit of health outcome (expressed in natural units such as life years) achieved. Where a CEA is presented as the primary economic evaluation, justify why the quantified health outcomes are not translated into QALYs and presented as a CUA.

Ensure that the incremental health outcome (e.g. life years, accurate diagnosis or other health event) presented in a CEA is patient-relevant. Present the outcome measure that is most closely and validly representative of the overall health of the patient, from their perspective, and in the context of the disease or condition for which they are receiving the proposed health technology. For investigative technologies where there is a personal utility arising from knowledge of the test result (such as those that claim to assist with reproductive planning), outcomes could include at-risk couples identified, or couples whose risk status is identified. Justify the choice of outcome and describe the extent to which the outcome captures all relevant health considerations.

Where a combination of outcomes (either intermediate or final outcomes, or both) are relevant to the patient, capture these collectively. Ideally, these would be transformed into QALYs and combined in a CUA, rather than presenting CEAs for multiple outcomes. Where this is not possible, additional CCAs may be useful.

Cost-consequences analysis

A CCA compares the incremental costs of the proposed health technology with the comparator, and presents the various incremental differences in a range of relevant (disaggregated) outcomes. A CCA can be useful when the proposed health technology is demonstrated to have a different profile of effects that are not adequately captured by a single outcome measure, and where there might be trade-offs in effectiveness and safety between the intervention and the comparator.

Generally, a CCA should not be presented on its own, but it may be useful as a supplementary analysis to a CUA or a CEA. Disaggregated analyses may demonstrate changes in patterns of health care resource use or specific health outcomes of interest that are not obvious in an aggregated evaluation.

Cost-benefit analysis

Cost-benefit analysis does not incorporate the breadth of considerations that are relevant to MSAC decision-making. There are limitations to the process of eliciting monetary valuations of health, particularly in the context of the Australian health care system where individuals do not face market prices. A cost-benefit analysis is unlikely to be helpful to the MSAC decision-making process.

TG 17.6 Generation of the base case

Within-study economic evaluation

A trial-based evaluation is sufficient to provide the base case of the economic evaluation if the evidence presented in Section 2 of the assessment report:

- recruited patients who are representative of those for whom listing is sought
- tested the proposed health technology in the circumstances of use expected to apply to the requested MBS listing
- directly measured and reported patient-relevant end points over an appropriate time horizon.

Modelled economic evaluation (including stepped adjustments to a trial-based evaluation)

If evidence for clinical effectiveness was synthesised from multiple sources, or if the trial(s) did not provide evidence that sufficiently measures the full clinical and economic performance of the proposed health technology compared with its main comparator in the Australian setting, use modelling and/or adjustments to the trial data to generate the base-case economic evaluation.

Justify and make transparent any translations of the primary effectiveness data and additional assumptions used in the model. Construct economic models in a way that allows the results to be presented sequentially before and after key translational steps.

The stepped approach may include some or all of the following stages:

- Present the outcomes and costs as identified in the key trial(s) (see Technical Guidance 21 and Technical Guidance 22).
- Adjust treatment effects on health care resource use and health outcomes, as would be anticipated in the Australian setting and MBS population defined in the proposed item descriptor (see Technical Guidance 19). This may involve one or more steps – for example,
  - re-estimate the treatment effect in the MBS population (e.g. use selected subgroups or weighted trial outcomes to improve applicability to the Australian setting)
  - incorporate Australian circumstances of use or clinical practice (e.g. with respect to patterns of resource use)
  - incorporate other necessary and justifiable assumptions to improve the representativeness of the model (e.g. resource use or outcomes associated with adverse event data, or subsequent treatment lines that are not captured in the trial data or previous translations).
- Extrapolate health care resource use and health outcomes (for the proposed MBS use) as required over the appropriate time horizon (see Technical Guidance 20).
- Transform health outcomes, if necessary, to the final outcomes used in the economic evaluation (e.g. using utility weights to obtain QALYs) (detailed in Technical Guidance 21).

The stages included in the stepped approach may vary depending on the nature of the available data. The base-case result is represented by the final incremental costs and outcomes, and the incremental cost-effectiveness ratio after the evidence from the main trial(s) has been translated.

For investigative technologies, this may include sequential incorporation of the linked evidence (e.g. assuming perfect test accuracy and change in management due to the test result in the first step, then sequentially relaxing these assumptions). This enables MSAC to identify which steps of the linked evidence have the greatest effect on the cost-effectiveness of the test.

A table should be presented in the assessment report that summarises the steps undertaken in the economic analysis.

Technical Guidance 18 Model development process

KEY CONSIDERATIONS

- **Report the processes for developing the structure of the model, present the final model structure and explain how this differs from a conceptual model. Identify structural assumptions (see TG 18.1).**
- **Justify the choice of the time horizon (TG 18.2).**
- **Explain the how the model computationally functions (TG 18.3).**
- **Provide the sources of the input data, and justification for the selection of the inputs (TG 18.4).**

The model structure should capture all relevant and important health states or clinical events along the disease or condition pathway, and should be consistent with the treatment and disease or condition algorithms created in response to Technical Guidance 2. For investigative technologies, the model structure may need to account for prevalence of disease (or risk stratification for a prognostic technology), test accuracy (including cost and health outcome implications for patients who receive a false result or those in whom testing fails), change in management, and the effect of change in management, where relevant (see Figure 26 for an example of a model structure for a diagnostic test).

For evaluations in which it is reasonable to model multiple distinct populations (such as where a test identifies multiple distinct diseases with differing treatments and prognoses), the model structuring process should be performed for each distinct population (with some indication as to how these are structurally combined). In such circumstances, the model should be structured to allow the cost-effectiveness of the health technology to be determined for both disaggregated and aggregated populations. In the case of testing for heritable diseases, the model should be structured to allow the cost-effectiveness to be explored under incremental expansion of the test from index cases, to index cases plus first-degree relatives, and to index cases plus first- and second-degree relatives, and so on, as considered relevant by the PICO Advisory Sub-Committee.

The model structuring process should be transparent and clearly described. This process includes model conceptualisation, choice of computational method and consideration of other structural assumptions.[55-61]

Assumptions incorporated into the model structure should be explicitly specified. State how these assumptions have been tested in sensitivity analyses (see Technical Guidance 25).

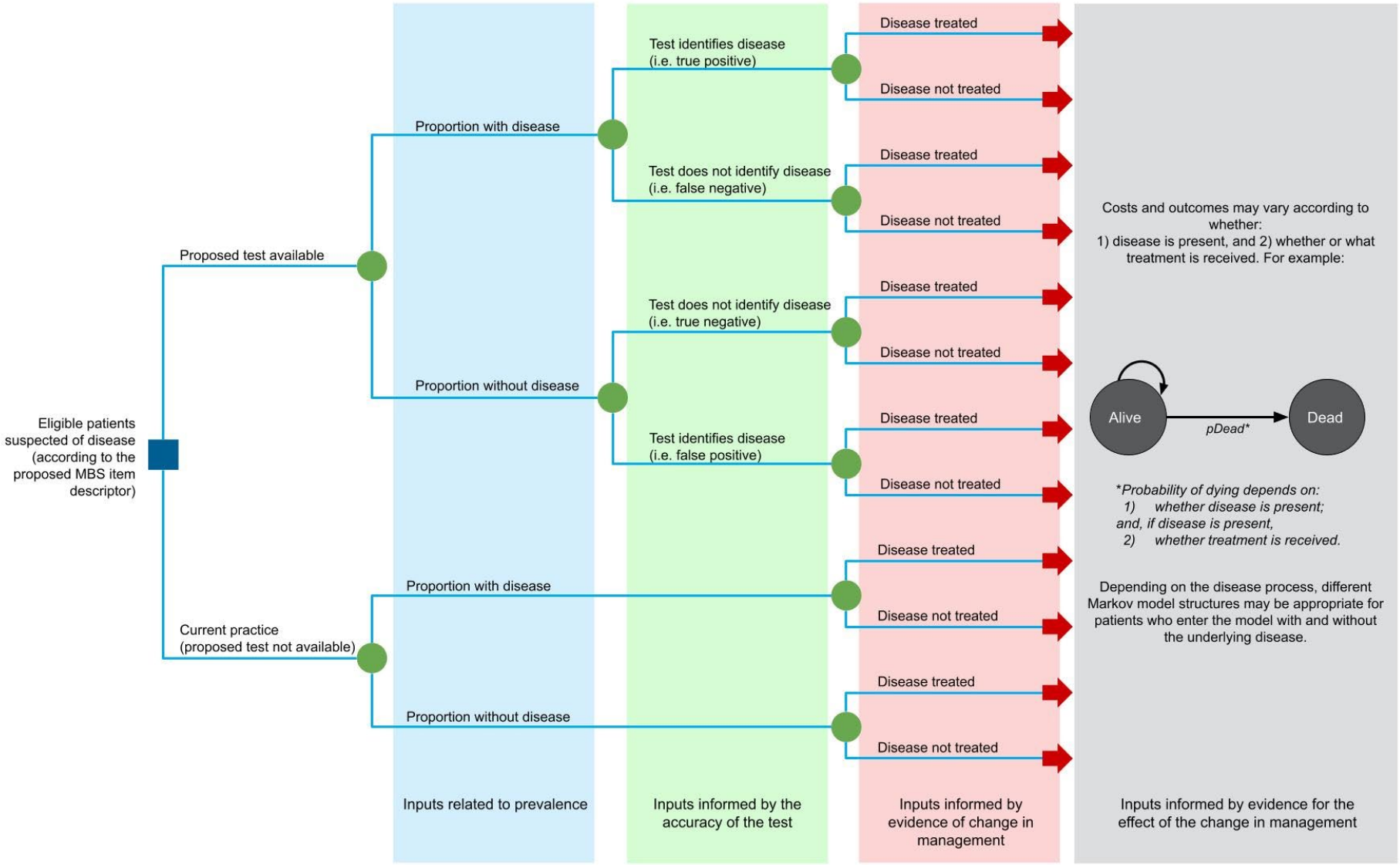


Figure 26 Example model structure for an investigative health technology

TG 18.1 Model conceptualisation process

The model conceptualisation process should be clearly described. The process should be driven by the assessment question, rather than by data availability. The following summarises this process to aid a transparent approach to model conceptualisation.

Literature review

Present the results of a literature search for economic evaluations of similar decision analyses (in terms of similarity to the treatment algorithm and/or the proposed and similar health technologies), focusing on the structure of the existing models. This may include MSAC Public Summary Documents or other reports of similar technologies previously considered by MSAC, and models considered by other HTA agencies.

Present any additional literature (e.g. additional clinical trials, clinical guidelines, natural history studies, burden of disease studies, surveys) that informs the model structure and that has not already been presented in Sections 1 or 2 of the assessment report. Provide copies of the original sources of all data not already presented in Section 2, or expert opinion used in the model, in an attachment.

The review of the literature should also identify whether there are important patient attributes that may influence the risk of experiencing subsequent events or disease progression, as this may inform the choice of computational method (see TG 18.3).

Conceptual model

A figure depicting the conceptual model should be clearly presented. This should include all clinically relevant and significant health states/events that were identified from the review of the literature. Significant health states/events are defined with respect to the strength of their relationship with the condition of interest, as well as their potential impact on associated costs and/or economically important health outcomes such as QALYs. The health states/events should be disaggregated if there are likely to be important differences between the disaggregated states/events in terms of disease progression, associated costs or associated health outcomes (e.g. QALYs). These health states/events should form the basis of the model structure used in the economic analysis.

Final model structure

The conceptual model should be reviewed within the context of the available data. If adequate input data are not available to populate the model as conceptualised, alternative model structures should be identified that better conform to the available data. The face validity of these alternative model structures should be assessed.[62] The adaptations to the final model structure should be clearly justified and described, including discussion of any potential effects of these adaptations on the model outputs.

If multiple plausible model structures are identified (e.g. alternative health states/events), these should be clearly presented and tested as part of structural sensitivity analyses. The impact of these alternative plausible structural assumptions on model predictions should be clearly discussed (see Technical Guidance 25).

Other structural choices/assumptions

Other structural assumptions used in the model should be fully documented and justified, including how these have been tested in structural sensitivity analyses. Examples of other structural assumptions reported in the literature include:[56]

- the relationship between time and transition probabilities, including time dependency of probabilities (e.g. if an event is more likely to occur with time in a given health state)
- which model transitions the proposed health technology has an effect on (such as affecting transitions related to the initial incidence of disease or an event, but not affecting transitions after the disease or event has occurred)
- the duration of treatment effects beyond the observed period in the empirical data (e.g. if the treatment effect is assumed to continue beyond the observed period)
- the choice of statistical method for estimating health outcomes beyond the empirical data (see Technical Guidance 20).

TG 18.2 Time horizon

Define and justify the time horizon over which the costs and outcomes of the proposed health technology and its main comparator are estimated. Ensure that the time horizon captures all important differences in costs and outcomes between the intervention and the comparator as a result of the choice of treatment, but does not extend unnecessarily beyond this. The same time horizon should be used for both costs and health outcomes.

Where interventions do not affect mortality and have temporary health or quality-of-life effects, a relatively short time horizon may be appropriate.

Where there is evidence that a health technology affects mortality or long-term/ongoing quality of life, a lifetime time horizon is appropriate. Note that a lifetime time horizon relates to the life expectancy of the relevant patient population and reflects the time span required for nearly all of the model cohort to die. If the patient cohort initiated in the model has a broad distribution of ages or prognoses, explain the impact of this distribution on the model time horizon.

The validity of the lifetime horizon is determined by the population of the model and the inputs; it is not an independently nominated duration. Inputs that are not realistic will result in the model predicting an implausible duration of outcomes or survival and, thus, an implausible lifetime time horizon. The assessment of plausibility is also critical when considering how the model extrapolates data to reach the nominated lifetime time horizon (see Technical Guidance 20).

As a modelled time horizon extends – in absolute terms and relative to available data – it is associated with increasing inherent uncertainty. Therefore, economic claims based on models with very extended time horizons and predominantly extrapolated benefits will be less certain and are likely to be less convincing to MSAC. Technical Guidance 20 and Technical Guidance 25 address the extrapolation of costs and outcomes for an extended time horizon and the associated uncertainty.

TG 18.3 Computational methods

If a trial-based economic evaluation is being undertaken using individual patient data on costs and outcomes from a clinical trial(s), describe the methods and software used to do this.

For model-based economic evaluations, identify the modelling technique that is most appropriate for implementing the final model structure(s).[63] Generally, select the least complicated modelling

technique for which the specified model structure can be feasibly implemented, moving from decision trees to cohort-based state transition models to individual-level modelling techniques.

For some technologies (e.g. investigative technologies), approaches that combine decision trees with other modelling techniques, such as cohort-based state transition models, might be appropriate.

Decision trees

Decision trees are useful for models with short time horizons. General spreadsheet software (e.g. Excel) or specialist software (e.g. TreeAge) can be used. Follow good-practice guidelines for using decision trees.[64]

Cohort-based state transition (or Markov) models

Use cohort-based state transition models to represent longer time horizons for models that can be represented using a manageable number of health states under the constraints of the Markovian (memoryless) assumption. General spreadsheet software (e.g. Excel) or specialist software (e.g. TreeAge) can be used.

Follow good-practice guidelines for using state transition models.[65] In particular, consider the following questions when implementing a cohort-based state transition model:

- Is it reasonable to assume that the transition probabilities from each defined health state are independent of states that may have been experienced before entering that health state? Health states that describe pathways through the model can be used to represent the effects of previous events on subsequent transition probabilities.
- Do transition probabilities vary according to how long individuals have remained in each health state? Tunnel states may be required to represent time-varying transition probabilities.
- Is the eligible population homogeneous, or is variation in patients normally distributed? This issue commonly refers to the age of the eligible population, but may include other factors. If relevant factors are not normally distributed, run separate analyses of the model and aggregate the outputs, or consider using a microsimulation model.
- What is the likely impact of alternative cycle lengths on the model outputs? Describe the factors determining the selected cycle length.

A half-cycle correction is the default approach to representing the time of transition between states, although an alternative correction factor may be proposed with justification.

Partitioned survival analysis (or area under the curve modelling)

Partitioned survival analysis models are conceptually similar to Markov models in that they are characterised by a series of health states with associated state values. However, health state membership is not estimated using transition probabilities; rather, it is derived from a set of independently modelled non-mutually exclusive survival curves (e.g. overall survival, progression-free survival).[66, 67] Where effectiveness outcomes are reported using non-mutually exclusive survival curves, a partitioned survival approach may be used. Depending on the maturity of the data available, statistical extrapolation beyond the observed data may be required to model outcomes to the nominated time horizon (see Technical Guidance 20).

Where a partitioned survival analysis approach is used, justification for the key structural assumptions associated with this approach should be provided (i.e. that all end points, including overall survival, are modelled and extrapolated independently, and that transitions between health states are not explicitly modelled).[67]

Individual-level (or microsimulation) models

Use individual-level modelling approaches only when a defined model structure cannot be feasibly implemented as a cohort-based model. Describe the characteristics of the model structure that prevent a cohort-based model being used. Potential factors include baseline heterogeneity, continuous disease or condition markers, time-varying event rates, and the influence of previous events on subsequent event rates.[68] Also describe how incorporation of these features in an individual-level model are expected to produce a more accurate representation of the disease or condition pathways, costs and patient outcomes.

The most common individual-level approaches include individual-based state transition and discrete event simulation models. Follow published guidelines on good research practices for applying these models.[65, 69] Discuss any requirements for specialist software with the Department in advance.

Other modelling techniques

If the results from simpler models are robust enough to produce plausible sensitivity and scenario analyses, it is not necessary to use more complex modelling techniques.[70] If an alternative modelling technique is used, describe and justify how the approach leads to more accurate and valid results. For example, in the clinical area of infectious diseases, the use of dynamic transition models or agent-based models to represent herd immunity may be justified if a simple nondynamic model will not demonstrate cost-effectiveness accurately enough.

Note that more complex modelling techniques may be less transparent, and the model assumptions less certain. This might result in MSAC having less confidence in the cost-effectiveness claim. **Discuss the use of complex modelling techniques (including any specialist software) with the Department in advance.**

TG 18.4 Input data

Where possible, input data should be sourced from the evidence presented in Section 2 of the assessment report. At the 'source-of-evidence' level, identify which model inputs are derived from the clinical evidence presented in Section 2, and which are derived from alternative data sources. Where multiple sources of data were identified in Section 2 to inform a particular parameter, the justification for the input used in the base-case analysis should be provided in the relevant subsection of the assessment report, including which alternative sources of input data are used in sensitivity analyses.

Justification to support the approach used in the economic evaluation is required if one or more studies presented in Section 2:

- had reliability issues (e.g. due to inadequate concealment of randomisation, or inadequate blinding of subjective outcomes)
- reported few or no patient-relevant outcomes
- were of insufficient duration to detect the most patient-relevant outcomes
- reported outcomes that could not be translated into the economic analysis.

Where relevant applicability issues with clinical data from Section 2 are identified, these should be discussed and translated to the Australian population and setting, if necessary, in Section 3A.1.3 of the assessment report.

If data beyond the clinical evidence presented in Section 2 are used to populate the model input parameters, describe the methods used to identify the additional data – for example, whether

systematic or ad hoc reviews of the literature were undertaken, or how relevant primary data sources, including registries and observational studies, were identified. The method of identifying the data should be robust and transparent. Where multiple sources of data exist, the source of the input used in the base case should be described and justified.

Applicability concerns (and any translation issues) relating to additional data should be described in the relevant subsection of the assessment report.

TG 18.5 Fully editable electronic copy of the economic evaluation

Provide an electronic copy of the economic evaluation. The economic evaluation should be constructed in line with best practices.[71] Ensure that all variables can be changed independently, including allowing the base case of the economic evaluation to be respecified and a new set of sensitivity analyses to be conducted with each respecified base case. Ensure that the economic evaluation can produce results following respecification of variables within reasonable running times. To help understand the electronic copy of the economic evaluation, apply clear and unambiguous labels to values, and cross-reference data sources.

The following software packages do not need prearrangement with the Department:

- TreeAge Pro
- Excel, including @RISK.

Use of other specialist software, including advanced features and plug-ins for Excel, must be prearranged with the Department before submission.

Technical Guidance 19 Population and setting

KEY CONSIDERATIONS

- **The population in the economic evaluation reflects the target Australian population (TG 19.1).**
- **Describe the model population and applicability issues associated with the clinical evidence (either identified in response to Section 2, or sourced separately to inform the economic evaluation) (TG 19.2).**

TG 19.1 Demographic and patient characteristics, and circumstances of use

The setting of the economic evaluation should be the Australian health care setting, with the modelled population representing the target Australian population indicated for use of the proposed health technology. The circumstances of use should be consistent with the clinical management algorithm and the indication specified in the proposed item descriptor (Section 1).

Describe the demographic and clinical characteristics of the modelled population using summary statistics, including information on distributions around the central estimate (e.g. standard deviations, confidence intervals). Relevant patient and clinical characteristics may include age, sex, ethnicity, medical condition and severity of the medical condition, and comorbidities. Indicate which patient characteristics are incorporated explicitly, and which are implicit (associated with use of other data) or not included.

For investigative technologies, the modelled population includes all patients eligible for the test – not just those that the test aims to identify. The prevalence(s) of the target (e.g. disease, subtype or pathological variant) in the tested population used in the model should be reported and should be consistent with that identified in Section 2.

Describe and justify how heterogeneity in patient characteristics (if relevant) is represented in the economic evaluation. Heterogeneity could arise where multiple distinct populations are identified by a test (and so multiple indications are modelled), or when the test identifies a heritable disease (and so the eligible populations modelled include those suspected of being an index case, and the relatives of those in whom the heritable disease is identified).

Provide details of any additional circumstances of use relating to the proposed health technology that are relevant to the model setting or population, and detail how they are incorporated into the model. These may include:

- the position of the service in the overall algorithm for diagnosing, treating or managing the disease or condition (e.g. prevention, first-line treatment, second-line treatment)
- any limitations on the duration or frequency of delivery of the service; for example, in a 24-hour period, or a 12- or 24-month period
- any required co-delivered medical services or treatments (including any additional diagnostic tests required)
- any contraindicated medical services or treatments
- any unique characteristics of the referrer or provider (e.g. specific qualifications or training)
- any specific requirements in terms of geography, facilities or location of delivery of the service (e.g. limited to the hospital setting or to approved laboratories, or any specific equipment or facilities that need to be available).

TG 19.2 Applicability issues and translation studies associated with the clinical evidence

If differences between the clinical evidence setting(s) and the Australian setting are identified in Section 2 of the assessment report as being potentially important, design a translation study for each difference. This includes differences in the populations, disease or condition, circumstances of use, or treatments between the evidence presented and what would be expected if the proposed health technology is reimbursed according to the proposed MBS item descriptor and in accordance with the proposed clinical management algorithm. For investigative technologies, applicability issues arising from a potential change in the spectrum of disease identified by the test, or the transitivity across the studies included in the linked evidence approach, should be addressed.

Table 34, Appendix 6, lists examples of factors that, when different across settings, may result in a difference in treatment effect, adverse events or patient management across those settings.

Each translation study should determine whether quantitative adjustments to model inputs are necessary and, if so, the nature of the appropriate translation. If there are inadequate data for a translation study, identify this as an issue that will remain a source of uncertainty in the model.

The translation study should include:

- the issue and the specific question to be addressed
- the data used and their sources (justify the choice of data if there are multiple possible sources)
- the methods of analysis (common methods are described below), with sufficient details to enable independent verification of the analysis
- the results, which for therapeutic technologies might include an estimate of the comparative treatment effect (both relative and absolute) and the 95% confidence interval, and a description of how (or whether) the findings are applied in the model
- a description of any residual uncertainty, and sensitivity analyses that are proposed to address this uncertainty.

Take care when converting relative treatment effects or estimates of accuracy across jurisdictions with different baseline risks. Use measures of sensitivity and specificity, rather than positive predictive value and negative predictive value. Ensure that the baseline risk (i.e. prognostic characteristics) of patients is the same in the trial evidence and the target MBS population, or that patients are not expected to respond better to the proposed health technology or the main comparator in one setting compared to another.

Common methods for translation include subgroup analyses, regression analyses or meta-regression, or use of other published studies. Justify the selected approach.

Subgroup analysis

For subgroup analyses, follow the methods outlined in Technical Guidance 6.

Regression or meta-regression

Regression analysis has an advantage over stratified analyses based on subgroups because it can examine more than one covariate (or difference between the clinical trial participants and the target MBS population) simultaneously. Where multiple trials are available, use a meta-regression, if appropriate. Meta-regression may be used at the study level or at the individual patient level (where

the study is entered as a covariate). Only use a meta-regression at the study level if the number of trials is large (5–10 trials for each covariate examined).

If a regression analysis is used, present and interpret the results in the main body of the assessment report, and provide the following additional details in an attachment:

- a clear description of the regression method, the associated assumptions, how these assumptions were tested and the results of the tests
- the statistical commands or syntax used in the analysis, with a description of the variables (including a description of the thresholds used to define categorical variables)
- the direct output from the statistical program
- the dataset used in the statistical program (or a justification, if the dataset is not provided).

Published studies

If it is not possible to inform translation using the direct clinical evidence for the intervention, describe the reasons and seek relevant published data. Systematically identify published studies concerning the proposed health technology (or comparator) in the proposed eligible population. Present the search strategy and selection criteria in an attachment.

Report the relevant findings from the included studies. Describe the findings in relation to the proposed health technology and apply the findings to inform the translation.

Technical Guidance 20 Model transition probabilities, variables and extrapolation

KEY CONSIDERATIONS

- **Transition probabilities (or other methods for establishing health state membership) determine how modelled patients move through the health states in the model. Explain the derivation of transition probabilities (TG 20.1).**
- **Justify the need to extend transitions beyond the duration of the clinical study data, and explain methods for extrapolation (TG 20.2).**

TG 20.1 Transition probabilities and variables

Transition probabilities inform the movement of patients between health states in decision trees or state transition models. For investigative technologies, this also includes decision-tree parameters related to test accuracy and changes in clinical management, where relevant. In a discrete event simulation, time-to-event parameters are analogous to transition probabilities. Transition probabilities or time-to-event parameters may differ by treatment or by how long a patient has been in a particular health state (time-varying probabilities).

Transition probabilities that differ by treatment are generally estimated using the clinical evidence described in Section 2 of the assessment report (with applicability translation, as per Technical Guidance 19, as appropriate). Cross-reference the relevant subsections for the clinical evidence and note whether further translation studies or extrapolations are required.

Other transition probabilities may be required that describe the progression of a disease or condition following an intermediate modelled event, and for which the same transition probabilities are applied, regardless of treatment allocation. Where external sources of data (other than the clinical trials from Section 2) are used to inform transition probabilities (or other variables) in the model, assess the applicability of these sources of data to the Australian setting. Note and justify whether the data are applicable, require translation (in which case, follow the approach detailed in Technical Guidance 19) or are a source of uncertainty within the model.

Detail where the model uses other variables instead of, or in addition to, transition probabilities (such as allocation to a medical management pathway), and justify the source of these input variables in the same manner. Do not include variables associated with the valuation of outcomes or costs; these are described in response to Technical Guidance 21 and Technical Guidance 22, respectively.

Describe and justify the methods used to identify and analyse relevant data to derive transition probabilities and variables.

For each transition probability or variable, present the point estimate and interval estimates (e.g. 95% confidence intervals). Follow good-practice guidelines when choosing the methods to derive interval estimates (e.g. using probability distributions based on agreed statistical methods for alternative types of input parameters).[72] Ensure that values taken from all sources of evidence are appropriately adjusted to represent the transitions required by the model structure.[73] For example, translate reported rates or cumulative probabilities to the probabilities for timeframes associated with a model cycle, if necessary.

Occasionally, secondary outcomes and other trial-derived data (e.g. adverse event rates) are relevant to outcomes and/or resource use in the economic model, and point estimates are

numerically different across the arms but not statistically significantly different. This may reflect either no 'real' difference, or that the trial had insufficient power to demonstrate a difference statistically. Explain the approach used to inform the probability in the base-case model (e.g. whether it has been pooled across arms or differentiated between arms), and justify with supporting evidence, if available. Examine the alternative approach in a sensitivity analysis.

Assess the potential correlation between transition probabilities and/or variables. Correlation between parameters is explored further in Technical Guidance 25 for uncertainty analysis.

TG 20.2 Extrapolation

Extrapolation may be justified when all important differences in costs and outcomes between the intervention and comparator(s) groups are not represented over the time horizon for which observed data are available. Detail any extrapolations of data that are required for the base-case economic model.

Extrapolating time-to-event data

Where extrapolation of time-to-event data is required, use observed time-to-event data in preference to modelled data up to the time point at which the observed data become unreliable as a result of small numbers of patients remaining event-free.

Describe and justify the selected time point beyond which extrapolated transition probabilities are applied. External data may be used to justify the selected time point – for example, the point at which one or more of the curves fitted to the clinical trial data deviates from a curve fitted to observational data from a similar patient cohort with a larger sample over a longer follow-up period. Test alternative truncation points in the sensitivity analysis.

Derive appropriately estimated parametric survival curves based on the observed data (using individual patient data, if available) to extrapolate transition probabilities beyond the data truncation point.

If extrapolated transition probabilities have been applied:

- Discuss whether an assumption of proportional hazards is appropriate beyond the observed data.
- Fit a range of alternative survival models to the observed data (e.g. exponential, Weibull, log-normal, log-logistic, gamma, Gompertz). Include more flexible extrapolation approaches with multiple points of inflexion (e.g. piecewise spline models) to better facilitate extrapolation based on the section of the Kaplan–Meier curve that is most representative of long-term survival.[74]
- Assess and discuss goodness of fit using visual inspection, Akaike's information criterion and Bayesian information criterion. Justify the most appropriate model for the base case and test a number of the best-fitting models in the sensitivity analysis.
- Discuss the plausibility of the predictions in the unobserved period (e.g. the ongoing hazard ratio and/or treatment effect, the point of convergence, and/or residual survival in each arm).

The treatment effect resulting from the independent extrapolation of the survival curves should be plotted over the time horizon of the model. If the treatment effect is maintained or increasing, and this is not clinically plausible, apply a hazard ratio such that the intervention and comparator curves converge at a plausible time point. The assessment of plausibility should be linked to the justification of the time horizon (see Technical Guidance 18).

When considering the extrapolated treatment effect, give explicit consideration to clinical decisions about stopping or continuing treatment. State and justify all assumptions in this regard, and apply them consistently when modelling respective treatment costs.

Numerous sources of advice on extrapolation techniques for economic evaluation are available in the literature.[75-80]

Other individual patient extrapolation issues

For categorical data that describe the experience of multiple intermediate or outcome events, use a 2-stage process of modelling the time to any event, combined with a multinomial logistic model to define the probabilities of the aggregate event being each of the competing events. Include a time covariate in the multinomial logistic model to represent time-varying probabilities, if possible. The other option is to fit independent competing risks time-to-event models for each event, but this approach is likely to overestimate parameter uncertainty as a result of the assumed independence of the multiple events modelled.

For continuous variables, format the data into categories, or use a generalised estimating equation model.

Extrapolating published time-to-event data

If individual patient time-to-event data are not available, extrapolate survival probabilities from published Kaplan–Meier curves using graph digitiser software. Fit alternative constant (i.e. exponential) or monotonically increasing and decreasing (e.g. Weibull or Gompertz) hazard functions to the extracted survival data, from the last point of inflexion to the time point at which the observed data become unreliable because of small numbers of patients remaining event-free.

Present tests of the relative and absolute goodness of fit of the alternative curves, and use the best-fitting curve in the base case. Test the alternative models in sensitivity analyses.

Use of data from other non-randomised studies to extrapolate beyond the evidence

Data from other non-randomised studies may sometimes be useful to extrapolate beyond the results of the clinical evidence presented in Section 2. This is because the included studies might have been of insufficient size or duration to capture the full impact of therapy on the outcomes of the disease, or the typical resource use measured in an overseas trial might need adjustment to reflect patterns of resource use in Australia. In contrast, other non-randomised studies might involve longer follow-up for an active main comparator, or the natural history of the medical condition if the main comparator is not an active intervention. Given that the data from non-randomised studies are subject to bias, be cautious when using assumptions based on those data during modelling.

When presenting data from other non-randomised studies for extrapolation purposes in a modelled economic evaluation, demonstrate that a systematic approach has been taken to search for, locate and select the non-randomised studies for presentation. The selection process should be presented and justified. Provide a report of each study in a technical document or attachment. The results of the non-randomised study might contribute to finding and justifying a variable in the economic evaluation. This variable might vary from a single point estimate to a regression formula. The results of the non-randomised study might also help identify risk factors that contribute to the expected risks of the comparator arm in a model.

When indicating which results are being extrapolated, explain how the extrapolations are achieved by the model for the streams of costs and outcomes for the proposed health technology and the main comparator. In particular, if non-comparative data are used (e.g. from single-arm studies), it is

necessary to make an assumption about how the other arm in the model would change. The usual practice, in the absence of empirical evidence to the contrary, is to assume that the comparator arm would change so that the relative risk between the 2 arms measured in the randomised trial(s) remains constant across the duration of therapy. Justify the use of this (or any other) assumption in the model presented in the assessment report.

Technical Guidance 21 Health outcomes

KEY CONSIDERATIONS

- **Economic evaluations report health outcomes. State the final health outcome presented in the model results. Explain the inputs and any translation studies required to generate the final health outcomes (TG 21.1).**

TG 21.1 Health outcomes

Nominate and justify the final health outcome that is considered to best reflect the comparative clinical performance of the interventions and will form the denominator unit in the base-case ICER.

Detail the health outcome(s) (intermediate and/or final) that inform the final outcome in the economic evaluation. Explain whether these were reported directly in the clinical evaluation (Section 2) and, if not, summarise the transformations involved to obtain the final outcome.

If available, use quality-of-life or utility data reported in Section 2 to estimate QALYs in the model. If alternative indirect methods are used to estimate QALYs when direct data are available, justify the use of these methods. Present both sets of methods and results, and compare the interpretation.

Present the results of any utility study as the point estimate of the mean elicited utility weight for each health state, and include its standard deviation and 95% confidence interval, where available.

If a claim is made for a change in a nonhealth outcome, or the assessment report identifies health-related outcomes in people other than the patient receiving treatment (e.g. quality-of-life benefits for family, decreased carer burden), these generally should not be included in the base-case evaluation, but could be included in supplementary analyses (see Appendix 10).

Use of quality-of-life data from the clinical trials to estimate QALYs

Estimates of quality of life or utility from the evidence presented in Section 2 may inform direct estimates of QALY gains in the intervention and comparator populations, or inform utility values applied to health states in an economic model.

If a MAUI has been used in a study included in Section 2 to estimate utility weights, state where and when the scoring algorithm was derived, and consider how applicable it is to the general Australian population. It is preferred that Australian-based preference weights are used in the scoring algorithm for calculating utility weights.

If the initial patient-reported outcome measure is not a MAUI, provide details of the measure used and justification of its use in Section 2. Describe a validated method of mapping the results into preference weights (see 'Mapping of generic and disease-specific scales' below). State whether Australian-based value sets are incorporated. If there is no reliable method of transforming the patient-reported outcome data into utility weights for the model, describe why this is not possible and detail whether the patient-reported outcome data from the trial can still be used to inform or validate the economic model.

Consider the duration over which the patient-reported outcome measure informing utilities was administered compared with the duration of the condition of interest. If a generic MAUI or patient-reported outcome measure was used, consider whether it captures all important disease- or condition-specific factors that might be relevant.

Address the following questions when incorporating trial-based patient-reported outcome data into the economic model:

- Are the participants representative of the population for whom listing is requested? (Refer to Section 3A.1.3 of the assessment report, as needed.)
- If quality of life is not the primary outcome, is the trial adequately powered to detect a difference in the survey results? As with all secondary outcomes, assess the results with reference to the conclusion from the primary analysis of the trial.
- Is there a 'healthy cohort effect'? (i.e. where the sickest patients are least likely to complete patient-reported outcome data forms, and therefore the data obtained are biased towards healthier patients). Consider the responder numbers and drop-outs. While generally associated with an overestimate of utility weights, the direction of any associated bias may depend on whether the treatment and comparator are associated with different utilities, the relative extent of the effect across different arms and health states, and the time spent in different health states. Identify any impact of bias on the overall ICER.
- Is there potential for systematic bias where progressed health states are defined by nonsymptomatic events (i.e. identified by investigations that may or may not reflect clinical practice)? Provide details.
- Is it appropriate to pool patient-reported outcome data across arms of a trial? This may be appropriate where patient numbers are small and for post-treatment states, but not in other circumstances where treatment (rather than disease or condition) directly affects quality of life (e.g. because of serious adverse events and any associated long-term implications, or imposed limitations). Justify the approach and, where possible, present results with and without pooling.
- Is there a risk of bias from a regression to the mean effect?[81] This may be more likely in instances where quality of life for the control arm is drawn from a trial other than a randomised controlled trial (e.g. from a pre-intervention population).

Use of other sources of data to estimate utility weights

Where utility weights or QALY changes cannot be directly estimated from data collected in the clinical studies from Section 2, or there are significant concerns about the reliability and relevance of trial-based utility, transform the Section 2 health outcomes to estimate QALY gains (e.g. by applying utility weights to the time spent in different health states that represent the experience of clinical outcomes).

Additional studies (either published or commissioned for the assessment report) may be needed to estimate utility weights for health states in the economic model. These studies should be identified, and copies provided.

Describe the source(s) and method(s) (as described in the following sections) used to generate externally derived health state utilities and justify their inclusion in the model.[82] Depending on the clinical context and available data, there may be more than one acceptable source of utility weights. In this case, reflect the uncertainty in selecting an optimal source of weights by reporting the sensitivity of the result to switching between the various sources.

Address the questions regarding quality-of-life data derived from the clinical trials (listed above) that are applicable to any utility estimates obtained from alternative sources and methods.

Mapping of generic and disease-specific scales

Non-preference-based patient-reported outcome measures will need to be transformed into preference-based measures, using a mapping algorithm, to estimate utilities. Where this occurs, provide the source of the mapping algorithm. Describe the estimation sample (e.g. population demographic and clinical characteristics, sample size) and whether there is an external validation sample. Provide details of the initial patient reported outcome measure and target measures (e.g. index, dimensional), and the statistical model and methods used to estimate the mapping algorithm. Detail the statistical associations or operations that constitute the algorithm. Discuss methods used to measure the algorithm performance and validity. Present the resulting predicted utilities with their associated uncertainty. Discuss the applicability to the data presented in the assessment report, particularly in relation to the sample in which the algorithm was developed.

Scenario-based methods to indirectly elicit utility weights

Scenario-based methods use vignettes to describe the symptoms of a health state in a sample population, usually a representative general population sample, from which utility weights are elicited using an accepted preference-based method. Methods to elicit preferences include the standard gamble, time trade-off and discrete choice experiments, and other stated preference methods.

If using a scenario-based utility valuation to value health outcomes beyond the time horizon of a trial, include one or more health states captured and valued within the trial in the scenario-based study to validate the commonality of the trial-based and scenario-based utility weights.

Present supporting evidence for any claim that a scenario-based approach has increased sensitivity to identify real differences in utility.

Describe all stages of a scenario-based study in detail and explain efforts to minimise potential bias. It is difficult to minimise the many sources of analyst bias that are intrinsic to the scenario-based utility approach. Sources of bias include the non-blinded nature of the construction and presentation of the scenarios (e.g. incomplete inclusion and differential focus on alternative aspects of quality of life), the design of the methods to elicit values, and the analysis and interpretation of the results.

Population-matching study method to indirectly elicit utility weights

This form of utility study involves recruiting a separate sample of patients with characteristics similar to those enrolled in the clinical trials reported in Section 2. Matched patients complete a MAUI reflecting their current health state, which informs the estimation of utility weights for the health states in the economic model. See Technical Guidance 6 for further detail on MAUIs.

Potential sources of bias for such studies include the possibility of systematic differences between the clinical study participants and the matched patients, and the inability to blind the sampled patients from the objectives of the study. If there are important symptomatic toxicities, the sampled patients should possibly have been exposed to the health technology and its toxicities at the time the MAUI is completed.

Matched patients should complete other patient-reported outcome measures that were completed by the trial participants, and the results of this concurrent instrument should be used to more closely match utility study participants to the clinical study population.

Published sources of utility weights

Utility estimates may be available from the literature. The validity of the derived utility weights depends on the applied elicitation methods and the relevance of the study populations. Present details of search strategies, and inclusion and exclusion criteria used to identify relevant utility studies. Assess the validity of all identified studies, including:[82]

- how representative the health state in each identified study is of the health state in the economic evaluation (including the type and severity of symptoms, and the duration of the health state)
- how the health state was captured (e.g. MAUI, scenario-based)
- how the preference was elicited (e.g. standard gamble, time trade-off)
- what sample was chosen to respond to the MAUI questionnaire or scenario (e.g. the general public, patients, carers, health care professionals)
- the country in which the utility data were collected
- what assessment was made of the nature and direction of bias that might arise, given the sample and methods (report the variance in the utility estimates and response rates, extent of missing data or data lost to follow-up, and study type, i.e. observational study or randomised controlled trial)
- how the sensitivity analyses examined variation in the identified utility options.

The original published study for utilities should be cited, not a previous economic study that used this evidence.

Using different published studies to inform utility weights for alternative health states is discouraged because of the potential for inconsistency in the methods (e.g. instrument) and populations from which utilities were derived.

Utilities for concurrent clinical events can be estimated by:[82]

- subtracting the sum of the estimated utility decrements for overlapping events from the estimated utility in the absence of an event (additive method)
- multiplying the utility in the absence of an event by the product of the ratios of utilities for individuals with the clinical events to utilities for individuals who do not experience the clinical events (multiplicative method)
- using the lowest utility for all the clinical events (minimum method).

Good-practice guidelines for using health state utilities currently recommend the multiplicative method.[82] Alternate approaches may be presented in supplementary analyses, if relevant.

Presentation of outcomes and health utility value information

If presenting a CUA, summarise information on all modelled health outcomes (e.g. intermediate, final outcomes and events) contributing to the final health outcome in the economic evaluation, and any associated utilities or disutilities. A format for presenting the minimum information required is suggested in Table 14.

Table 14 Identification of health outcomes used in the model

Health state or event	Mean utility (SD and/or 95% CI) or QALY	Nature of estimate and any translations	Source of estimate	Alternative estimates of utility value (and sources)	Average application in the model: proposed health technology	Average application in the model: comparator
[Health state 1]	[Utility estimates for health state 1]	[e.g. EQ5D data (Australian value set)]	[e.g. From Trial 001 (see Section 2)]	[e.g. Nonpooled data from study]	[e.g. days/months]	[e.g. days/months]
[Health state 2]	[Utility estimates for health state 2]	[e.g. Scenario-based study using standard gamble method]	[e.g. External publication: Smith et al. 2010]	[e.g. External publication: Jones et al. 2008]	[e.g. days/months]	[e.g. days/months]
[Event 1]	[x QALYs per event]	[e.g. Scenario-based study using time trade-off method]	[e.g. Commissioned study (study report provided in attachment)]	[e.g. External publication: Jones et al. 2008]	[no. of events]	[no. of events]

CI = confidence interval; QALY = quality-adjusted life year; SD = standard deviation

Technical Guidance 22 Health care resource use and costs

KEY CONSIDERATIONS

- **Identify costs associated with the use of the health technology and downstream implications. Explain which costs are incorporated into the the model and justify the selection of the source for these costs (TG 22.1).**

TG 22.1 Health care resource use and costs

For within-trial analyses, identify the health care resource items that have a change in use associated with substituting the proposed health technology for the main comparator.

For model-based evaluations, estimate cost weights representing the resources used within a relevant time period (e.g. a model cycle for a state transition model) for every health state. Alternative health state costs may be defined for patients receiving the intervention and patients receiving the comparator – for example, to account for differences in adverse event rates.

Health care resource items

Where appropriate, consider the following resource items:

- medical services (i.e. procedures, diagnostic tests and investigative services), including the proposed and comparator health technologies if they are medical services
- public and/or private hospital services
- medicines, including pharmaceutical benefits
- blood products
- community-based services (e.g. attendances by specialists, general practitioners or allied health care professionals)
- any other direct medical costs.

Consider whether there are resource differences between who can request the proposed health technology and the main comparator (e.g. if the proposed health technology can only be requested by a specialist, whereas the comparator can be requested by a general practitioner).

For pathology services, consider whether use of patient episode initiation, specimen referral or block retrieval services would differ substantially between the intervention and the comparator. Where relevant, include the cost of obtaining a new sample and retesting.

For each resource item, define the natural units and quantify the number of natural units provided to patients in each treatment group, or to patients remaining in a health state for a relevant time period (e.g. number of services provided, number of packs of medicine dispensed, number of general practitioner consultations, number of episodes of hospital admission).

Use of the intervention and comparator services is generally derived from the clinical studies reported in Section 2. However, if a therapeutic health technology is provided multiple times over the treatment course and studies have incomplete follow-up, this may represent a truncated mean and require adjustment. Justify and explain any calculation of the cost per patient per year, as necessary, for therapeutic health technologies used episodically. If relevant, incorporate wastage in the model, because it is a consumption and therefore an incurred cost.

For estimates of health care resource item use, describe and justify the basis of the estimate, and specify the information source. Consider the applicability of the data to the modelled setting. Measure prospectively the pattern of health care resource use in the course of a clinical study by:

- retrospectively reviewing relevant records or through linking data with claims data
- administering a questionnaire or survey
- using diaries.

Distinguish between data on resource use that are directly derived from the primary evidence, and extrapolations or modelling of resource use beyond that available from the primary evidence. Justify any choice to use data that are not consistent with data from the primary evidence, particularly where this has an important impact on incremental costs, as revealed in the sensitivity analyses.

If appropriate, exclude types of health care resources that would not have a material influence on the conclusion of the economic evaluation. This may be because the cost is very small, or the cost largely cancels out between the intervention and the comparator(s). If resources are excluded for this purpose, state this and justify their exclusion, and outline how the exclusion affects the incremental cost of the intervention.

Occasionally, because of the medical condition under treatment or the age of the patients, consideration of non–health care costs such as social services (e.g. home help, day care, Meals on Wheels, private travel to access health care) or costs to other sectors might be relevant (see also Appendix 10). If incorporation of such non–health care resources is relevant for a supplementary analysis, adapt the general principles described in this TG section to generate and present these variables.

Allocation of prices (unit costs) to resources

Present all unit prices and costs in Australian dollars with a consistent year of analysis (which should be stated and be as close as possible to the submission date of the assessment report).

Section 3 adopts a broad perspective for the valuation of health care resources, so include all contributions to the costs of health care resources in the economic evaluation – including those paid for by patients, governments, health insurance agencies and any other part of society. Generally, the source of costs recommended by the Pharmaceutical Benefits Advisory Committee (PBAC) *Manual of resource items and their associated unit costs*[a] should be used. However, it is preferred that, where possible, the economic evaluation includes patient out-of-pocket costs (i.e. average charges above the schedule fee) in the unit cost of MBS-funded health technologies (including the proposed service if it is to be funded through the MBS). It is recognised that it may be difficult to obtain or estimate these costs, and sensitivity analyses should be presented. The unit cost of blood products should be derived from the National Product Price List.[b]

If there are important reasons to use unit prices other than those that are recommended, present these as a sensitivity analysis, justify each, and describe its source or generation. Ensure that any different unit price is consistent with the broad perspective of including all contributions to the costs of health care resources.

a www.pbs.gov.au/info/industry/useful-resources/manual

b www.blood.gov.au/national-product-list

Detail all alternative costs, their sources and any assumptions about them. If multiple estimates are identified, justify the estimate used in the base case and present alternative plausible estimates in sensitivity analyses.

If using historical estimates of costs, detail the information sources and the methods used to estimate them. Justify the use of the historical cost source as relevant and the best estimate available. Use the most relevant Australian price index (e.g. total health and health industry–specific price indexes published by the Australian Institute of Health and Welfare) to adjust for inflation and estimate current prices.

Explain the process of converting costs from other jurisdictions to current Australian prices.

Value future costs at current prices (i.e. do not allow for future price inflation in the calculations), consistent with using a constant price year in the economic evaluation.

Presentation of resource use and cost information

A format for summarising the minimum dataset of health care resource items and their associated unit costs relevant to the economic evaluation is suggested in Table 15. The table shows samples for each identified category. These are consistent with the PBAC *Manual of resource items and their associated unit costs*,[a] but are not comprehensive of all types of health care resource items, natural units of measurement or sources of unit costs.

Present all steps taken to calculate costs in the economic evaluation in a way that allows the calculations to be independently verified.

If a complete presentation of costs is very large, present the calculations in an accompanying technical document. Cross-reference between the calculations and the main body of the assessment report, and include an electronic version of the detailed calculations.

a www.pbs.gov.au/info/industry/useful-resources/manual

Table 15 Indicative list of health care resource items, unit costs and usage included in the economic evaluation

Type of resource item	Subtype of resource item	Natural unit of measurement	Unit cost (AUD)	Source of unit cost	Usage for the proposed health technology	Usage for the comparator
Medical services	Proposed health technology	Service rendered	x	Proposed cost of the health technology	[add usage]	[add usage]
	Comparator health technology	Service rendered	x	MBS schedule fee for item code according to current MBS, if MBS-listed service	[add usage]	[add usage]
	Other medical services	Service rendered	x	MBS schedule fee for item code according to current MBS, if MBS-listed service	[add usage]	[add usage]
Medicines	Medicine	Prescription dispensed	x	PBS dispensed price for item code according to current PBS, if PBS-listed medicine	[add usage]	[add usage]
Hospital services	Hospital admission	Episode for identified AR-DRG	x	Average cost weight for DRG item code according to current AR-DRG Public Sector Estimated Cost Weights	[add usage]	[add usage]
Residential care	ACFI category	Daily	x	Daily ACFI subsidy rate plus basic daily care fee	[add usage]	[add usage]

ACFI = Aged Care Funding Instrument; AR-DRG = Australian Refined Diagnosis Related Group; AUD = Australian dollars; DRG = Diagnosis Related Group; MBS = Medicare Benefits Schedule; PBS = Pharmaceutical Benefits Scheme

Technical Guidance 23 Model validation

KEY CONSIDERATIONS

- **Methods for validating an economic model to improve the certainty in the model results are described. Provide and comment on the validity of model traces generated by the model (TG 23.1).**
- **Describe other methods for validating the economic model (TG 23.2).**

Validation of an economic model to demonstrate that the generated results represent what they are intended to represent is best practice. It helps to reduce some of the uncertainty associated with modelling, and a more thoroughly validated model allows more confidence in its predictions.

TG 23.1 Operational validation of the economic model

Model traces for the proposed health technology and its comparator provide a clear depiction of the implications of the model. They can inform the face validity of the model logic, computerisation and external validity.

For models that include multiple indications or populations, model traces should be presented per indication/population. For investigative technologies, separate model traces for patients who do and do not have an appropriate change in management might also be informative (due to the dilution effect associated with investigative technologies, whereby the test may affect management in a subset of patients tested).

Use traces to track patients through the model and demonstrate that the logic of the model is correct. Present traces representing the proportions of the cohorts in each health state over time, and the cumulative sum of the undiscounted costs and outcomes (e.g. QALYs) over time. If applicable, state the number of events over time where patient-relevant events occur within a health state. Comment on whether each of the model traces is logical – for example, where a lifetime model is appropriate, ensure that any traces of overall survival practically converge to zero at or before the time horizon of the model (see Technical Guidance 18 and Technical Guidance 20).

Compare model traces with corresponding empirical data, where possible, to identify whether outcomes are consistent. Consider both data sources used in the model (dependent validation) and data sources not used in the model (independent validation). For example, compare predicted clinical events with observed data on the natural history of the medical condition. Comment on and explain any differences indicated by these comparisons.

In addition, compare modelled outcomes against outcomes from similar models identified in Section 3A.2.1 of the assessment report as a cross-validation tool to identify consistencies (or differences that can be explained).

TG 23.2 Other validation techniques

Present or cross-reference any other completed model validation exercises. The Assessment of the Validation Status of Health-Economic Decision Models (AdViSHE) Study Group describes a range of validation processes, and these should be considered.[62]

Technical Guidance 24 Results of the base-case economic evaluation

KEY CONSIDERATIONS

- **The presentation of the results of the economic evaluation should aim to show the impact of individual components and assumptions in the model. Several different presentations can increase this visibility.**
- **Present the intervention costs per patient (TG 24.1).**
- **Present the stepped results of the base-case analysis, incorporating translation studies and assumptions in a stepped process (TG 24.2).**
- **Present disaggregated and aggregated results of the analysis (TG 24.3).**
- **Summarise the base-case results (TG 24.4).**
- **If the health technology may be used in different scenarios (such as in a cascade testing population), present the results of additional scenarios so that the impact on the ICER of including the additional scenarios is transparent (TG 24.5).**

TG 24.1 Intervention costs per patient

Present the expected costs of the proposed health technology and comparator (individually) per patient. For therapeutic health technologies, the costs per patient per course for an acute or self-limited therapy, or per patient per year for a chronic or continuing therapy, should be reported. This estimate should be consistent with estimates of per-patient use in Section 4 of the assessment report.

TG 24.2 Stepped presentation of results

If the model translates clinical data, present the results of the key steps involved in transforming the comparative data (from Section 2) into the modelled base-case estimate of incremental cost-effectiveness.

Begin with an analysis of costs and outcomes that are directly associated with the comparative data presented in Section 2. Where the following procedures are undertaken to estimate the base case, sequentially present re-estimated costs and outcomes (and interim results) for each step:

- transformation(s) for applicability
- extrapolation of data over longer time periods
- additional data or assumptions
- transformation of clinical outcomes to final health outcomes (QALYs).

For investigative technologies, consider aligning the initial steps to sequentially incorporate evidence from each of the linkages. For example, first present costs and outcomes associated with test use based on test analytical data, then add change in management information, clinical outcomes and, finally, translated (e.g. extrapolated and/or transformed) health outcomes. Incorporate additional translations as required for applicability where relevant.

Identify the steps or assumptions of the model that have important impacts on the ICER.

Table 16 shows an example of how to present a stepped analysis incorporating evidence translations. Table 17 shows an example of steps that may be relevant where the intervention is an

investigative technology and linked evidence is required to estimate all health outcome and resource changes.

Table 16 Presentation of the stepped derivation of the base-case economic evaluation from the clinical study data

Steps (only included if undertaken)	Proposed health technology costs	Comparator costs	Incremental costs	Proposed health technology health outcomes	Comparator health outcomes	Incremental health outcomes	Incremental cost-effectiveness ratio
Comparative study data (as presented in Section 2); Setting: (trial setting); Time horizon: (trial follow-up)	[A][a]	[B][a]	[A – B]	[C] (surrogate or clinical outcome)[b]	[D] (surrogate or clinical outcome)[b]	[C – D] (surrogate or clinical outcome)	$[A – B]/[C – D] per [surrogate or clinical outcome]
Study evidence transformed from surrogate to clinical outcome (C→E, D→F)[c]	[A]	[B]	[A – B]	[E] (clinical outcome)	[F] (clinical outcome)	[E – F] (clinical outcome)	$[A – B] / [E – F] per [clinical outcome]
Study evidence transformed to clinical outcome and translated to the Australian population and/or Australian setting (may need multiple steps)	[modified A][d]	[modified B][d]	[modified A – modified B]	[modified E][e]	[modified F][e]	[modified E – modified F]	$[modified A – modified B]/[modified E – modified F] per [clinical outcome]
Study evidence transformed to clinical outcome, translated to the Australian population/setting, and extrapolated to the appropriate time horizon	[modified & extrapolated A] = [G]	[modified & extrapolated B] = [H]	[G – H]	[modified & extrapolated E] = [I]	[modified & extrapolated F] = [J]	[I – J]	$[G – H]/[I – J] per [clinical outcome]
Study evidence transformed to clinical outcome, translated to the Australian population/setting, extrapolated, and with additional assumptions or modelled information	(G + *w*) = [K][f]	(H + *x*) = [L][f]	[K – L]	(I + *y*) = [M][g]	(J + *z*) = [N][g]	[M – N]	$[K – L]/[M – N] per [clinical outcome]

Steps (only included if undertaken)	Proposed health technology costs	Comparator costs	Incremental costs	Proposed health technology health outcomes	Comparator health outcomes	Incremental health outcomes	Incremental cost-effectiveness ratio
Study evidence translated to clinical outcomes, the Australian population/setting, extrapolated, with additional modelling and transformed into a relevant health outcome (e.g. QALYs) (M→O, N→P)	K	L	[K – L]	[O]	[P]	[O – P]	$[K – L]/[O – P] per QALY

QALY = quality-adjusted life year

a If resource data are not provided, estimate resource use and apply costs (Australian $) within the study period.

b Key outcome(s) from comparative data (presented in Section 2) used to generate 'treatment effect' in the economic evaluation, without any modification (may be a surrogate, intermediate or a final clinical outcome).

c If the key outcome(s) from comparative data (presented in Section 2) are surrogate outcomes, evidence to justify the transformation of the surrogate outcome to the clinical outcome and the method used should be fully documented in Section 2.

d Include here any modelled changes in the provision of resources that would occur in the Australian health care setting.

e Include here any transformations to estimated outcomes to increase applicability to the Australian population or setting.

f Re-estimate of costs after including additional data or assumptions (*w* and *x*) that were not captured in the key comparative clinical data (e.g. adverse events or second-line treatments).

g Re-estimate of outcomes after including additional data or assumptions (*y* and *z*) that were not captured in the key comparative clinical data (e.g. adverse events or second-line treatments).

Table 17 Presentation of the stepped derivation of the base-case economic evaluation, investigative technology example

Steps (only included if undertaken)	Proposed health technology costs	Comparator costs	Incremental costs	Proposed health technology health outcomes	Comparator health outcomes	Incremental health outcomes	Incremental cost-effectiveness ratio
Comparative diagnostic accuracy, as applied to the prevalence in the eligible Australian population[a] Time horizon: time to reach a diagnosis	Cost of the proposed test [A]	Cost of the comparator service [B]	[A – B]	TP: [ ]%, FP: [ ]%, TN: [ ]%, FN: [ ]% Total correct diagnoses: [C]%	TP: [ ]%, FP: [ ]%, TN: [ ]%, FN: [ ]% Total correct diagnoses: [D]%	[C – D] (correct diagnosis)	$[A – B]/[C – D] per [correct diagnosis]
Incorporation of repeat or confirmatory testing, which may affect final diagnostic conclusions Time horizon: time to reach a diagnosis	Include cost of additional testing, resampling, where relevant [E]	Include cost of additional testing, resampling, where relevant [F]	[E – F]	Total correct final diagnoses: [G]%	Total correct final diagnoses: [H]%	[G – H] (correct final diagnosis)	$[E – F]/[G – H] per [correct final diagnosis]
Uptake of treatment, or other change in clinical management, by final test result Time horizon: time to treatment allocation decision	Include cost of treatment [I]	Include cost of treatment [J]	[I – J]	Correct treatment allocation: [K]%	Correct treatment allocation: [L]%	[K – L]	$[I – J]/[K – L] per [correct treatment allocation]

Steps (only included if undertaken)	Proposed health technology costs	Comparator costs	Incremental costs	Proposed health technology health outcomes	Comparator health outcomes	Incremental health outcomes	Incremental cost-effectiveness ratio
Incorporation of effectiveness of treatment (e.g. survival benefit) translated to the Australian population and/or setting, and extrapolated to the appropriate time horizon (may need multiple steps) Time horizon: appropriate time horizon to capture differences in costs and outcomes due to changes in treatment allocation decisions (e.g. lifetime)	Include costs due to time horizon extension (e.g. disease progression or management) [M]	Include costs due to time horizon extension (e.g. disease progression or management) [N]	[M – N]	Life years gained [O]	Life years gained [P]	[O – P]	$[M – N]/[O – P] per [life year gained]
Outcomes transformed into a relevant health outcome (e.g. QALYs) (O→Q, P→R)	M	N	[M – N]	[Q]	[R]	[Q – R]	$[M – N]/[Q – R] per QALY

FN = false negative; FP = false positive; QALY = quality-adjusted life year; TN = true negative; TP = true positive

a Trial-based accuracy and prevalence estimates could be presented as a prior first step, and then translated to the proposed setting (i.e. most applicable estimates of accuracy and prevalence in the proposed setting).

The order of the steps for the translation of the trial-based economic evaluation may vary.

The final row of Table 16 and Table 17 incorporates all translation studies and additional modelling to complete the impacts of translating the trial-based economic evaluation into a modelled economic evaluation. Ensure that this corresponds to the base-case ICER.

The stepped presentation informs the face validity of the results, and identifies assumptions and approaches to be examined in more detail in sensitivity analyses. For example, if the main impact is achieved by extrapolating the final outcome over time, undertake comprehensive sensitivity analyses around the extrapolation methods.

Present the base-case incremental cost, incremental effectiveness and ICER (calculated as the incremental costs divided by the incremental health outcomes).

TG 24.3 Disaggregated and aggregated base-case results

If a decision-tree model is used, present a detailed disaggregation of costs incurred at each branch by resource type for the intervention and comparator groups. For state transition models, present disaggregated discounted costs by resource type for each health state for the intervention and comparator groups. In all models, report the proportions of patients predicted to experience alternative target clinical outcomes in the intervention and comparator groups.

Alternative examples of tables showing disaggregated costs are provided in Table 18 and Table 19.

Table 18 Health care resource items: disaggregated summary of cost impacts included in the economic evaluation

Type of resource item	Subtype of resource item	Costs[a] for proposed health technology	Costs[a] for main comparator	Incremental cost[a]	% of total incremental cost[a]
Medical services	Type of medical service				
	Health state 1	$x1	$y1	$x1 – $y1	z1%
	Health state 2	$x2	$y2	$x2 – $y2	z2%
	[etc]	$xk	$yk	$xk – $yk	zk%
	Total	∑$x	∑$y	∑$x – ∑$y	∑z%
Medicines	PBS medicine				
	Health state 1				
	Health state 2				
	[etc]				
	Total				
	Non-PBS medicine				
	Health state 1				
	Health state 2				
	[etc]				
	Total				
Hospital services	Hospital admission				
	Health state 1				
	Health state 2				
	[etc]				
	Total				
Residential care	ACFI category	$x	$y	$x – $y	z%
	Total	$x	A\$y	$x – $y	100%

ACFI = Aged Care Funding Instrument; PBS = Pharmaceutical Benefits Scheme

a Indicate clearly whether cost values are discounted costs (use of discounted costs is appropriate).

Table 19 Health states: disaggregated summary of cost impacts included in the economic evaluation

Health state in model	Resource use by health state (modelled)	Costs for proposed health technology	Costs for main comparator	Incremental cost	% of total incremental cost
Health state 1	Resource type 1	$x1	$y1	$x1 – $y1	z1%
	Resource type 2	$x2	$y2	$x2 – $y2	z2%
	[etc]	$x etc	$y etc	$x etc – $y etc	z etc
	Total for health state 1	∑$x	∑$y	∑$x – ∑$y	∑z%
Health state 2	Resource type 1	$xx1	$yy1	$xx1 – $yy1	zz1%
	Resource type k	$xxk	$yyk	$xxk – $yyk	zzk%
	Total for health state 2	∑$xx	∑$yy	∑$xx – ∑$yy	∑zz
[etc]	[etc]	[etc]	[etc]	[etc]	[etc]
Total	–	∑$x + ∑$xx etc	∑$y + ∑$yy etc	(∑$x + ∑$xx etc) – (∑$y + ∑$yy etc)	100%

– = not required

Similarly, an example of a table showing outcomes disaggregated by health state is given in Table 20.

Table 20 Health states: disaggregated summary of health outcomes included in the economic evaluation

Health state in model	Outcome for proposed health technology	Outcome for main comparator	Incremental outcome	% of total incremental outcome
Health state 1	x1	y1	x1 – y1	z1%
Health state 2	x2	y2	x2 – y2	z2%
[etc]	[x etc]	[y etc]	[x etc – y etc]	[z etc]
Total	x	y	x – y	100%

Identify which health states and resources contribute to the greatest incremental differences between the proposed health technology and the comparator.

TG 24.4 Summary of base-case results

Summarise the base-case estimate of the incremental outcome(s), incremental cost and the cost-effectiveness ratio(s) obtained in the economic evaluation(s), including both CUA and CEA where relevant. For multi-indication models, the results of the CEA should be presented across the total proposed population eligible for the health technology. If indications can be reasonably excluded from the population eligible for the proposed health technology, results of the CEA disaggregated by indication can be presented as alternative listing scenarios (see TG 24.5).

Comment on whether there is likely to be bias in the base-case estimate of the ICER (e.g. an over- or underestimate of costs or outcomes, that was identifiable but not quantifiable) and the likely overall direction of that potential bias.

If the ICER is based on an outcome other than life years or QALYs gained, summarise any other health outcome effects (benefits or harms) that are associated with the intervention but not captured in the outcome (and which may not be quantifiable). If additional health outcomes effects can be estimated, present a summary of relevant health outcomes in the format of a CCA. Compare the presented results with any previous MSAC decisions based on the same measure of outcome.

TG 24.5 Alternative listing scenarios

Present alternative listing scenarios. For multi-indication models, this may include presenting the results of the economic analysis disaggregated by indication, or excluding indications.

For genetic testing for heritable conditions, alternative listing scenarios should be presented where testing is expanded incrementally from index cases only, through to index cases plus first-degree biological relatives, index cases plus first- and second-degree biological relatives and potentially index cases plus first-, second- and third-degree biological relatives. The marginal cost-effectiveness of expanding the populations eligible for the test should also be presented.

For plausible alternative listing scenarios, key sensitivity analyses (Technical Guidance 25) should be presented in Section 3A.9 of the assessment report.

Technical Guidance 25 Uncertainty analysis: model inputs, structure and assumptions

KEY CONSIDERATIONS

- **Careful identification and testing of sources of uncertainty in the model is important to increase confidence in the model results.**
- **Identify parameter, translational and structural souces of uncertainty (TG 25.1).**
- **Present sensitivity and scenario analyses using appropriate alternative inputs or assumptions to explore the impact of the uncertainty on the model results (TG 25.2).**
- **Present relevant multivariate sensitivity analyses (TG 25.3).**
- **Summarise the results of the sensitivity analyses (TG 25.4).**

TG 25.1 Identifying and defining uncertainty in the model

Present univariate deterministic sensitivity analyses for all input parameters, or natural groups of input parameters (e.g. cost or utility weights for all target clinical outcomes) using plausible alternatives. The following requests are based on good-practice guidelines for model parameter estimation and uncertainty analysis.[72]

Parameter uncertainty

Use commonly adopted statistical standards to represent the uncertainty around the true value of each uncertain input parameter. For example, beta distributions are a natural match for transition probabilities; log-normal for relative risks or hazard ratios; logistic distributions to calculate odds ratios; and gamma or log-normal for costs and utility parameters.

Justify using alternative distributions. Use interval estimates (e.g. 95% confidence intervals) derived from fitted probability distributions to define the ranges of the parameter values tested in the deterministic sensitivity analyses.

Where there is very little information on a parameter, adopt a conservative approach by defining a broad range of possible parameter values. Never exclude parameters from uncertainty analysis on the grounds that there is insufficient information to estimate uncertainty.

Consider correlation between input parameter values. If applicable, represent the joint uncertainty around the true values of 2 or more input parameters in the uncertainty analyses. It is preferable to represent the joint uncertainty around transition probabilities in the intervention group and the comparator group by applying a relative treatment effect parameter. If a relative treatment effect parameter is not applicable, individual-level data for the comparator and intervention could be bootstrapped to provide more realistic estimates of the joint uncertainty between these.[72]

The joint estimation of multiple input parameters when using regression analysis produces relevant correlation parameters. Otherwise, model calibration methods may be used to represent joint uncertainty around the true value of model input parameters.

Translational uncertainty

Where clinical data have required translation for applicability, transformation or extrapolation for incorporation into the model, systematically consider the assumptions used in the translation and identify any uncertainty in these assumptions. Identify plausible alternatives for testing in scenario analyses.

Examples of analyses that can be used when data or outcome translations are incorporated into the base-case analysis are presented in Table 21.

Table 21 Examples of potential sources of translational uncertainty in the economic model and suggested scenario analyses

Translations incorporated into base-case analysis	Suggested uncertainty analysis
Transformation of continuous outcome data to a dichotomous outcome	Alternative thresholds
Treatment effect with adjustment for switching	Treatment effect without adjustment for switching, and/or using an alternative adjustment technique
Treatment effect based on translation (e.g. subgroup analysis) following applicability study	Treatment effect based on intention-to-treat population
Treatment effect from selected source(s) of data	Alternative available source(s) of data, and/or meta-analysis of data as source of treatment effect
Transformation of a surrogate to a final outcome	Range of alternative plausible values (as derived from establishing STFO relationship)
Extrapolation of data beyond the trial	Alternative data truncation point(s), alternative choices of parametric model, or alternative assumptions regarding ongoing treatment effect
Utility values estimated from pooled within-trial data (or alternative approach)	Estimates based on individual arms (or the alternative approach)
Externally sourced utility values	Alternative values or sources

STFO = surrogate to final outcome

Structural uncertainty

If multiple plausible choices or assumptions are identified for the model structure, assess and present the potential impact of these on the model outputs. If a substantial impact is predicted, use a formal approach to characterise the structural uncertainty. Use scenario analyses to assess the impact of assumptions regarding the structure of the economic model, including alternative model structures identified in response to Technical Guidance 18, or alternative assumptions regarding the duration of the treatment effect or the choice of parametric model to extrapolate survival data. Report the results of each set of plausible structural assumptions. Alternatively, set parameters on structural assumptions where there is sufficient clinical evidence or expert opinion to do so.

Describe and justify the inclusion and exclusion of potential scenario analyses when making alternative assumptions about aspects of the model structure.

Include an analysis of the impact of the time horizon.

Use other scenario analyses to assess the effects of substantial use of the proposed health technology beyond the intended population and circumstances of use defined in the requested restriction. A wider population or circumstances would be expected to have different demographic, patient and usage characteristics than the target population and circumstances.

TG 25.2 Univariate sensitivity and scenario analyses

Tabulate all parameter values and assumptions included in the model, and present the results of univariate sensitivity and scenario analyses in a similar format to Table 22.

Use a tornado diagram to represent the relative effect of the uncertainty around alternative input parameters on the base-case incremental cost-effectiveness result.

Identify the input parameters and model assumptions to which the incremental cost-effectiveness results are most sensitive.

TG 25.3 Multivariate and probabilistic sensitivity analyses

Use multivariate sensitivity analyses to test the combined effects of uncertainty around the true values of input parameters. Perform the analyses using the parameters that the base-case incremental cost-effectiveness result was sensitive to in the univariate analyses. If the univariate analyses identify multiple parameters for testing in a multivariate analysis, consider incorporating changes in a stepped manner to allow MSAC to see the impact of each change on the resulting ICER.

Describe the multivariate sensitivity analyses to be undertaken, and present the results. Justify the inclusion and exclusion of parameters in these analyses.

A probabilistic sensitivity analysis (PSA) may be provided in addition to a deterministic sensitivity analysis. However, as translational and structural uncertainties have previously been more influential in MSAC deliberations than uncertainty regarding the precision of parameter estimates, multivariate analyses incorporating any translational and structural uncertainties should be prioritised above conducting and presenting a PSA.

If undertaking a PSA on a cohort-based state transition model, the number of iterations (sets of randomly sampled input parameter values included in the analysis) should provide stability in the model outputs across multiple analyses using alternative random number seeds. Provide the random seed associated with the presented results to enable replication, and also ensure that the model permits alternative seeds.

If undertaking a PSA on an individual-level model (e.g. a discrete event simulation), the number of iterations may be selected to balance stability of model outputs and a reasonable time required to undertake a PSA (e.g. a few hours, rather than a few days).

Use cost-effectiveness planes and acceptability curves to present the results of a PSA, as well as tabulating the interval estimates for the ICER or the incremental net benefits of the proposed health technology.

TG 25.4 Summary of the uncertainty analysis

Describe and justify a likely range of values within which the true estimate of the incremental cost-effectiveness of the proposed health technology is likely to lie, identifying the key sources of uncertainty. This range may be informed by a formal PSA, or by subjective interpretation of the presented deterministic sensitivity and scenario analyses.

Discuss the implications of the sensitivity and scenario analyses with respect to the certainty of the base-case ICER estimate. Discuss the likely overall effect of deficiencies in the evidence base on the reported cost-effectiveness of the proposed health technology.

Table 22 Results of the sensitivity and scenario analyses characterising the uncertainty around the ICER

Variable or assumption	Base-case value	Plausible alternative(s) or range of values	Incremental outcomes	Incremental costs	ICER	Description of impact on ICER
Base case			[base case]	[base case]	[base case]	
Discounting rate	Outcomes and costs = 5%	Outcomes and costs = 3.5% Outcomes and costs = 0%	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Plausible range of treatment effect, if modelled as a variable (e.g. hazard ratio or relative risk)	[add]	[e.g. upper and lower 95% confidence intervals around estimate]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Altered patient characteristics, if relevant	[add]	[e.g. different average age, disease or condition severity]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Transition or event probabilities	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Outcome-related assumptions or variables [Recommended examples: • alternative methods or sources of utility weights]	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Cost-related assumptions or variables	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Alternative extrapolation variables or assumptions [Recommended examples: • start point • choice of parametric model • assumption regarding ongoing treatment effect]	[e.g. maximum follow-up]	[e.g. median follow-up]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Any other translation assumptions [e.g. use of intention-to-treat/non-adjusted data]	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Alternative assumptions regarding model structure	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]

Variable or assumption	Base-case value	Plausible alternative(s) or range of values	Incremental outcomes	Incremental costs	ICER	Description of impact on ICER
Time horizon	[add]	[e.g. trial-based; 5, 10, 20 years, as appropriate]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]
Plausible alternatives for other variables or assumptions [e.g. including leakage beyond the requested restriction]	[add]	[add]	[alternative estimates]	[alternative estimates]	[alternative estimates]	[describe as required]

ICER = incremental cost-effectiveness ratio

Section 3B Cost minimisation

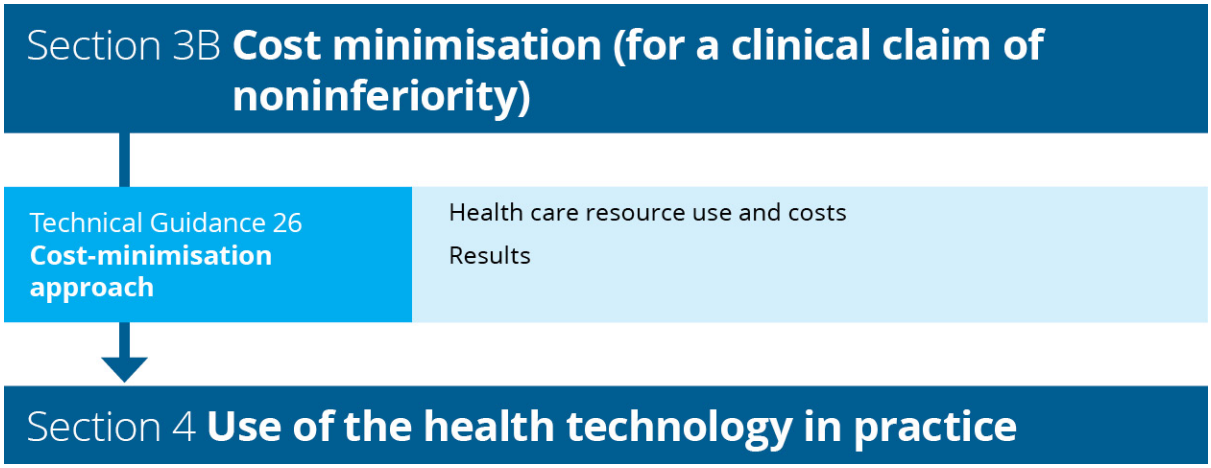
This section provides guidance on preparing Section 3 using a cost-minimisation approach (see Section 3, Introduction). Flowchart 7 shows the structure of the guidance in this section.

The assumption of noninferiority, with respect to both effectiveness and safety, needs to be well justified for the cost-minimisation approach to be accepted. Irrespective of the therapeutic claim, if the adverse effect profiles of a proposed health technology and its main comparator are significantly different, it is unlikely that the cost-minimisation approach will suffice. The implications of these differences, for both health outcomes (ideally, utility) and resource use, should be explored in a full economic evaluation.

The cost-minimisation approach has an abbreviated Section 3 that identifies differences between the proposed health technology and the comparator that are likely to result in a difference in health resource use. This includes identifying differences in:

- the costs of prescribing or administering the services
- the costs of monitoring or managing associated adverse events
- anything else that may impact health resource use.

A cost analysis compares costs only. It is strictly defined as a partial rather than a full economic evaluation, because it does not quantitatively assess comparative costs in a ratio over comparative effectiveness. Although less preferred than a full economic evaluation, cost analyses have sometimes been presented and found to be acceptable if the proposed health technology is demonstrated to be no worse in terms of effectiveness but has a superior safety profile compared with the main comparator.



Flowchart 7 Structure of the guidance for a cost-minimisation approach

Technical Guidance 26 Cost-minimisation approach

KEY CONSIDERATIONS

- **Identify the relevant costs (direct health technology costs and other relevant costs) (TG 26.1). Provide and justify the sources of data used.**
- **Report the results of the cost-minimisation approach (TG 26.2).**

TG 26.1 Health care resource use and costs

Direct health technology costs

Using guidance in Technical Guidance 22, estimate the direct health technology costs per patient. For therapeutic health technologies, the costs estimated should be per patient per course for an acute or self-limited therapy, or per patient per year for a chronic or continuing therapy. Use of the intervention and comparator therapies is generally derived from the clinical studies reported in Section 2.

For investigative technologies, it would generally be sufficient to cost the health technologies in each arm to the point of diagnosis. It would be difficult to justify a cost-minimisation approach assuming final health outcomes were equivalent if the analytical test outcomes/diagnostic outcomes were not also equivalent.

Additional costs and/or cost offsets

The nature of additional costs and/or cost offsets will differ in each MSAC application. Two common areas for additional costs or offsets are costs associated with prescribing or administration, and costs of managing adverse events; however, this does not preclude other possible cost offsets. Additional costs or offsets could also arise from subsequent changes in resource use due to changes in management (e.g. further downstream testing) that result from investigative technologies, provided these do not impact downstream costs and final health outcomes. Justify any other additional costs and/or cost offsets in terms of how they are realisable and/or patient relevant, and show how they differ between the options being considered in the cost-minimisation analysis.

Comparison of prescribing and administration profiles

Identify differences in the costs of prescribing or administering the proposed health technology and the comparator.

Listing a noninferior health technology might have cost consequences related to its differing mode of administration. These have sometimes arisen if the proposed health technology and its main comparator are available in different forms. If this applies, identify the types of other resources affected, estimate the extent to which the quantity of each type of resource provided (in its natural units of measurement) would change following a listing, and multiply by the relevant unit costs. Aggregate this with the health technology cost impact to estimate the net cost impact within the cost-minimisation analysis.

See also the PBAC *Manual of resource items and their associated unit costs*[a] for further detail on costing administration-related resource use.

a www.pbs.gov.au/info/industry/useful-resources/manual

Comparison of safety management profiles

Only use the cost-minimisation approach where the proposed health technology has a safety profile that is superior (preferably) or noninferior to the main comparator.

Identify any differences in the costs of monitoring or managing adverse events associated with the proposed health technology and the comparator.

If the proposed health technology is demonstrated to be no worse in terms of effectiveness, but has a superior safety profile to the main comparator, a price advantage for the proposed health technology over its main comparator could be sought on the basis of cost offsets due to reduced costs of monitoring for, or managing, adverse reactions. Provide evidence to support a claim that monitoring costs are reduced.

Where safety profiles are similar, but the proposed health technology has a reduced magnitude of adverse effects (severity or incidence), present a thorough description of the quantified differences in safety, with a justified estimate of any corresponding implications for resource use.

Where the adverse effect profiles of a proposed health technology and its main comparator are different in nature, a CEA or CUA is likely to be preferred (see Section 3A). However, a cost analysis may be acceptable to quantify a claim that the cost offsets from the reduction in health care resources required to treat the adverse events are sufficient to reduce the incremental cost to zero or a negative value.

See also the PBAC *Manual of resource items and their associated unit costs*[a] for further detail on resource use and costing associated with monitoring and managing adverse effects.

TG 26.2 Results

Results of the cost-minimisation approach

List all identified costs associated with both the proposed health technology and the comparator to estimate the net cost difference.

The economic claim should be that, at the price requested, the overall cost of therapy with the proposed health technology is the same as, or less than, the overall cost of therapy with the main comparator.

Sources of data

Provide copies of the original sources of all data (beyond those already presented in Section 2) or expert opinion used in the model in an attachment or technical document. Cross-reference data extracted from each source to the level of the page, table or figure number of the source document.

To enable independent verification of each analysis, provide an electronic copy of any computer-based calculations of the analysis.

a www.pbs.gov.au/info/industry/useful-resources/manual

Section 4 Use of the health technology in practice

Introduction

Section 4 of the assessment report should present a set of budget impact analyses, and provide the most likely extent of use and financial impact of the proposed health technology. These analyses are relevant to MSAC, the Australian Government and other committees and funding bodies that refer to MSAC. Section 4 is important for estimating the likely uptake of the proposed health technology in clinical practice and the financial impact of the service on the relevant funding program and the Australian Government budget. Depending on the funding context, this may also be used to negotiate risk-sharing arrangements.

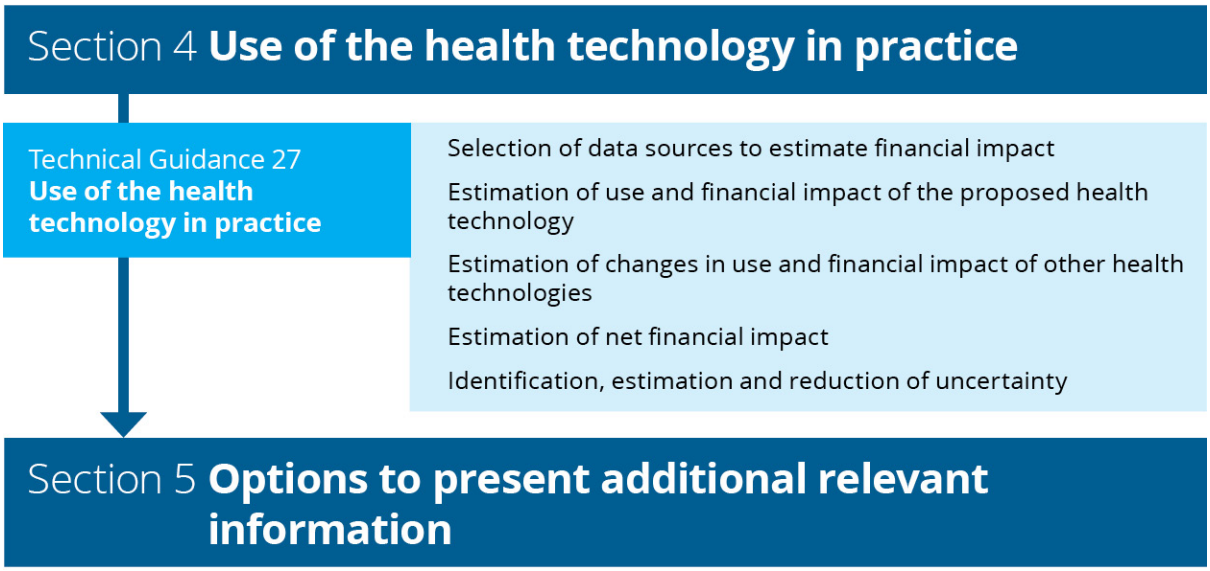
Flowchart 8 shows the structure of the guidance provided in this section for estimating likely use and financial impact of the proposed health technology.

The 2 broad approaches for developing utilisation and financial estimates are epidemiological and market-share analyses, although their use is not mutually exclusive. An epidemiological approach is usually preferred for generating utilisation and financial estimates if the assessment report includes a clinical claim of superiority to the comparator. However, a market-share approach might be preferred if the assessment report includes a clinical claim that the technology is noninferior.

The approach taken for estimating use and financial impact should be justified in the assessment report. Where data inputs from one approach (epidemiological or market share) are uncertain, concordance across both approaches should be demonstrated.

Ensure that any estimates of the extent of use of the proposed health technology (and other technologies affected by the listing of the proposed technology) in the Australian setting are consistent with evidence presented throughout the assessment report. Ensure that uptake of the health technology, change in the use of alternative health technologies and offsets are all consistent with the clinical place of the technology (Section 1), the use of the technology in the clinical evidence (where applicable; Section 2) and the circumstances presented in the economic evaluation (Section 3). Any discrepancies should be explained and justified.

Provide sufficient data in Section 4 so that the steps in the analyses can be interpreted. Where the calculations used to generate estimates are not transparent in the main body of the assessment report, present additional data.



Flowchart 8 **Structure of the guidance for estimating the use and financial impact of the proposed health technology in practice**

Specification of the relevant funding program

It is within MSAC's remit to consider assessment reports for health technologies that are funded through different funding programs, so the relevant funding program for the proposed health technology should be identified. The utilisation of the proposed health technology and the financial impact on the relevant funding program should be presented in Section 4.2 of the assessment report. Any consequential utilisation and financial changes to other items funded by the same funding program should be presented in Section 4.3. In Section 4.4, the estimated net financial impact on the relevant funding program should be reported. Changes in the utilisation and financial estimates of other health technologies funded by Australian Government health programs should be reported in Section 4.5.

In contrast to the economic evaluation presented in Section 3 of the assessment report, these financial analyses exclude health outcomes, do not use discounting, and exclude any resource item or copayment from a source other than the budget identified in Section 4.4. However, the financial impact on other Australian Government budgets can be presented in Section 4.5.

Epidemiological approach

An epidemiological approach estimates the number of people with the medical condition, and then estimates the use of the proposed health technology (see TG 27.2). It also estimates any consequential changes in use of other services (see TG 27.3) in the context of the patient group defined by the proposed item descriptor.

An epidemiological approach estimates the patients eligible for the proposed health technology; however, market-based data or market research may be required to establish estimates such as the rate of uptake of the health technology.

Market-share approach

The market-share approach estimates the extent of the current market represented by the proposed patient indication and, consequently, the share likely to be taken by the proposed health technology. It is likely to be the most suitable approach when the proposed health technology will completely substitute existing MBS-listed services.

In contrast with the epidemiological approach, the market-share approach allows an abbreviated presentation of information if there is an expectation of no market growth following listing. It also provides an alternative way of generating estimates to compare with the epidemiological approach.

The key issue with estimates built on the market-share approach is whether listing the proposed health technology on the MBS is likely to increase the current market or market growth rate. If not, a health technology listed on a cost-minimisation basis would usually have a negligible effect on the net financial impact on the MBS; however, it may have financial impacts on other parts of the Australian Government health budget. If the proposed health technology is likely to increase the market size or its growth rate, it is critical to estimate the extent of this likely increase.

Fully editable electronic copy of the financial impact analysis

The financial impact analysis should be constructed in an Excel workbook and provided with the assessment report to allow an independent assessment of the data. Ensure that the responses in Section 4 and the Excel workbook cross-reference each source that data were extracted from to generate the estimates in the analysis (to the level of the page, table or figure number of each

source document). Where commissioned data have been used, include the correspondence for the data request.

Ensure that the calculations flow through the spreadsheets, so that changes to any variable flow on to the results. To help understand the spreadsheets, apply clear and unambiguous labels to spreadsheet values, and cross-reference the data source. Where relevant, complex analyses or supporting data should be presented in separate spreadsheets. Provide clear and consistent formulas in the spreadsheets, to facilitate tracing and replicating the calculation flow.

Throughout Section 4, refer to the relevant spreadsheet number/title. Describe the approach, methods, assumptions and potential biases. Where possible, add comments to the Excel workbook to describe these factors, particularly if the approach is complex. Confidence in the estimates is reduced if the interpretation of calculations in the Excel workbook cannot be reconciled with the relevant assumptions or approach.

Technical Guidance 27 Use of the health technology in practice

KEY CONSIDERATIONS

- **If multiple sources of data are required to estimate the financial impact of the proposed health technology, discuss the concordance across these sources and present sensitivity analyses for the different estimates across the sources (TG 27.1).**
- **Describe the data and data source and their purpose in the analysis, as well as how the data (especially any international data) are relevant to the Australian setting. Discuss the methods used to derive each estimate from the source data, as well as any assumptions, limitations and biases in the approach taken (TG 27.1).**
- **Discuss the use and financial impact of the proposed health technology, changes in use and financial impact of other health technologies, and the likely impact of any uncertainty in the financial estimates (TG 27.2–27.5).**

TG 27.1 Selection of data sources used to estimate the financial impact of the proposed health technology

Available data sources

Data sources fall under the broad headings listed in Table 23; however, there might be other suitable data sources (some examples may be listed on the Pharmaceutical Benefits Scheme (PBS) website: *Sources of data for use in generating utilisation estimates*[a]).

For the market-share approach, relevant sources of data include MBS data, including those supplied by Services Australia relating to the MBS rebates paid and patient out-of-pocket costs, or data collected through the relevant funding program.

a www.pbs.gov.au/info/industry/useful-resources/sources

Table 23 Categories of data sources

Data type	Examples
Disease or condition epidemiological data (provide estimates of prevalence or incidence in the population)	Australian case or mortality registers that estimate the incidence or prevalence of a disease or condition Large, well-designed Australian studies that estimate the incidence or prevalence of a disease or condition Australian national health surveys that estimate the prevalence of a disease or condition Utilisation databases, including MBS data for other services in the proposed population, or state-based utilisation data where the proposed health technology is already in use
Market data	Quantitative description of the existing market, including estimates of change in the size of the market over time Estimates of relative market shares Estimates of the impact of the requested MBS listing on current treatment paradigms, based on similar previous listings
Commissioned data	Data requests to registries, epidemiological studies or utilisation studies Epidemiological studies

MBS = Medicare Benefits Schedule

Multiple sources of data may be required. In Section 4.1 of the assessment report:

- describe the data and data source
- explain the purpose of the data in the analysis
- describe how the data are relevant to the present Australian setting. Where data on overseas markets are provided, clearly state that Australian data were not available and discuss the applicability of international data to the Australian setting (with particular reference to the subsidy arrangements in the overseas jurisdiction)
- where there are multiple sources of data, discuss the concordance across these sources and present sensitivity analyses for the different estimates across the sources
- for each estimate derived from source data, summarise the methods used, and discuss any assumptions, limitations and biases in the approach taken.

Commissioned data

A commissioned study may be used to fill a gap in the data, and may include health technology usage surveys, data from disease or condition registries, or claims data. Clearly state the original purpose for the collection (e.g. the data were collected for the primary purpose of understanding treatment choices). When reporting the results of commissioned data, provide sufficient background and methodological information to adequately interpret the results.

See Appendix 9 for further guidance on presenting commissioned data from a survey of experts. Provide the method for identifying experts, the reasons for collecting the information, and any potential conflicts of interest of the respondents or the organisation undertaking the survey. Present the actual questions asked and the range of responses. Where the respondents are experts in treating specific diseases, provide an estimate of the number of patients they treat, what proportion this is of the expected number of patients in Australia, and the health area and setting in which the

respondents practise (e.g. public hospital, private hospital, community, regional area, inner urban area).

When analysing administrative data and registries, provide sufficient information about the method used to sample the dataset, the proportion of the affected population included in the dataset, rules for analysis, assumptions used (particularly where elements in the dataset are used as surrogates) and statistical methods (such as censoring or use of propensity scores).

TG 27.2 Estimation of use and financial impact of the proposed health technology

Justify any estimates of the incidence or prevalence of the disease or condition, or estimates of market growth over 6 years. Multiple factors may influence market growth, and it may not be appropriate to assume linear growth in the estimates, particularly if the proposed health technology is not the first entrant to the market for the specific indication. It is important to base projections on the number of patients, not services provided, wherever possible.

Epidemiological approach

Incidence or prevalence data

For an epidemiological approach, present the methods and assumptions for converting incidence or prevalence data to the number of patients likely to take up the proposed health technology each year.

The choice to use incidence or prevalence data depends on several factors, including the nature of the medical condition, its treatment and the available data. In general, treatments of short duration are best suited to incidence estimates, and long-term treatments (e.g. for chronic diseases or conditions) may be better suited to prevalence estimates. A combination of prevalence and incidence estimates may be required (e.g. intermittent treatments for a chronic condition).

Consider the current prevalent patient population in addition to the incident population – for example, a cancer therapy where there are patients receiving best supportive care before the proposed health technology becomes available. Only calculating the incident population would underestimate the likely number of patients treated in the early years of listing.

Estimate the number of patients with the medical condition

Estimate the likely number of patients in the 6 years following listing, using the incidence or prevalence approach, accounting for trends in disease or condition incidence or prevalence. If appropriate, present shorter periods (e.g. monthly or quarterly) in supporting spreadsheets and summarise annually for 6 years from listing. If using an incidence approach, also estimate the prevalent population (from years before listing) that may add to the eligible patient pool in year 1. Justify when the addition of a prevalent population is not required.

If the medical condition has a subjective element in its diagnosis, consider the impact of misdiagnosis that may render patients eligible for the proposed health technology.

Estimate the number of patients eligible for the proposed health technology

Using the annual numbers of patients with the medical condition for 6 years, estimate the proportions of patients in each year who would be expected to be eligible for the proposed health technology according to the proposed eligibility criteria.

Where the proposed eligibility criteria contain subjective elements, consider the risk of misclassification increasing the number of patients eligible for the proposed health technology.

Estimate the number of patients likely to use the proposed health technology

Using the annual numbers of eligible patients, estimate the proportions likely to take up the proposed health technology in each of the 6 years. Ensure that the estimates reflect the rate of uptake of the proposed service, and consider the impact of the use of other services or treatment options. For proposed MBS services, uptake should further be considered by the setting of use (i.e. private sector or public hospital sector). Analyses should account for billing of the MBS by public hospitals, where relevant.

Consider whether there are differences in out-of-pocket costs associated with the proposed health technology that may influence the rate of uptake.

Justify the estimate of uptake and assess variations to this estimate in sensitivity analyses.

Number of times the proposed health technology is delivered

The estimate of the number of services provided for each of the 6 years should account for, where applicable:

- the rate of uptake of the proposed health technology across the 6 years from listing (described previously)
- the number and frequency of use of the proposed health technology per patient.

Present each of the steps for estimating the number of units dispensed separately.

Market-share approach

Describe the market

To generate estimates of expected utilisation and costs, ensure that the market-share approach relies on health technology utilisation data or studies for currently available services that are likely to be substituted. This is the basis for predicting whether the market will change because of listing the proposed health technology.

Number of services provided by currently listed items

Estimate the units dispensed in the most recent 12 months of the relevant market.

Where possible, present the services provided *and* the number of patients this represents according to the evidence provided in Section 2. This will be particularly important where a market-share approach is being compared or used in conjunction with an epidemiological approach. However, if the number of services per patient per course of treatment is uncertain, do not back-calculate to estimate the number of patients, as it can introduce significant errors into the patient numbers.

Estimate the rate of growth in the market over 6 years following listing. Base this on historical trends in the market or other influences, but ensure that it is unrelated to the listing of the proposed health technology. Justify the estimate of market growth in the absence of the listing of the proposed service.

Where more than one service within the funding program is likely to be substituted, present the market share and rate of growth for each item, if required. Disaggregating the estimated growth according to each service is important if they are likely to have different rates of growth, are likely to be substituted differentially by the proposed health technology or have a different cost.

Estimate the market share

Estimate the rate of substitution in the market by the proposed health technology for each year of the 6 years following listing. Provide evidence, such as market uptake rates from other markets and the applicability of these markets to the Australian setting, to justify the estimate of market share. Clearly communicate and justify the likely extent of market uptake following listing of the proposed service.

Present a table in the assessment report for overall estimates, if appropriate. Also present a table in the Excel workbook, stratified by the individual health technologies likely to be substituted, and clearly show the steps for aggregating the data. Ensure that the proportions of each health technology likely to be substituted by the proposed service are clear on the spreadsheet.

Estimate the growth of the market after listing

Use historical data to estimate the number of units dispensed for the proposed health technology above the growth projected in the market, for each year of the 6 years following listing. Report both the expected increase in patient numbers, and the expected number of services of the proposed health technology.

Justify when no additional growth in the market is predicted. When the proposed service may be used in clinical practice to treat people who are intolerant to an existing listed service, or following failure with that service, it is likely that entry of the proposed service into the market will increase the overall number of people treated.

To increase the certainty in the estimated financial impact of listing the proposed service, provide references to data for similar circumstances in similar markets, and discuss risks associated with market growth.

Financial impact over 6 years

Present the total estimated financial impact of listing the proposed health technology with appropriate patient copayments subtracted. For proposed MBS items, the level of the MBS benefit paid depends on a number of factors:

- whether the service is provided as part of an episode of hospital or hospital-substitute treatment
- whether patients are bulk-billed
- capping of the patient copayment at the Greatest Permissible Gap amount for high-cost outpatient services (which means the proportion covered by the MBS increases with the cost of the service).

Ensure that these factors are considered, where important, in the estimated financial impact of listing the proposed health technology. If the proportion of services that attract the different levels of MBS benefit (i.e. 75% or 85%) are known, or if there is adequate justification to apply only one rebate, these proportions can be used. However, if the proportions are unknown, a pragmatic approach assuming the 80% level of MBS benefit may be used.

TG 27.3 Estimation of changes in use and financial impact of other health technologies

Identify health technologies likely to be affected

If using a market-share approach, services funded under the same program that are likely to be substituted will have been identified in Section 4.2 of the assessment report. However, identifying

other affected services within the program that will increase or decrease in usage may still be relevant.

Health technologies funded within the same program that are likely to be affected by the listing of the proposed health technology include:

- health technologies substituted by the proposed health technology
- other health technologies with decreased usage
- other health technologies with increased usage.

List all health technologies that fall into each of these 3 categories. Include those identified as comparators and as other relevant therapies in Section 1 of the assessment report. Where the proposed health technology is replacing a technology funded through a different program, or where patients are receiving best supportive care in the absence of the proposed technology, there will be no substituted technologies.

Health technologies funded within the same program for which usage is expected to increase or decrease after the listing of the proposed health technology include those that are:

- coadministered with substituted therapies or with the proposed health technology
- used to treat adverse reactions to substituted therapies or the proposed health technology
- used to treat the clinical end points that might be increased or reduced after the proposed health technology.

The impact of adverse reactions might have less weight if the evidence shows that they are of insufficient clinical importance to require management, or if they are similar for the proposed health technology and its comparator. Note if there is insufficient information available from trial results or extended assessment of comparative harms to include the impact of adverse reactions on expenditure.

Change in other health technologies funded within the same program over 6 years

Discuss the extent of change for each health technology within the same funding program that will be substituted, and for those that are expected to increase or decrease in usage after listing of the proposed health technology. Present and justify the change in the number of services provided for each of these technologies over 6 years. Reference how the estimates were generated and the data on which the estimates are based.

Justify any inconsistencies between the economic evaluation (Section 3) and Section 4 in terms of the identified health technologies or the estimated extent of change in usage over the 6 years following listing of the proposed health technology.

Financial impact over 6 years

Based on estimated utilisation changes, estimate the financial impact in each year of the 6 years following listing for each health technology funded within the same program that is substituted, decreased (i.e. cost offsets) or increased (i.e. on-costs). Refer to TG 27.2 for the suggested approach.

TG 27.4 Estimation of the net financial impact

Net financial impact on the relevant funding program

The net financial impact on the relevant funding program over 6 years should be presented in Section 4.4 of the assessment report. The estimate of net impact should account for the estimated cost of the proposed health technology (estimated in response to TG 27.2), the costs of increased usage of other health technologies and cost offsets for substituted health technologies likely to have reduced usage (estimated in response to TG 27.3).

Net financial impact on the Commonwealth health budget

Change in use and financial impact on other Commonwealth health budgets

Use the approach in TG 27.3 to identify health services funded through other Commonwealth health budgets that are likely to be affected by the listing of the proposed health technology.

Based on estimated utilisation changes, estimate the financial impact in each year of the 6 years following listing for each affected health service and per program (e.g. if multiple PBS items are expected to be affected by the listing of the proposed health technology, estimate the financial impact for the change in each item, and then overall to the PBS). Refer to TG 27.2 for the suggested approach to estimate the financial impact. Present costs with any appropriate patient copayment subtracted.

Overall financial impact on the Commonwealth health budget

Present the net financial impact on the health budget over 6 years, incorporating the impact on the relevant funding program estimated in Section 4.4 of the assessment report, and the changes in use and financial impact on other Commonwealth health budgets estimated in Section 4.5 of the assessment report.

Net financial impact on other health budgets

The decision to subsidise a health technology may result in financial impacts to other stakeholders in the health care system. Key examples that may be of relevance to MSAC decision making are impacts on state government funding, private health insurance and patient out-of-pocket expenses.

Accurate information for describing the impact on other stakeholders may be difficult to identify. As a minimum, an assessment should *identify* other stakeholders and qualitatively describe the financial impact of a funding decision for the proposed health technology. Where the financial impact is expected to be substantial, greater effort may be required to identify quantitative estimates.

If there is a risk that the decision to fund the health technology may result in patient out-of-pocket expenses, consider whether this is relevant to raise in response to Technical Guidance 29 (other relevant considerations).

TG 27.5 Identification, estimation and reduction of uncertainty in the financial estimates

Sources of uncertainty

Uncertainty arises when estimating utilisation and financial impact because of the potential for usage that differs from expectations, and usage that extends beyond the intended population defined by the item descriptor.

Address these sources of uncertainty and clearly differentiate the two. Where there is substantial uncertainty in the utilisation and financial estimates, particularly when this uncertainty is a result of

usage beyond the proposed population ('leakage'), minimise the impact of the uncertainty by proposing a risk-sharing arrangement.

Factors affecting uncertainty

The following subsections list some factors to consider when assessing uncertainties in predicted utilisation patterns and the financial impact resulting from listing of a proposed health technology. The lists are not exhaustive. Factors may arise from epidemiological data, expert opinion and assumptions used in quantifying predictions. Present details of these factors to increase understanding of the uncertainties present in utilisation estimates.

It might not be necessary to address any or all of these factors if the uncertainties are very small or of little importance to the overall cost to the MBS, so consider how relevant each of the factors might be.

Factors that could affect the extent of usage within the proposed population

Consideration of the following factors might provide relevant information on uncertainties within the proposed population:

- Promotion might result in greater identification of the proposed health technology, resulting in more health care practitioners considering patients for treatment.
- Indirect media exposure to consumers might result in some consumers being more aware of, and seeking to use, the proposed health technology. These patients might not be identified if a treated prevalence approach has been used.
- Outcomes of related research might have an impact on uptake of the proposed health technology. This could be positive or negative, and could emerge at the time the assessment report is lodged or be expected to occur within 5 years of listing.
- More health care practitioners and patients might seek treatment if the proposed health technology treats a medical condition for which the alternatives are considered to be substantially inferior to the proposed health technology (e.g. in terms of effectiveness, tolerability, or patient acceptability and convenience).
- Limited access to designated types of health care practitioners or to designated diagnostic procedures in the proposed population might limit uptake and utilisation.
- The duration of treatment might be longer than expected, compared to the time frame of the randomised trials, particularly when trials are truncated.
- Utilisation might be greater than expected, particularly in the case of medical conditions with episodic manifestations.
- There might be a likelihood of usage increasing over time.

Some of these factors might not be relevant in all assessment reports or might have a negligible impact on the overall estimates.

Factors that could affect the likelihood of usage beyond the proposed population

Some of the factors listed above might also affect the likelihood of usage beyond the proposed population. The following factors might also be considered to increase the likelihood of usage beyond the proposed population:

- The proposed population is a subset of the types of patients who are eligible according to indication(s) approved by the Therapeutic Goods Administration.

- The proposed population is a subset of the types of patients who were eligible for the randomised trial(s) published for the proposed health technology, or there are randomised trials demonstrating evidence in other medical conditions.
- The proposed population is a subset of the types of patients who have been subsidised by the applicant before lodgement of the assessment report (e.g. on compassionate grounds or as part of clinical studies).
- The proposed population is a subset of the types of patients for whom the applicant plans to promote use of the proposed health technology before or after the listing for MBS funding is implemented.
- The proposed population is a subset of the types of patients who have the underlying medical condition; in this case, identify whether:
  - there are any likely difficulties for health care practitioners in determining eligibility for the proposed health technology (e.g. a difficult differential diagnosis, ambiguity in the wording of the item descriptor, or poor precision or accuracy in a diagnostic test) that might result in misclassifications of eligible patients from the population with the underlying condition; and /or
  - patient advocacy groups are likely to have an influence on determination of eligibility by health care practitioners.

Impact of uncertainty

Address the following factors in any uncertainty consideration:

- the direction of impact on the estimate (underestimate or overestimate)
- the impact on the magnitude of the estimate (small or large)
- the likelihood that another estimate should replace the base-case estimate (probable or improbable).

Although quantitative estimates of uncertainty are preferred, provide approximate assessments if necessary. Note where the effects of some uncertainties are difficult to quantify. As a general principle, the more sensitive the overall financial impacts are to a particular source of uncertainty, the more important it is to minimise that uncertainty.

Reducing uncertainty

Uncertainty can be reduced by using data from multiple sources, if available, which is sometimes referred to as 'triangulation' (the use of multiple sources of data or multiple approaches to determine the consistency or otherwise of the conclusions from those sources or approaches). Where estimates derived from different sources are concordant, there might be more confidence, and less uncertainty, in the resulting estimates. Where estimates are discordant, the disparity between the estimates might contribute to the estimate of uncertainty.

Similar results following the use of different methodologies may also reduce uncertainty in the estimates. Examples include using estimates based on a market-share approach as well as an epidemiological approach; or using 'treated prevalence', where the prevalence of patients treated for a disease or condition, determined from an epidemiological database, is used as a surrogate for the true prevalence.

Summary of calculations

Summarise the results of any calculations (e.g. sensitivity or scenario analyses) to quantitatively examine the impact of uncertainty.

Section 5 Additional relevant information

Introduction

MSAC's Terms of Reference are to advise the Minister for Health on whether a medical service, health technology or health program should be publicly funded, and what circumstances, if any, should apply to such funding. This advice should be based on an assessment of the comparative safety, clinical effectiveness, cost-effectiveness and total cost using the best available evidence. However, MSAC are free to consider factors beyond the clinical, economic and financial implications of the proposed health technology. The purpose of Section 5 is to discuss *concepts* that may affect implementation of the proposed health technology or influence the decision-making, but which have not been captured through the evaluation of the comparative safety, effectiveness and cost-effectiveness of the technology.

This section provides only minimal guidance on how to assess these additional concepts and present the information; applicants and assessment groups can determine how best to present the additional information. The assessment of value of knowing and other relevant considerations will not be necessary for decision-making for all technologies. This section includes guidance on when this additional information may be relevant to MSAC decision-making.

Value of knowing

For the purposes of these guidelines, value of knowing is defined as the harms and benefits derived by a patient (or a patient's family or carers) from knowing the results of a test. The value of knowing is only considered relevant for the assessment of investigative technologies. The completion of this section is only required if the proposed test is more costly than the comparator, and the additional cost is not adequately justified by an impact on health.

Examples of tests for which consideration of the value of knowing may be necessary are:

- a test that can detect a disease for which there is no available treatment (the test results in an increase in cost but does not result in health gains)
- a test that can provide a prognosis although there is no clinical management that would alter the prognosis.

Benefits or impacts beyond individuals, family members or carers are not considered in the 'value of knowing' section (e.g. the value to a clinician of knowing the result to aid management decisions would be considered under the 'change in management' section).

Other relevant considerations

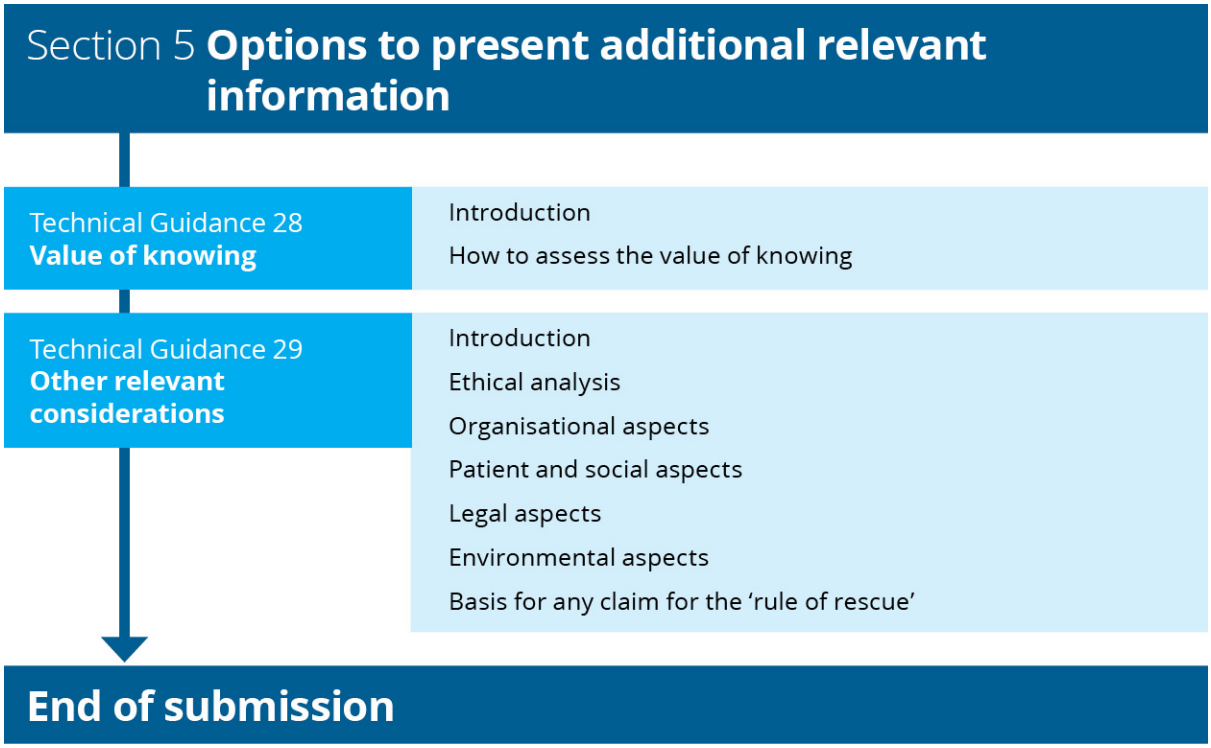
Other considerations that may be relevant include ethical principles such as equity and rule of rescue, and other factors such as organisational issues, social issues, legal issues and environmental issues.

While other relevant considerations may include benefits that fall within the value of knowing category, it also includes impacts that are broader than the individual, family members or carers.

There may be additional impacts of the proposed health technology (medical service) beyond the health care system. Some non–health care system costs and outcomes may be quantified and

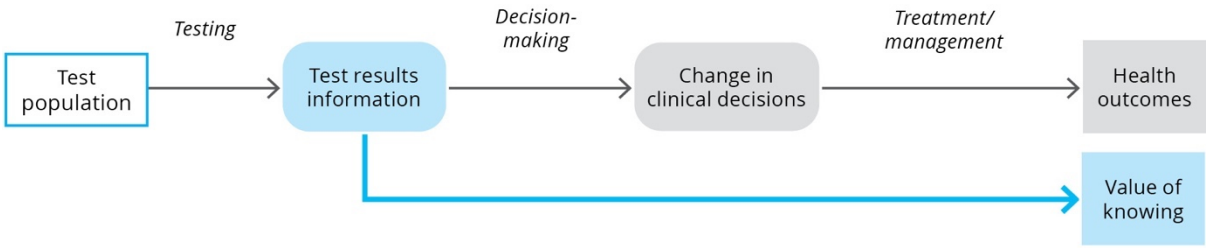
included in supplementary analyses in the economic analysis (see Appendix 6). However, some impacts may be less readily quantified (such as impacts on educational attainment).

Flowchart 9 shows the structure of the guidance provided in this section.



Flowchart 9 **Structure of the guidance on additional claims for value of knowing and other relevant considerations**

Technical Guidance 28 Value of knowing



KEY CONSIDERATIONS

For investigative technologies

Provide evidence supporting any claim that the key benefit of a test is from its impact on nonhealth outcomes (value of knowing). Discuss the benefits and harms of receiving the test information versus what would happen in the absence of the test (TG 28.2).

TG 28.1 Introduction

When the clinical utility of a test cannot be established, MSAC may consider nonhealth benefits and harms from the test under the concept of value of knowing.

For the purposes of these guidelines, the value of knowing encompasses any consequence for the wellbeing of a patient beyond the changes in the health outcomes attributed to changes in the health care provided. These additional outcomes may or may not be able to be demonstrated with quantitative data. The benefits or harms for people other than the patient receiving the health technology (e.g. their family members or carers) is also considered.

When Wilson and Jungner first described their criteria for screening tests in 1968, it was deemed essential that, for screening to be worthwhile, an effective treatment should be available.[83] However, since that time, there has been an understanding that information derived from tests may be used in a variety of ways, such as reproductive planning or changing behaviours relating to sectors outside of health.

Most tests have some value to patients and/or family members by providing a greater degree of certainty about, for example, a diagnosis, risk level or prognosis. In most assessments, the qualitative benefit of the information itself need not be demonstrated, as the clinical claim rests on how the information is used by a clinician in a shared decision-making process with the patient to alter clinical management. However, some test information may not lead to any change in clinical management or health outcomes. Value may still be derived from the test results if they result in, for example, avoiding a lengthy diagnostic odyssey, being able to plan for end of life, engagement with others with a similar diagnosis, or financial support being available because the diagnosis allows access to disability schemes.

Box 6 Example of benefits from value of knowing

Genetic testing to identify X-linked retinitis pigmentosa may be important to identify whether a child is likely to become blind or not. Although no treatment is currently available to prevent or treat the vision loss, the prognostic information provided by a genetic test identifying the risk of this condition

may permit planning for education and career,[84] finding suitable housing, initiating access to the National Disability Insurance Scheme, and connecting with support services and social supports.

Knowledge of test results may also have a negative impact on patients and their family members. Psychological distress following test results is common. The absence of an appropriate treatment or preventive measure for the identified condition may mean that a test is more harmful than beneficial. Information on the pre- and post-testing level of distress for those with positive results, negative results and ambiguous results may be important to consider.

Box 7 Example of harms from value of knowing

Research has demonstrated that testing for Huntington disease and heritable cancers causes more stress if there is an ambiguous result than if there is a positive result.[7, 85] There may also be guilt for passing on heritable diseases, or survivor guilt for unaffected siblings.[7]

The assessment must therefore not presume that additional information is always beneficial, and provide evidence of the impact.

TG 28.2 How to assess the value of knowing

If a claim is made that the key benefit of a test is from the impact of the test on nonhealth outcomes (value of knowing), the evidence supporting this claim must be provided. The benefits and harms of receiving the test information versus what would happen in the absence of the test should be discussed.

Present a summary table of the benefits and harms of the test compared with the comparator (often no test or clinical diagnosis) (see example in Table 24). Support each proposed benefit or harm with reason and evidence. Evidence may have been directly generated for the proposed technology or may have been generated for an alternative technology with similar characteristics. Demonstrate the applicability of evidence from other technologies to the current assessment.

Table 24 Example summary of benefits and harms of proposed test versus comparator

Benefit or harm	Proposed test	Comparator
Reduce the 'diagnostic odyssey'	Availability of genetic testing for pathogenic variants in the *SMN1* gene in infants or children displaying unexplained hypotonia.	Imaging, nerve conduction tests, muscle biopsy, electromyography and blood tests.
Diagnostic delay	The availability of genetic testing results in earlier diagnosis. Even in countries with genetic testing, diagnostic delay is several months for SMA I and almost 4 years for SMA III. Diagnostic delay may result in psychological stress for the caregiver who is unable to determine the cause of their child's illness or access treatment to help them.	
Interventions[a]	Early diagnosis permits early intervention. Irreversible degeneration occurs in the first 6 months in SMA I. Delays in diagnosis prevent early intervention.	Delayed, may reduce efficacy of interventions. Knowing that an earlier diagnosis would have resulted in improved outcomes may result in anger or grief for a caregiver.
Access to support	Early diagnosis permits access to funding through national schemes (NDIS, carer support). Value derived from connection with	Delayed.

Benefit or harm	Proposed test	Comparator
	others.	
Career and life decisions	Different decisions regarding work and life are available after genetic test provides indication of prognosis.	Delayed.
Psychological impact of a diagnosis of a fatal disease	A diagnosis of SMA is accompanied with grief and psychological stress. Availability of genetic testing permits an earlier diagnosis.	Clinical diagnosis is likely to remain uncertain for some time, but will be accompanied by a similar impact as a genetic diagnosis.

NDIS = National Disability Insurance Scheme; SMA = spinal muscular atrophy

a The reported impact of permitting earlier intervention is more appropriately discussed with quantitative evidence of the clinical utility of the test. However, having a certain course of action/ treatment earlier due to earlier diagnosis may be of value to the caregiver. In addition, knowing that interventions are more likely to be efficacious if provided earlier may have negative impacts on the caregiver who is unable to access a timely diagnosis due to the absence of genetic testing.

Evidence to support these proposed value of knowing benefits (or harms) can be discussed at length following the summary table.

Quantitative evidence that allows MSAC to consider the proportion of patients who experience the benefits and harms, and the magnitude of such benefits/harms, is desirable, although it is acknowledged that this evidence may not be possible to generate or identify.

Qualitative evidence may be presented in support of quantitative evidence or where quantitative evidence is of poor quality or absent. Various strategies are available for synthesising qualitative evidence, such as narrative summary, thematic analysis, grounded theory, meta-ethnography, realist synthesis and content analysis.[86] MSAC has no stated preference for how qualitative information should be analysed or presented (although both harms and benefits, for both the proposed test and the comparative scenario, should be considered). Two suggested sources of guidance are the Cochrane Qualitative and Implementation Methods Group,[87] and the National Institute for Health and Care Excellence (NICE).[88]

Where a test may have implications involving personal values (such as termination of an affected foetus), MSAC considers that the value of the test is to provide the prospective parents with information permitting an informed decision on available options. Presenting only the data for one option (such as the termination of a pregnancy) ignores the value of knowing associated with other available options, such as preparation and life planning, which may be equally as valuable to those who choose this option. While quantitative estimates of actions taken that follow the test results may be useful, it is important to explore the full impact of the value of knowing (benefits and harms) across all of the options that are presented following a test result.

Benefits and harms to the individual patient should be presented separately from benefits and harms to their family members/carers.

If the benefits of the test are seen in sectors outside of health, they would not be able to be incorporated into a cost-utility analysis; however, they could potentially be incorporated into a cost-consequences analysis. For advice regarding the cost-consequences approach to claims of value of knowing, see Technical Guidance 17.

Technical Guidance 29 Other relevant considerations

KEY CONSIDERATIONS

- **Identify and discuss any factors that (TG 29.1):**
  - **– are unique to the proposed technology or circumstances of use, such that MSAC is unlikely to have considered the factors previously in the same context**
  - **– have a considerable impact on the way the clinical and economic results are interpreted**
  - **– have been included in the ratified PICO confirmation for further assessment.**
- **These factors may include ethical, organisational, legal, environmental, patient and social aspects of the proposed health technology (TG 29.2–29.6).**
- **As with other relevant considerations, the rule of rescue supplements, rather than substitutes for, the evidence-based consideration of comparative cost-effectiveness (TG 29.7).**

TG 29.1 Introduction to identification and presentation of other relevant considerations

Additional information relevant to decision-making that is not captured elsewhere in the assessment may be captured in Section 5. Evidence in this section should be clearly presented and reasoned. Where possible, evidence should be generated using high-quality methods or sourced systematically.

The goal of the additional relevant information section of an assessment report is to identify and report on factors that:

- are unique to the proposed technology, circumstances of use or funding arrangement, such that MSAC is unlikely to have considered the factors previously in the same context
- have a considerable impact on the way the clinical and economic results are interpreted
- have been included in the ratified PICO confirmation for further assessment.

It is not always possible to verify whether information applicable to Section 5 meets the above criteria, and broader information can be presented. Ensure that the most important features of Section 5 are presented in the main body of the assessment report, and provide broader considerations (if assessed) in an appendix.

Provide a summary of important findings, explain their relevance to decision-making, and describe the source of the information. If possible, present this summary in a tabular format (see example in Table 25). Provide longer analysis or explanations in an appendix.

Table 25 Example summary of additional relevant factors

Type of consideration	Factors for consideration and their impact on the value of the proposed technology	Source(s) of evidence
Ethical, patient and social considerations	The use of the proposed technology is limited to highly specialised centres located in capital cities. Evidence is available that people living with the condition value local treatment options and report anxiety relating to travel and lack of support structures. This was consistent with feedback from consumers on the PICO confirmation. The availability of the technology raises issues of equity. Improving consumer support by subsidising travel, and enabling a support person to travel, has been shown to lessen anxiety in consumers who must travel to distant centres for care.	Review of literature identified as relevant for an ethical analysis during the primary search of the proposed technology. Responses to the public consultation of the PICO confirmation. See Appendix 5 of the assessment report for further detail of the searches and analysis.
	Patients with the target disease require ongoing care to assist with activities of daily living. Evidence indicates that approximately 10% of patients require some assistance in earlier stages, increasing to almost all patients in later stages when institutional care is often required. Submissions from consumers in response to the PICO confirmation identified that care in the home is often provided by a relative, and that accessing professional care is costly and difficult to organise. The proposed technology has been reported to slow progression of the disease; however, studies included in Section 2 have not directly reported on the impact on the level of care provided or the impact on the carer. Evidence supports that a reduction in the need for care and the ability to live independently are valued by patients. Evidence has also reported that interventions that increase patient autonomy have a positive impact on the primary carer. Some of the value identified relating to the improvement in function and maintenance of autonomy are captured by outcomes presented in the clinical assessment (quality of life and functional assessment PROM), and may be captured in the utility values entered into the economic model. However, it is unlikely that the utility measure used (EQ-5D) is adequately sensitive to identify the full value of the benefits experienced by patients.	Review of the literature identified during the primary search for effectiveness and safety of the proposed technology. Targeted search of literature reporting on carers for patients with the disease. Review of disease organisations' websites. Identification of a high-quality study on a different disease with similar characteristics. Article describing the sensitivity of MAUIs in the target disease. Further detail of the searches and analysis are presented in Appendix 5 of the assessment report.
Organisational considerations	Due to the equity issue and potential negative impact on rural and regional consumers identified above, the implementation of the technology should consider measures to mitigate the impact of distance and travel.	Based on the above evidence, described in greater detail in Appendix 5.
	It is unclear whether the number of patients identified as eligible for the technology could feasibly receive the technology due to capacity constraints. In the initial years, it is expected that there would be a prevalent population that may be eligible, resulting in higher demand for the technology. Methods to increase capacity as well as appropriate criteria for triaging may	Comparison of the forecast numbers (Section 4) with the estimated throughput (expert input from specialist hospital).

Type of consideration	Factors for consideration and their impact on the value of the proposed technology	Source(s) of evidence
	be required.	

MAUI = multiattribute utility instrument; PROM = patient-reported outcome measure

TG 29.2 Ethical analysis

'Implementing new technologies in health care can have morally relevant consequences. Technologies carry with them values that can challenge the current mores and attitudes of society. Every HTA requires many value-based decisions to be made during the assessment process'.[89]

Ethical issues may include aspects of a technology not captured by standard measures of safety, effectiveness and cost-effectiveness, but which may affect the value of the technology. Ethical issues may also include factors that stakeholders should keep in mind when funding, producing, delivering or using the health technology. There is no agreed method for undertaking ethical analysis in HTA; however, several frameworks exist.[90]

One of these frameworks is the EUnetHTA core model.[a] According to this model, ethical issues that may be relevant in the assessment of health technologies are:

1. **The balance of benefits and harms** – Consider whether there may be hidden or unintended consequences or implications for patients or other groups that are not captured in assessments of safety and effectiveness. Consider whether there may have been any ethical obstacles impeding those assessments.
2. **Autonomy** – Consider whether the value of the health technology is augmented by its impact on the autonomy of patients or other groups (i.e. the right and capacity of people to direct their own lives). Is the health technology of particular value because it helps to restore or promote the autonomy of patients who are particularly vulnerable or who may have a reduced capacity for exercising autonomy? Conversely, could the health technology result in a reduction in autonomy and thereby be of lesser value? Are there additional interventions that may be required to ensure the target population can provide valid (i.e. informed and voluntary) consent to receive or refuse the health technology?
3. **Respect** – Could the health technology have implications for matters of human dignity, stigma, privacy, or moral, religious or cultural conviction or tradition? Could widespread use of the health technology change our perception of certain persons?[91]
4. **Equity** – Could the implementation of the health technology have impacts on equitable access to care across the target population? Would government subsidy of the medical intervention affect the distribution of health care resources in problematic ways?
5. **The HTA and its implications** – Are there ethical issues or implications relating to the choice of endpoints, populations or comparators in the assessment? Are there ethical problems relating to the assumptions in the economic evaluation? In particular, are there important respects in which the health technology may be of greater or lesser value that the economic evaluation has not captured.[92]

Another commonly used tool for identifying and analysing ethical issues in HTA has been developed by Hofmann et al., based on the Socratic approach.[93]

a eunethta.eu/hta-core-model

TG 29.3 Organisational aspects

Organisational aspects (according to the EUnetHTA core model) are the ways in which different kinds of resources need to be organised when implementing a technology, and the consequences that flow from that in the organisation of the health care system. These resources include human skills, attitudes, material artefacts, money and work culture. Examples of organisational issues that arise are work processes and patient/participant flow, quality and sustainability assurance, communication and cooperation, centralisation, acceptance of a technology, and managerial structure.

EUnetHTA have identified that the organisational domain should include 5 topics, each containing 2 to 6 issues (15 issues in total). These topics generally represent the most important organisational issues; however, their relevance depends on the specific intervention.

The different topics discussed in the organisational domain in the EUnetHTA core model are:

1. **Health delivery process** – How does the technology affect the current work processes? What kind of patient/participant flow is associated with the new technology? What kind of involvement must be mobilised for patients, caregivers and others? What kind of process ensures proper education and training of staff? What kinds of cooperation and communication of activities must be mobilised? And in what way is the quality assurance and monitoring system of the new technology organised?
2. **Structure of the health care system** – How do decentralisation or centralisation requirements influence the implementation of the technology? What are the processes for ensuring access to the new technology for patients?
3. **Process-related costs** – What are the costs of processes related to acquiring and setting up the new technology? How does the technology modify the need for other technologies and use of resources? What is the likely budget impact of implementing the technology?
4. **Management** – What management problems and opportunities are attached to the technology? Who decides which people are eligible to receive the test/treatment and on what basis?
5. **Culture** – How is the technology accepted? How are other interest groups taken into consideration during the implementation of this technology?

A systematic search of the literature is unlikely to efficiently identify relevant organisational issues. However, potential organisational issues that are identified during the various steps in the assessment process (e.g. in the application phase, during the development of the PICO confirmation, during the literature search, during the assessment of safety and effectiveness) should be discussed.

Qualitative research identified during the main literature search for the proposed technology may assist in understanding how patients perceive health, how they make decisions regarding health service usage, and the culture of communities in relation to implementing changes and overcoming barriers. Clinical practice guidelines and other HTAs could also contain information useful for identifying possible implementation and organisational issues.

TG 29.4 Patient and social aspects

In the EUnetHTA core model, patient aspects relate to issues relevant to patients/individuals and their caregivers, whereas social aspects are related to social groups. Individuals who receive the intervention and their caregivers can provide unique perspectives on the experiences, attitudes, expectations and values regarding the intervention, and regarding health, illness, service delivery and treatments, which can inform the HTA.

Literature on these aspects is often referred to as 'patient-based evidence'[94] and includes evidence from patients, individuals, caregivers and social groups about the burden of living with the condition being studied, experiences of current practice or current health technologies, or experiences with and expectations of the health technology being studied.

The PICO confirmation and main body of the assessment report discuss relevant outcomes (as part of the PICO). Part of the justification for choosing outcomes is that they capture the most patient-relevant effects of the disease and impacts of the treatment. Therefore, some information relating to patient or social issues will need to be presented in the justification of the relevant outcomes. Perform additional targeted searches of HTA reports or recent studies to complement key patient or social issues identified during the assessment of effectiveness or safety, or at the PICO confirmation stage. Key issues are those that are novel or unique to the proposed technology, or likely to influence the interpretation of the clinical or economic analyses.

Evidence on perspectives of the patient and other stakeholders that is identified during searches for the effectiveness and safety of the proposed technology may also be relevant to present.

TG 29.5 Legal aspects

The objective of the legal aspects domain in the EUnetHTA core model is to assist HTA assessors in detecting rules and regulations that should be taken into consideration when evaluating the implications and consequences of implementing a health technology. As technologies rapidly change, policy- and decision-makers are required to know the legal implications of implementing or not implementing a technology. Some of the legal aspects outlined by the EUnetHTA model may be especially important to assess for digital technologies (e.g. ensuring patient data are appropriately secure).

Some elements relating to legal issues are also relevant in the ethical domain. The different topics that could be addressed when discussing legal aspects of the health technology are:

1. **Autonomy of the patient** – Who is allowed to give informed consent for the technology for minors and incompetent persons? What legal requirements are in place for providing appropriate information to the user? How should this be addressed when implementing the health technology?
2. **Privacy of the patient** – Does the use of the health technology produce additional information that is not directly related to the patient's care and may violate their right to privacy? What do the laws say regarding informing relatives about the results? What are the laws regarding the security of patient data and how should this be addressed when implementing the health technology?
3. **Equality in health care** – What do laws require regarding processes or resources that would facilitate equal access to the health technology? What are the consequences of rules and regulations around equal access to the technology?
4. **Ethical aspects** – Does the implementation of the technology affect the realisation of basic human rights? And can the implementation of the health technology give rise to ethical challenges that have not yet been considered in existing laws and regulations?
5. **Authorisation and safety** – What rules and laws are present around safety of the technology and how should this be addressed when implementing the health technology?
6. **Ownership and liability** – What should be known and reported about intellectual property rights and potential licensing fees, and about the regulations regarding the manufacturer's guarantee?

Who would be responsible if the health technology fails or provides false results? What would the medicolegal consequences be for 'overrelying' on test results?

7. **Regulation of the market** – What are the laws surrounding price control mechanisms of the technology? Are there laws or regulations regarding acquisition and use of the technology? Are there legal restrictions for marketing the health technology to users? What should be known around legal issues in cases of new technologies where current legislation is not directly applicable? Are there concerns about conflicts of interest regarding the preparation of binding rules and their implementation?

If any of the topics concerning legal issues stated above are relevant for the proposed health intervention, they should be discussed.

Information on legal aspects of health technologies can be found in journals (e.g. *Health Economics*, *Policy and Law*, *Medical Law International*, *Medical Law Review*, *Medicine and Law*), or websites such as the Federal Register of Legislation.[a]

TG 29.6 Environmental aspects

If there are particular concerns regarding the environmental impact of the proposed health technology, or a key benefit in the way the proposed technology reduces the environmental impact compared with the comparator, these can be outlined (e.g. if there is a reduction in the amount of radioactive waste generated, or a reduction of emissions related to transportation or the manufacturing process).

TG 29.7 Basis for any claim for the 'rule of rescue'

The 4 factors described below apply in exceptional circumstances and are particularly influential in favour of listing. When all 4 factors apply concurrently, this is called the 'rule of rescue':

- No alternative exists in Australia to treat patients with the specific circumstances of the medical condition that meet the criteria of the requested restriction. This means that there are no nonpharmacological or pharmacological interventions for these patients.
- The medical condition defined by the requested restriction is severe, progressive and expected to lead to premature death. The more severe the condition or the closer a person with the condition is to death, the more influential the rule of rescue might be in MSAC's consideration.
- The medical condition defined by the requested restriction applies to only a very small number of patients. Again, the fewer the patients, the more influential the rule of rescue might be in MSAC's consideration. However, MSAC is also mindful that the MBS is a community-based scheme and cannot cater for individual circumstances.
- The proposed technology provides a worthwhile clinical improvement sufficient to qualify as a rescue from the medical condition. The greater the rescue, the more influential the rule of rescue might be in MSAC's consideration.

As with other relevant considerations, the rule of rescue supplements, rather than substitutes for, the evidence-based consideration of comparative cost-effectiveness. A consideration of the rule of rescue is only necessary if MSAC would be inclined to reject a submission after considering the technology's comparative cost-effectiveness (and any other relevant considerations). MSAC would

a http://www.legislation.gov.au/

then consider whether the rule of rescue is sufficiently influential to change the decision in favour of a recommendation to list.

This guidance on the rule of rescue is deliberately narrow. MSAC is concerned that broadening the criteria would reduce the relative influence of the rule of rescue. In other words, the greater the proportion of technologies the rule of rescue is applied to, the smaller its average impact in favour of listing.

One issue that has arisen concerning the rule of rescue is that it cannot be considered for a second health technology that becomes available to treat the medical condition that is considered to meet the requirements of the rule. This is because, by definition, the second technology does not meet the essential first factor (i.e. that there is currently no alternative intervention). This causes a difficulty if listing of the second technology is sought on a cost-minimisation basis.

Appendix 1 Assessment frameworks

Assessment framework for noninferiority based on change in management

If the proposed test reports on a different parameter than the comparator, an assessment of comparative test accuracy is not possible. However, if the clinical interpretation of the results is the same, concordance on the clinical interpretation (or categorisation) is required. This would be evident if the same management decisions were made for the same test subjects regardless of the test used. Evidence that there would be no changes in management compared with the comparator test may permit noninferior health outcomes to be inferred (assuming there are no differences in test safety).

An assessment framework for a claim of noninferiority based on concordance of decision-making is shown in Figure 27.

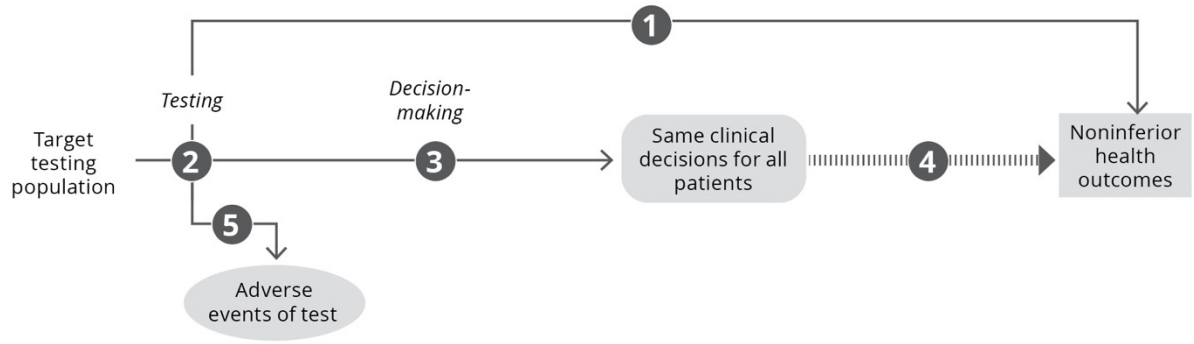


Figure 27 **Assessment framework that has been truncated at decision-making with the inference that concordant decision-making will result in the same health outcomes**

Assessment questions for a claim of noninferiority for a test based on concordance of decision-making (Figure 27)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

1 Does the use of [the proposed test] in [the target population] result in the same or better [key health outcomes, e.g. survival, quality of life] compared with [the main comparator]? (If adequate direct from test to health outcomes evidence is available, go to Assessment question 5).

LINKED EVIDENCE

2 *Not able to compare test accuracy as tests report on different parameters or biomarkers*.

❸ Does the use of [the proposed test] in [the target population] result in the same clinical decisions compared with [the main comparator]?

- Is [the proposed test strategy] concordant with [the main comparator test strategy] in the categorisation of the test results? (If different parameters are measured by the proposed test and the main comparator, they may still be categorised similarly – e.g. into low, moderate or high risk). Is the categorisation of test results a validated tool for decision-making? Is there evidence that clinicians will behave the same way regardless of the test used to inform the categorisation?
- Is [the proposed test strategy] concordant with [the main comparator test strategy] for clinical decision-making? How would decision-making nonconcordance impact patient health outcomes?

❹ Inference that decision-making concordant with existing comparator test strategy will result in noninferior health outcomes.

❺ What are the harms of [the proposed test] and of [the main comparator]?

Assessment framework for triage testing

If the proposed test is a triage test, such that it reduces the number of patients who will require a more definitive test, the final categorisation of patients or the final decision-making following the proposed test strategy is of interest. This means that those ruled out from having the condition by the triage test will need to be followed to see whether they ultimately have the condition and/or present for the definitive test at a later date.

Where the final classification of patients following triage testing plus definitive testing versus definitive testing alone is identical (or concordant), the framework may be truncated at the final classification (and a claim of noninferiority is appropriate). If the final classification of patients is not concordant, or there are differences in the timing of the final classification, the framework will need to include the impacts of subsequent steps in a linked evidence approach to establish the impact of the triage test on health outcomes.

A key consideration with triage testing might be that a greater proportion of the test population will adhere to the triage test, particularly if the triage test is less invasive than the definitive test (e.g. a blood test versus a biopsy). Test uptake, or adherence or compliance is unlikely to be informed by a direct trial of the proposed testing strategy versus the comparator test (as adherence within a trial setting is often artificially high).

A second consideration will relate to the pathway for patients with the condition (true positives) who are determined to be negative by the triage test. Typically, there is a delay in the eventual diagnosis, and the impact of this delay due to the reduced accuracy of the triage test should be explored.

An assessment framework for a claim of noninferiority for use of a triage test is shown in Figure 28.

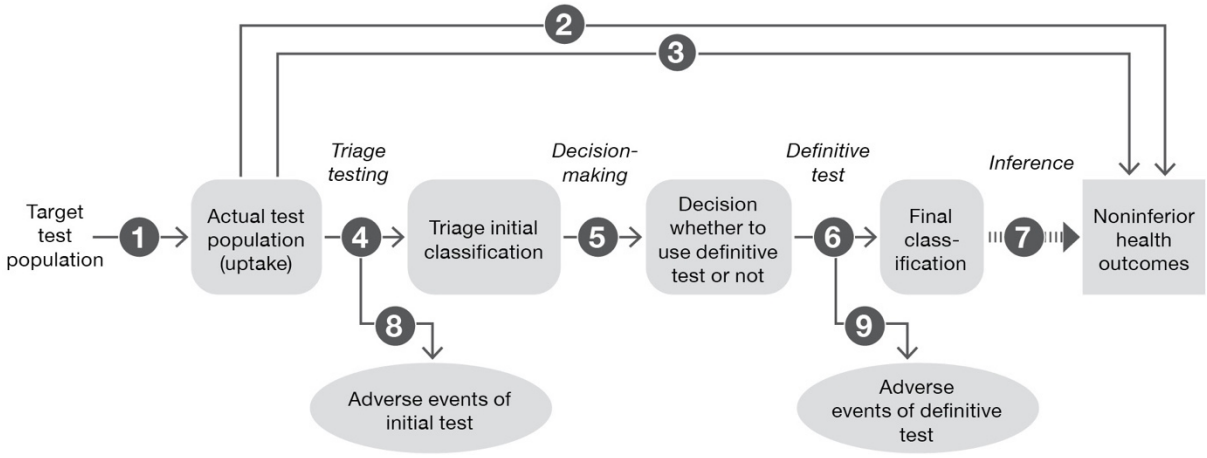


Figure 28 **Assessment framework that has been truncated at the final classification of test results (following a triage and definitive test) with the inference that the final classification will result in the same health outcomes**

Assessment questions for a claim of noninferiority for use of a triage test compared with the main comparator (definitive test) (Figure 28)

1 Does the use of the proposed triage test change the uptake rate for testing compared with the current testing regimen?

- If there are differences in the populations who receive the triage test compared with the comparator, are these differences likely to be associated with the result of the tests (e.g. are high-risk patients more likely to receive the test, or low-risk patients more likely to be noncompliant)?
- What is the impact of differences in uptake rates on the final test results?

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

2 Does the use of [the proposed triage test] in [the target population] result in the same or better [key health outcomes, e.g. survival, quality of life] compared with [the main comparator/definitive test]? (If adequate direct from test to health outcomes evidence is available, go to Assessment question 8).

LINKED EVIDENCE

3 Does the use of [the proposed triage test] in [the target population] result in the same categorisation (positive/negative, presence/absence, high risk/low risk) or the same clinical decisions compared with [the main comparator]?

4 What is the test accuracy of [the proposed triage test] compared with [a reference standard/the comparator definitive test]? What is the nature of the incorrect classifications (i.e. ratio of false positives to false negatives) from using the triage test? What are the clinical consequences of the false negative triage result? *In an asymptomatic population, when the triage test is a screening test, the consequences of false positive testing are also important.*

❺ Does information from [the proposed triage test] result in a change in investigative thinking and change in the individuals who are referred for the definitive test?

❻ *[If the definitive test is not established practice or is not the reference standard]* What is the test accuracy of [the definitive test] in terms of sensitivity and specificity?

❼ Inference that the same final classification of patients (all patients classified using the definitive test/comparator are classified similarly if the triage test were introduced) will result in the same health outcomes.

❽ What are the harms of [the proposed triage test]?

❾ What are the harms of [the definitive test/comparator]?

Assessment framework for a more definitive test

If the proposed test is replacing more than one test (i.e. the proposed test is more definitive or able to test multiple parameters concurrently), the added value to decision-making relates to either the final categorisation of patients (described above for triage testing), or the decision-making following the proposed test compared with the decision-making following the full test strategy it is intended to replace.

Assessment framework for monitoring

Investigative technologies intended to be used for monitoring are assessed in a similar way to diagnostic tests. However, a key difference is that monitoring tests are commonly repeated at intervals to detect a condition that would affect clinical decision-making.

The assessment framework remains similar to that for other types of investigative technologies; however, the assessment questions are expanded to incorporate the characteristics of a monitoring technology.

If monitoring is followed by a confirmatory test, the first change in management would be to undertake this test, and a subsequent step would include management decisions for treatment. Both steps in the change in management must be assessed in the evaluation of a monitoring test.

A key uncertainty regarding monitoring involves whether the monitoring test results in a change in management compared with current clinical practice. For this reason, more emphasis should be placed on robust evidence to support a change in management compared with current practice. As the change in management may have occurred because the condition is detected earlier than it would have been using standard clinical practice, there is a risk of a change in the spectrum of the disease being detected. The applicability of treatment evidence that has been derived from standard clinical practice is therefore a concern. As with the assessment of all tests, comparative direct from test to health outcomes evidence is preferred.

An assessment framework adapted for a monitoring test is shown in Figure 29.

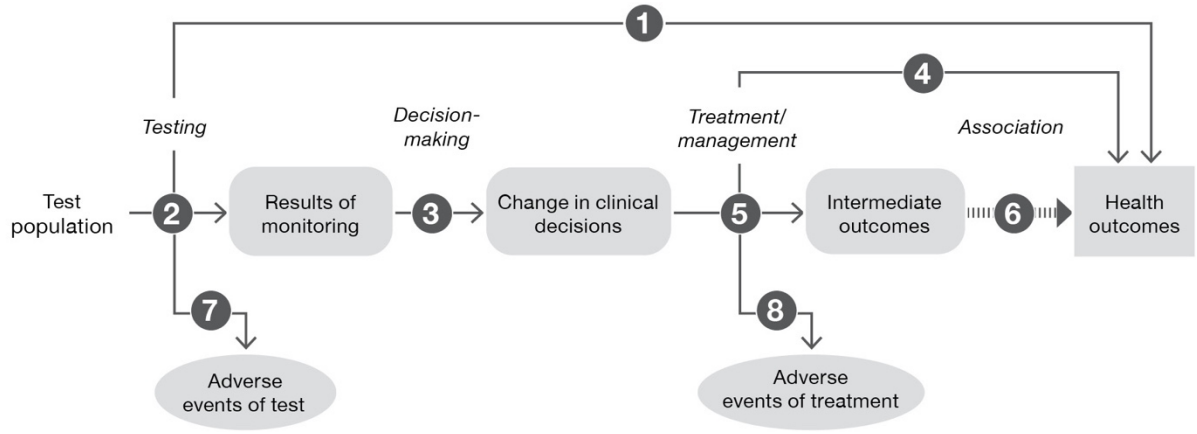


Figure 29 Assessment framework adapted for a monitoring test

Assessment questions for a claim of superiority relating to the use of a monitoring test (Figure 29)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

1 Does the use of the test strategy in place of the current test strategy (comparator) result in the claimed superior health outcomes?

- If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard?

LINKED EVIDENCE

2 How does the information from the proposed test differ from that of the comparator? What is the concordance of the findings from the proposed test relative to the comparator? What is the accuracy of the proposed test (against a relevant reference standard) compared with the comparator?

- If there is a change in the timing at which information becomes available, the results of monitoring will contain a time component. Is there evidence that the proposed test will result in a change in the timing of the detection of a condition compared with the comparator?
- If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard? This applies to concordance of both the test results and the testing protocol (periodicity of the testing).

3 Does the availability of new information from the proposed test lead to a change in management of the patient (compared to the information gained from the comparator)?

- The change in information provided by the proposed test may represent different test results and/or different timing of test results.

4 Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention, or differences in the timing of treatment/intervention), result in the claimed health outcomes?

- Has the treatment/management been provided to a population with the same spectrum of disease that the proposed test identifies? Is it biologically plausible that the

treatment/management will be as effective in the population with this spectrum of disease? *When the proposed test detects patients earlier than the comparator the treatment effect evidence may have been based on a population that was identified at a later time point.*

❺ Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention, or differences in the timing of treatment/intervention), result in the claimed surrogate outcomes?

❻ Is the observed change in surrogate outcomes associated with a concomitant change in the claimed health outcomes, and how strong is the association?

❼ What are the adverse events associated with the proposed test strategy and the comparative test strategy?

- Include downstream adverse events associated with any changes to subsequent testing (such as confirmatory testing).

❽ What are the adverse events associated with the treatments/interventions that lead from the management decisions informed by the test and by the comparator?

Assessment framework for multifactorial algorithms, black-box and self-learning algorithms

Algorithms are a broad category of investigative technologies that commonly include risk scores, nomograms, prognostic scores, and, more recently, self-learning software that processes genomic data, physiological data or imaging data to provide a diagnosis or an estimate of risk of a condition. The key characteristic of an algorithm is that the method of categorisation of patients (how the algorithm weights measured parameters to provide an estimate) is not easily, or cannot be, understood. For this reason, in some circumstances, there is no obvious reference standard against which the algorithm can be compared. For all types of algorithms, and particularly those for which the final step of the linked evidence approach (treatment) has uncertain applicability, direct from test to health outcomes evidence is preferred.

The approach to the assessment of an algorithm varies depending on:

- the clinical claim and the purpose of the test. If the test is prognostic or predictive, it will require longitudinal data, whereas a diagnostic test *may* require cross-sectional data.
- the presence of a reference standard. If the algorithm is being used to detect something that can be verified clinically (such as the presence of a tumour on imaging), the test accuracy can be established, and the applicability of downstream changes may be assessed. In the absence of a relevant reference standard, the accuracy of the test cannot be determined, and direct from test to health outcomes evidence will be required.
- the applicability of the training and validation dataset to the Australian population. Algorithm results by subgroups of interest are required to establish whether there is a risk of the algorithm failing in different populations.
- the applicability of the final step (intervention or treatment) in a linked evidence approach to the population identified by the algorithm (e.g. change in spectrum of disease).

A subsequent step to the assessment of self-learning dynamic algorithms will relate to the safeguards that are in place to ensure the algorithm remains applicable as it continues to evolve once it is available in clinical practice.

An assessment framework for a claim of superiority relating to use of a multifactorial algorithm is shown in Figure 30.

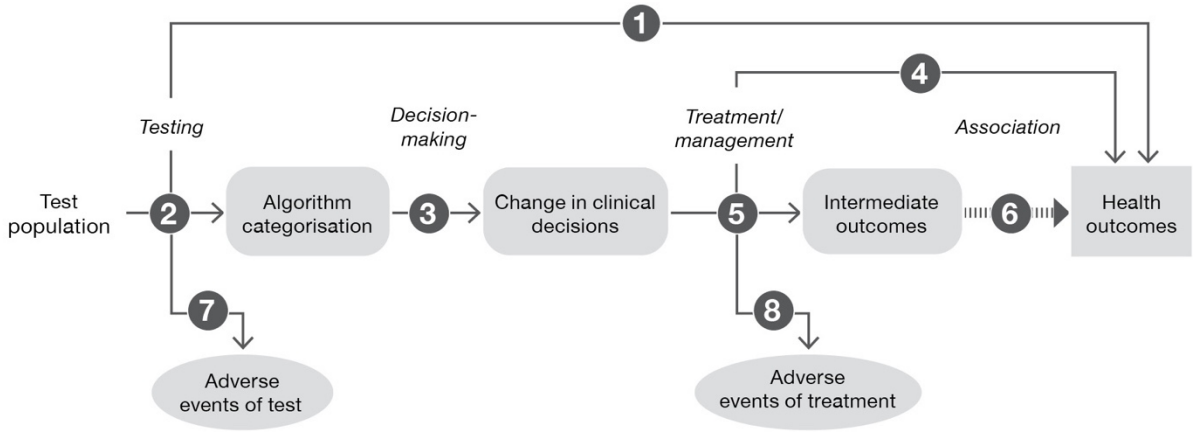


Figure 30 Assessment framework adapted for a multifactorial algorithm

Assessment questions for a claim of superiority relating to use of a multifactorial algorithm (Figure 30)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

1 Does the use of the test strategy in place of the current test strategy (comparator) result in the claimed superior health outcomes?

- If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard? Concordance is measured on the categorisation made by the algorithm. Where an appropriate reference standard is unavailable, a very high concordance would be required to determine that additional tests should be eligible for the same funding arrangements. In the absence of very high concordance or robust concordance data, alternative tests cannot leverage the direct from test to health outcomes evidence of the clinical utility standard.

LINKED EVIDENCE

2 How does the information from the proposed test differ from that of the comparator? That is, how do patient classifications differ using the algorithm versus standard practice?

- If there are multiple tests in clinical practice likely to be able to utilise the same funding arrangements, are these tests concordant with the proposed test and/or clinical utility standard?
- Are the data that were used to construct and validate the algorithm applicable to the target setting? Are there any populations missing from the training or validation datasets? Are there any key differences in the test accuracy across population subgroups?
- Is there a risk that the classification of patients will change over time (is the algorithm dynamic)? What safeguards are in place to ensure that changes to the algorithm are appropriate, or represent an improvement in accuracy?

❸ Does the availability of new information from the proposed test result in a change in management of the patient (compared to the information gained from the comparator/standard practice)?

❹ Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention), result in the claimed health outcomes?

- Has the treatment/management been provided to a population with the same spectrum of disease that the proposed test identifies?

❺ Do the differences in the management derived from the proposed test, relative to the comparator (e.g. differences in treatment/intervention), result in the claimed surrogate outcomes?

❻ Is the observed change in surrogate outcomes associated with a concomitant change in the claimed health outcomes, and how strong is the association?

❼ What are the adverse events associated with the proposed test strategy and the comparative test strategy?

❽ What are the adverse events associated with the treatments/interventions that lead from the management decisions informed by the test and by the comparator?

Assessment framework for universal screening tests

Due to the low prevalence of conditions that are tested for in population or universal screening, there is a high risk that the harms of the tests and the harms associated with false positives may outweigh the value of earlier detection. A further complication relates to the detection of conditions prior to clinical suspicion, based on symptoms or high-risk parameters. Earlier detection of the disease may have little influence on treatment outcomes, or may result in earlier treatments without any evidence that earlier intervention is more effective.

Consequently, universal or asymptomatic screening tests require direct from test to health outcomes evidence of the utility of the screening test. This may include direct from test to health outcomes evidence, or direct from test to an intermediate outcome evidence that can be robustly translated to a health outcome.

Further considerations for screening tests are presented in TG 15.1.

An assessment framework adapted for a population or universal screening test is shown in Figure 31.

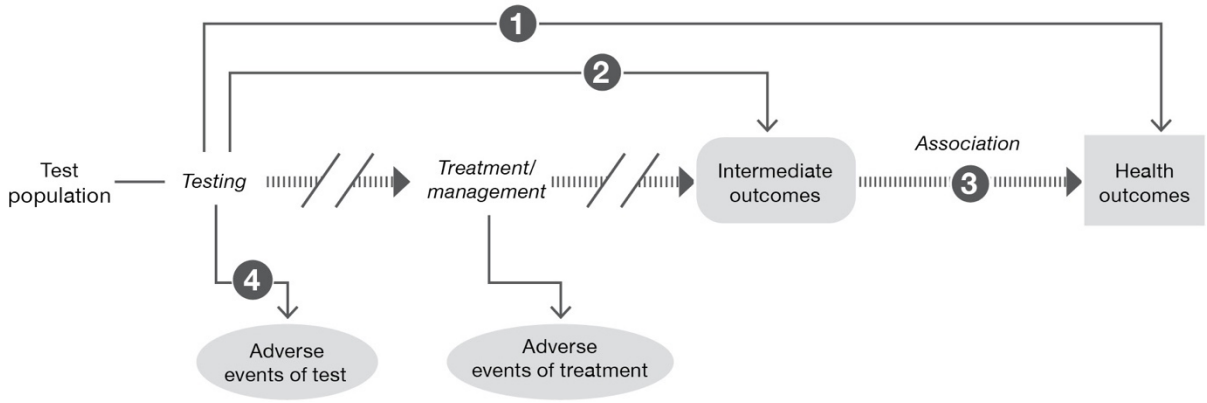


Figure 31 Assessment framework adapted for a population or universal screening test

Assessment questions for a claim of superiority relating to use of a population or universal screening test (Figure 31)

DIRECT FROM TEST TO HEALTH OUTCOMES EVIDENCE

1 Does the use of the test strategy in place of the current test strategy (comparator) result in the claimed superior health outcomes?

2 Does the use of the test strategy in place of the current test strategy (comparator) result in a change in intermediate or surrogate outcomes?

3 Is there evidence to support the validity of the translation of the intermediate outcomes to health outcomes for the populations identified by the proposed test?

4 What are the adverse events associated with the proposed test strategy and the comparative test strategy?

5 What are the adverse events associated with the treatments/interventions that lead from the management decisions informed by the test and by the comparator?

Appendix 2 Literature search methods

The primary objective of Section 2 is to provide the 'best evidence' to answer the assessment questions presented in Section 1. This appendix describes the search methods to ensure all relevant studies have been included in the clinical evaluation, as is appropriate for a full HTA.

Abbreviated search methods may be appropriate for an exemplar/facilitated assessment (see TG 5.2).

Search terms for therapeutic technologies

In most cases, the process of identifying the 'best evidence' is to identify all randomised controlled trials (RCTs) that compare the proposed therapeutic technology with the main comparator(s). However, there are situations where no RCTs will be available and other 'lower level' study designs are acceptable.[10] For instance, if the intervention has already been used for a number of years, if it is infeasible to perform randomised studies (e.g. if equipoise is lacking), if patients are unlikely to participate in a randomised study (given the already available data), if there is no alternative treatment, or if the intervention applies to a rare condition, only lower level evidence may be available.[95] If no direct randomised comparisons are located, indirect comparisons of randomised trials and/or nonrandomised studies will be required.

If no comparative evidence is identified, noncomparative literature should be assessed for both the intervention and the comparator to allow MSAC to make conclusions regarding the comparative effectiveness and safety of the technologies.

Search strategy

Develop a search strategy to address the assessment questions presented in Section 1.

An appropriate search strategy should have the following characteristics:

- It should include a search for studies involving either the proposed therapeutic technology or the main comparator(s). This approach would permit the identification of trials required to perform an indirect comparison. If comparative evidence is identified (e.g. a direct RCT), it may not be necessary to perform a search for the main comparator.
- It should not be restricted by study type.
- When a device is involved in the proposed therapeutic technology, the search should not be restricted by a particular manufacturer's device. In some circumstances, where the MBS item is intended to be restricted to a specific device, restricting the search to this device may be appropriate.

Search filters and additional search terms should be used with caution. Additional search filters or terms may include:

- randomised studies or systematic reviews
  - It is generally not appropriate to limit searches to include only randomised trials or systematic reviews.
  - If, during the PICO confirmation or during scoping searches, high-quality randomised studies are identified that adequately address the assessment questions, a filter that excludes

nonrandomised studies from the search may be appropriate (e.g. Cochrane Highly Sensitive Search Strategies[a]).
  - Limiting to only randomised studies or systematic reviews should only be necessary if the number of references retrieved is otherwise unmanageable.
  - A justification for restricting the search strategy to randomised studies or systematic reviews would include 1) the number of citations retrieved in the absence of the filter, and 2) careful consideration of the applicability of the high-level evidence to the Australian setting (see 'Supplementary evidence' below).
- population
  - The search may be limited to identify studies that include participants with characteristics that overlap with those of the target population. In general, this approach is only relevant if the proposed health technology and/or the main comparator(s) are used across multiple populations/indications that are not relevant to the assessment. Care should be taken when excluding studies in different populations or indications, particularly if adverse events may be generalisable to the proposed use of the health technology.
- date range
  - The search period may be limited to the earliest use of the proposed health technology and the main comparator(s). If comparative evidence is identified, the search period may only need to extend to the earliest use of the proposed health technology.
  - If the proposed technology or comparator has changed over time, consider whether limiting the search period to recent literature is justifiable (such as the last 10 years).
  - If a relevant and high-quality systematic review is identified, a search period designed to identify new information may be appropriate.
- language
  - Searches may be limited to articles published in English or with reliable translations.
- publication type
  - Conference abstracts would only be accepted as evidence under exceptional circumstances.

For most assessments, a broad search strategy is appropriate. In circumstances where a large number of studies is identified that address the critical outcomes, studies of lower quality may be excluded from the search results. Search strategies that are limited by study type are generally not appropriate to remove lower quality studies.

If a focused search strategy is used, explain why. Justify why the included literature is adequate to address the effectiveness and safety of the proposed health technology, and that important studies have not been missed. Higher level evidence (RCTs or systematic reviews) may not report long-term safety or all relevant patient outcomes, or may not be applicable to the Australian clinical setting. Therefore, describe any gaps in the evidence, or uncertainties, associated with applying a focused search. Describe methods used to supplement the high-level evidence, if required.

Supplementary evidence

Although randomised trials may provide the most robust estimates of comparative effectiveness and safety, they may not, by themselves, provide the 'best evidence' or complete evidence. Well conducted nonrandomised studies or indirect comparisons may be informative and/or constitute the 'best evidence' for addressing the assessment question(s).

a training.cochrane.org/handbook/current/chapter-04-technical-supplement-searching-and-selecting-studies#section-3-6-1

Consider including supplementary evidence if the use of the included highest level of evidence has the following concerns:

- inadequate applicability to the Australian setting
- omission of important population groups
- differences in the circumstances of use of the proposed health technology or comparator(s)
- omission of important patient-relevant outcomes
- does not report long-term effects or safety
- does not assess user proficiency.

Explain the decision to include supplementary evidence beyond the highest level of evidence.

Present all the relevant search strategies in a technical appendix to the assessment report.

Presentation of the search strategy

The clear presentation of the search terms used improves the transparency of the approach. Tabulating the search terms can assist with presentation (Table 26). Present a table of the search terms for each bibliographic database or data source, and for each search (if more than one search is performed). The presentation of the search terms must explain how each of the terms interacts with other terms (i.e. Boolean operators).

Table 26 Search terms for the literature review (therapeutic technology)

Category	Description	Search terms
Study design (if justified)	[insert description of category]	[e.g. Cochrane Highly Sensitive Search Strategies for identifying randomised trials in MEDLINE, or MeSH and text word terms for nonrandomised study designs]
Population	[insert description of category]	[include MeSH terms, text words and synonyms for the target population/disease/condition]
Intervention	[insert description of category]	[include known proprietary and nonproprietary names, MeSH terms]
Comparator	[insert description of category]	[include known proprietary and nonproprietary names, MeSH terms]

MeSH = medical subject headings

Search terms for investigative technologies

The 'best evidence' for assessing a test would include studies that randomise participants to receive the proposed test or the test comparator and report on final health outcomes. However, studies that generate such direct from test to health outcomes evidence are uncommon, and additional searches are likely to be required to complete a linked evidence approach.

Search strategy

Develop a search strategy to address the assessment questions presented in Section 1.

PICO assessment questions, based on the assessment framework, will usually include questions relating to direct evidence of the test impact on health outcomes, as well as linked steps, including test accuracy, test concordance, change in management and the impact of change in management on health outcomes.

A broad search strategy that includes terms to identify the index test will identify studies (if available) that report on direct from test to health outcomes evidence compared to the comparator, and, if taking a linked evidence approach, the test-related evidence up to change in management. A separate search will be required to identify the impact of change in management (treatment, interventions) on health outcomes (i.e. a treatment-related search) (Figure 32). If direct from test to health outcomes evidence is available for the intervention but does not provide information on the incremental clinical utility (i.e. direct evidence of health outcomes after the comparator test strategy), a separate search could be performed for this information to allow an indirect comparison.

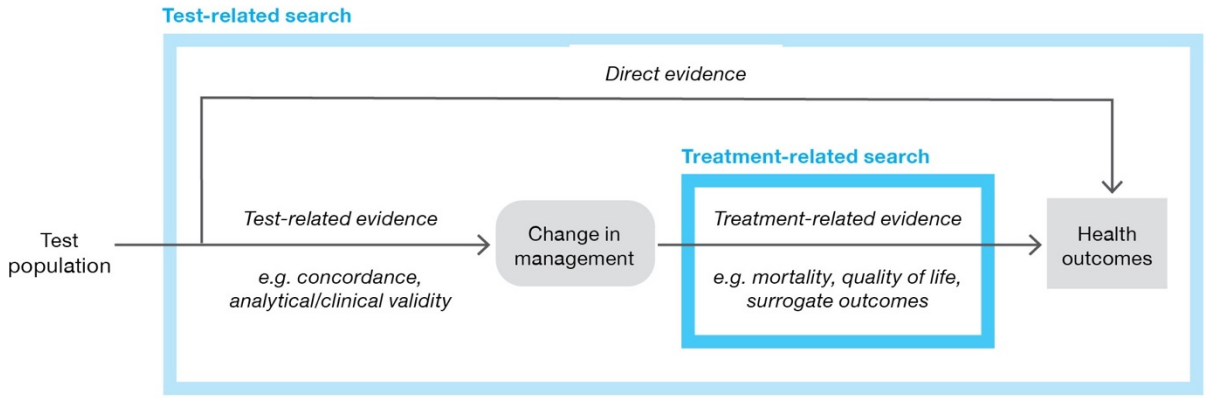


Figure 32 **Components of the assessment identified by test-related searches and treatment-related searches**

Test-related search (health impact of test, impact test has on management of patient, and accuracy of test)

An appropriate search strategy would usually have the following characteristics:

- include terms to identify the proposed test
- include terms to identify the comparator (if studies directly comparing the proposed test and comparator are not available)
- include terms to identify the target condition to be detected
- not restricted by study type.

It is preferable to start with a broad search strategy and exclude studies that are not relevant during the screening phase of the assessment.

Search filters and additional search terms should be used with caution. In circumstances where the number of results retrieved is large and unmanageable in the timeframes available, there are several options for using search filters:

- randomised studies or systematic reviews
  - It is generally not appropriate to limit searches to include only randomised trials or systematic reviews. These study types are unlikely to provide information required to undertake a linked approach.

- date range
  - The search period may be limited to the earliest use of the proposed test and the main comparator(s). If comparative evidence is identified, the search period may only need to extend to the earliest use of the proposed technology.
  - If the proposed technology or comparator have changed over time, consider whether limiting the search period to recent literature is justifiable (such as the last 10 years).
  - If a relevant and high-quality systematic review is identified, a search period designed to identify only new information may be appropriate.
- language
  - Searches may be limited to articles published in English or with reliable translations.
- Publication type
  - Conference abstracts would only be accepted as evidence under exceptional circumstances.

It is generally not appropriate to limit search strategies by study type. If the number of relevant studies identified is large, it is preferable to exclude studies of lower quality at the screening phase of the literature review.

The following guidance is relevant to searches designed for identifying test-related articles:

- Use a wide range of text words for each of the concepts (including synonyms, related terms and variant spellings). Filters for specific terms should be avoided.
- Use truncations and wildcards to capture variations in terms.
- Customise search strategies for each database (either manually or using a tool such as Polyglot Search Syntax Translator[a])
- Do not rely on controlled vocabulary (subject headings) alone, and do not limit searches by filters for test performance (e.g. sensitivity, specificity, concordance) as they do not capture change in management studies or direct from test to health outcomes evidence.
- Explode terms when the option is available.
- Use preliminary searches to identify a range of search terms.

All search strategies used should be saved and reported separately for each database searched, including which filters were used (if any). The date that the search was conducted should also be reported, and how many records were retrieved for each database searched.

Search for the impact of a change in management (treatment-related search)

Approaches that truncate the assessment framework (e.g. claims of noninferiority that can be established by comparing test characteristics) do not need to provide evidence of treatment effectiveness following the test.

If a full linked evidence approach is required, evidence of the impact of a change in management on health outcomes is unlikely to be identified with a search that applies the proposed test as a search term.[b] A separate search for the impact of the change in management is therefore required. The type of searches will be influenced by the change in management identified in the PICO confirmation clinical management algorithm or identified in the assessment of change in management. Different

a sr-accelerator.com/ - /polyglot

b By definition, if this evidence was available, it would be considered direct from test to health outcomes evidence of clinical utility, hence not requiring a linked approach.

sets of patients are likely to vary in the changes in management resulting from the test. A separate search may be required for each change in management that would occur following the receipt of the test results (Figure 33).

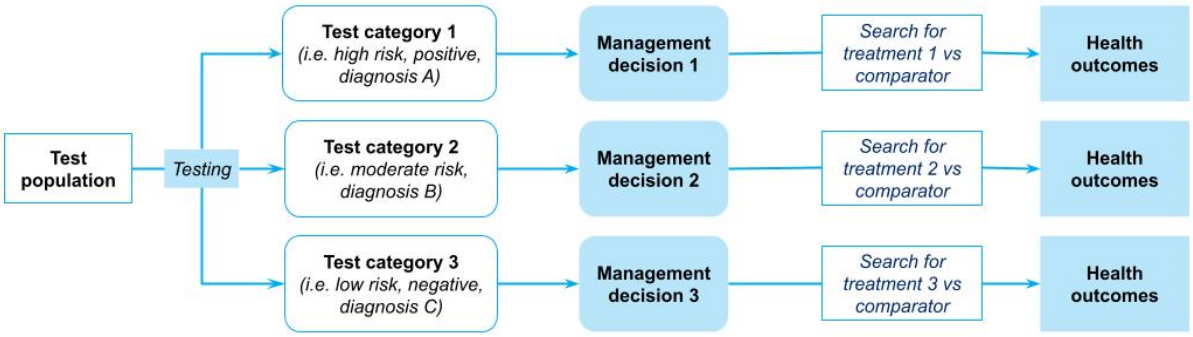


Figure 33 Demonstration of the requirement for multiple searches for the impact of the change in management

Examples of changes in management and the types of searches required to assess the harms/benefits of these include:

- If the proposed test provides an earlier diagnosis than the comparative test strategy, comparison of the benefits/harms of early versus late treatment would be appropriate. Search terms should include terms relevant for the target population and the treatment, as well as terms that relate to the timing of treatment (e.g. early, late, misclassified, delayed).
- If the proposed test reclassifies the stage of disease, influencing the type of interventions chosen, the effectiveness of different interventions for that stage of disease could be assessed (ideally comparing the treatment that the patient would have received in the absence of the proposed test classification). Search terms should include the target population (possibly with additional terms for stages of disease), and the treatment and/or comparator terms. If scoping searches find high-level evidence is available, searches could be limited to systematic reviews or RCTs.
- If the proposed test results in a patient receiving a different diagnosis than they would have received otherwise, then, if possible, the effectiveness of treatment for that disease should be compared against the treatment the patient would have received in the absence of that diagnosis. Limit searches to the highest level of evidence identified in scoping searches.
- If the proposed test results in the avoidance of invasive testing for some patients, search terms related to the harms of that subsequent test in a relevant population would be required.

In general, it is important to use judgement to determine and document relevant search strategies for assessing the impact of change in management. Although best practice would be to perform a systematic review to address the impact of the change in management, judgement may be used to determine whether existing systematic reviews are sufficient, or whether a rapid review of high-level evidence (e.g. using one database, with study design filters) may be appropriate. Preference should be given to higher quality and more recent evidence.

Search for information on the value of knowing

The methods for assessing the value of knowing are still in development. Studies that directly discuss the value of knowing results from the proposed test in the correct population may possibly be identified through the test-related search, as described above. However, bibliographic databases

with a psychological focus should be searched to supplement the medical bibliographic databases if the clinical claim relies on the value of knowing.

In the absence of directly relevant literature, strategies could include:

- broadening the type of evidence included to capture qualitative studies, opinion pieces, editorials, consumer input, etc.
- generalising from broader or other populations/interventions that are likely to have similar consequences.

Presentation of the search strategy

The clear presentation of the search terms used improves the transparency of the approach. Tabulating the search terms can assist with presentation (Table 27). Present a table of the search terms for each bibliographic database or data source, and for each search (if more than one search is performed). The presentation of the search terms must explain how each of the terms interacts with other terms (i.e. Boolean operators).

Table 27 Search terms for the literature review (investigative technology)

Category	Description	Search terms
Study design (if justified)	[insert description of category]	[e.g. Cochrane Highly Sensitive Search Strategies for identifying randomised trials in MEDLINE, or MeSH and text word terms for nonrandomised study designs]
Population	[insert description of category]	[include MeSH terms, text words and synonyms for the target population/disease/condition]
Intervention	[insert description of category]	[include known proprietary and nonproprietary names, MeSH terms]
Comparator (if required)	[insert description of category]	[include known proprietary and nonproprietary names, MeSH terms]

MeSH = medical subject headings

Sources of evidence

As a minimum, search the following sources:

- the published literature in bibliographic databases (at least MEDLINE, EMBASE and Cochrane library)
- registers of randomised trials
- HTA agency websites or the HTA database
- any unpublished studies on file (for an applicant developed assessment)
- reference lists of all relevant articles that are obtained.

The clear presentation of the sources searched improves the transparency of the approach. Tabulating the search strategies used can assist with presentation (Table 28).

The selection of data sources to search will be guided by the review topic. Include additional databases that may be relevant (e.g. PsycInfo for mental health literature).

In addition to bibliographic databases, trial registers and internal study reports from manufacturers/sponsors are an important source of relevant studies. Manually searching the reference lists of included studies (also called pearling or backward citation searching) may also identify relevant studies. Furthermore, potentially relevant studies may be found by searching for studies that have since cited an included study (forward citation searching);[96] this can be done by looking up the included study and searching the 'cited by' (Google Scholar) or 'times cited' (Web of Science) list for the study.

The methodological standards for the conduct of new Cochrane Intervention Reviews are an appropriate source of guidance for performing a high-quality systematic literature search.[97]

Table 28 Record of search strategies

Source	Date searched	Date span of search
MEDLINE (via PubMed)	[insert date]	[insert dates]
EMBASE (e.g. Embase.com)	[insert date]	[insert dates]
Cochrane Library[a]	[insert date]	[insert dates]
ClinicalTrials.gov	[insert date]	[insert dates]
International Clinical Trials Registry Platform[b]	[insert date]	[insert dates]
Australian Clinical Trials Registry	[insert date]	[insert dates]
Prospective Register of Systematic Reviews (PROSPERO)	[insert date]	[insert dates]
Internal registries	[insert date]	[insert dates]
Other (state other sources[c])	[insert date]	[insert dates]

a Includes the Cochrane Database of Systematic Reviews, the Cochrane Central Register of Controlled Trials and the Health Technology Assessment database

b International Clinical Trials Registry Platform[a]

c Report on the details of supplementary searches, including manual checking of the references in retrieved papers, searches of the Therapeutic Goods Administration dossier and searches of grey literature.

Study exclusion

Following the literature search, exclude studies that:

- describe an incorrect intervention (such as when the health technology is used beyond the use described in the requested MBS item descriptor)
- do not include the target population (not enough patients are enrolled who would be eligible for the proposed health technology according to the requested MBS item descriptor)
- do not make comparisons with the relevant comparator(s). This step is not relevant if comparative evidence is not identified and noncomparative studies are required.

a www.who.int/ictrp/en

- do not report a relevant outcome.

After studies have been excluded on the basis of the exclusion criteria (such as the PICO criteria mentioned above, incorrect study type, language or publication year), consider excluding studies on the basis of quality. This approach may be difficult to justify if the number of relevant studies included is not large. Studies excluded on the basis of quality (or other reasons, such as being unable to extract data) should be presented separately from those that are excluded on the basis of not meeting the PICO criteria.

For large reference lists that include a variety of study designs and qualities, identify the studies that represent the highest quality evidence and determine whether they are adequate to answer the assessment question. Where the higher quality studies are inadequate, include lower quality studies to supplement the evidence base. Quality may relate to study design, conduct or size.

Describe and justify the stepwise approach taken to exclude studies. Studies that are otherwise eligible for inclusion (but excluded due to study design, conduct or size) that contradict the results of the included RCTs should be identified and discussed. If randomised studies are included, consider providing results from large comparative observational studies as supplementary evidence.

Published systematic reviews and meta-analyses

When assessing a therapeutic intervention, it is preferable to extract individual studies from published meta-analyses and compare each study against the study selection criteria. Exclude any studies that do not meet the criteria. Discuss the decision to include the treatment effect from a published systematic review.

When assessing a test and searching for the possible (health) impact of change in management, systematic reviews and meta-analyses are often included. During this step of the linked evidence approach, a rapid review is often performed and preference is given to higher quality and more recent evidence. Therefore, in this situation, it may be acceptable to extract the results of the systematic review without extracting data from the individual studies.

PRISMA flowchart

For an investigative technology, if a linked evidence approach is taken, consider whether a single or multiple PRISMA flowchart(s) is necessary. Typically, at least 2 separate searches will be required for a linked evidence approach (one to capture test-related articles and one to capture the impact of change in management), and it may be more appropriate to present separate PRISMA flowcharts for each search. However, it may be appropriate to use a single PRISMA flowchart for presenting all test-related included studies. If this is done, present in the flowchart the number of studies included for each assessment question (diagnostic accuracy, predictive accuracy, concordance, safety, change in management etc.).

An adapted PRISMA flowchart for presenting the screening process in MSAC assessment reports is shown in Figure 34.

The adapted PRISMA flowchart has a 3-step process for study selection, in which studies are excluded:

1. based on title and abstract, or when the article cannot be retrieved
2. after retrieval of full-text articles

3. based on clearly specified reasons other than the exclusion criteria described in the 'Study exclusion' section above. Provide justification for each exclusion at this point.

Clearly depict the reasons for study exclusion in the PRISMA flowchart.

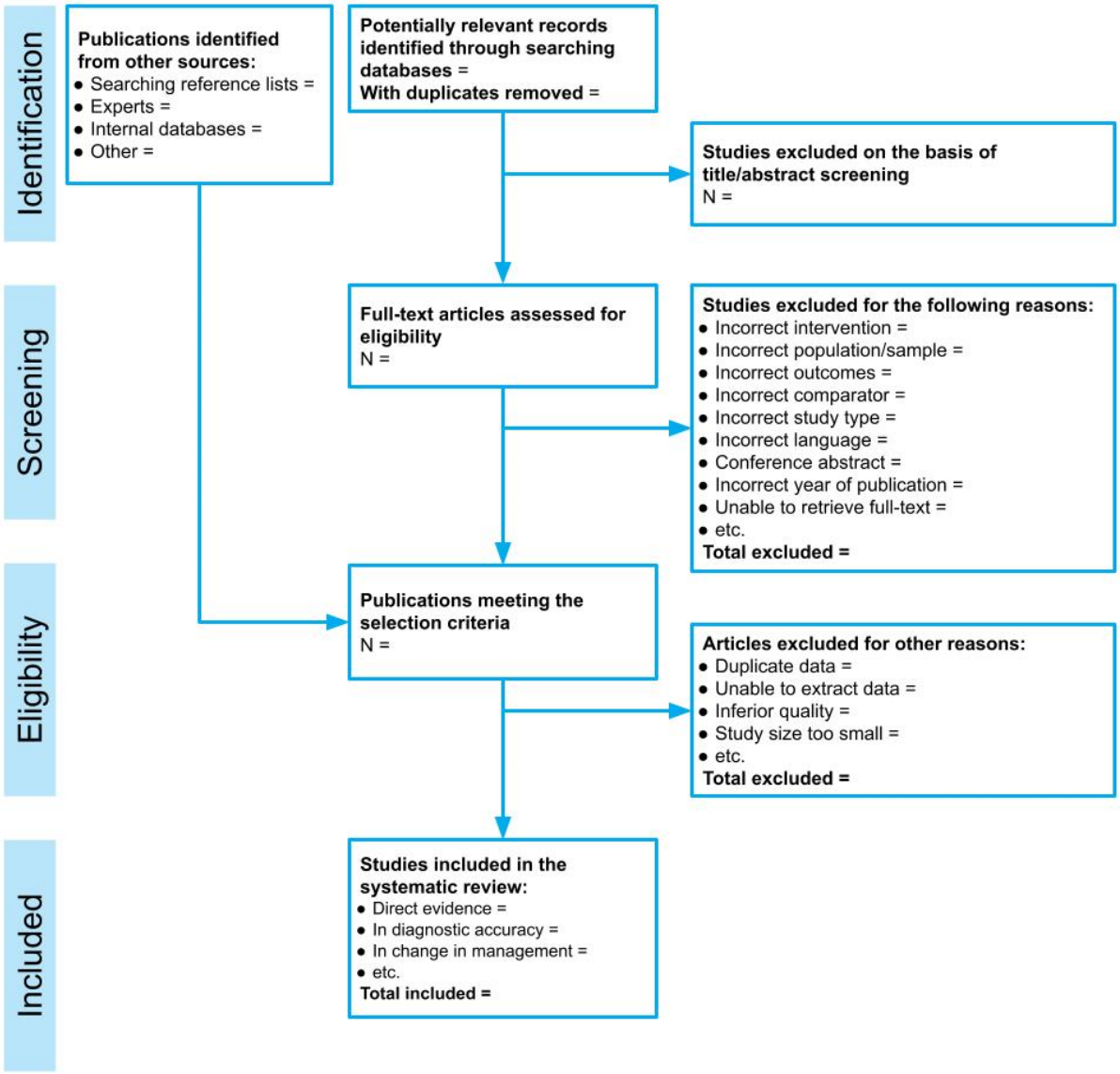


Figure 34 **Adapted PRISMA flowchart for presenting screening of studies for MSAC assessment reports**[98, 99]

Copies of included studies

For assessment reports that will undergo a commentary route (i.e. be reviewed by an independent review group) prior to consideration by MSAC, to facilitate the critique provide full-text copies of all the included studies. For assessment reports contracted by the Department of Health, provide full-text articles from 3 key studies.

If internal reports (commonly for manufacturer-led studies) have been included, the full study report is required.

Provide reputable translations of trial reports that are published in languages other than English.

Appendix 3 Risk of bias

The objective of Appendix 3 is to describe appropriate approaches for considering the risk of bias in the studies identified in the assessment report.

Bias is a deviation from the true underlying effect of an intervention as a consequence of issues in study design, study conduct, data collection, data analysis, interpretation of the results, reporting or publication.[100]

The key purpose of assessing the risk of bias of the included studies is to:

- provide MSAC with a clear idea of which studies are of greater scientific rigour
- assist in the discussion and interpretation of the results.

All studies that are included in the assessment report for the purpose of answering the PICO assessment questions should be assessed for risk of bias.

While the assessment of bias may culminate in a summary statement for each study (i.e. low or high risk of bias), this is not the sole or most important output from the assessment of risk of bias. The assessment of the risk of bias assists in identifying key issues that may have affected the treatment effect observed in the studies. These issues are then raised during the interpretation of the synthesis of the evidence.

The choice of tool for assessing risk of bias should be appropriate for the study design, and should be published, structured and (ideally) validated. A list of risk of bias tools developed for different study designs is included in Table 29. These are intended as examples of tools that are commonly used, and not to state what should be used in the assessment. Many risk of bias tools do not differentiate the possible impact of bias on different outcomes. For example, subjective outcomes may be more susceptible to unmasking or open label designs than objective outcomes. When considering the risk of bias in a study, it is important to consider it in the context of the impact the bias might have on the outcomes of interest. This may lead to different judgements of risk of bias across different outcomes in the same study.

Risk of bias may also be assessed for qualitative studies and ethical analyses, if these studies are likely to be important for decision-making.

Table 29 Common risk of bias tools used for different study designs

Study type	Applicable risk of bias assessment tools	Link/reference
Systematic reviews	ROBIS (2016)	www.bristol.ac.uk/population-health-sciences/projects/robis
	AMSTAR-2 (2017)	amstar.ca
	NHLBI systematic review checklist	www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
Randomised controlled trials	Cochrane Risk of Bias 2.0 Tool (2019)	www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2

Study type	Applicable risk of bias assessment tools	Link/reference
	SIGN checklist for RCTs (2014)	www.sign.ac.uk/what-we-do/methodology/checklists/
	NHLBI controlled intervention checklist	www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
Nonrandomised studies	ROBINS-I (nonrandomised studies of interventions) (2016)	www.riskofbias.info
	Newcastle-Ottawa Scale (NOS) (1999)	www.ohri.ca/programs/clinical_epidemiology/oxford.asp
	Downs and Black checklist (1998)	Downs S and Black N (1998) 'The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions', *Journal of Epidemiology and Community Health*, 52(6):377–384.
Nonrandomised studies – cohort studies	SIGN checklist for cohort studies (2014)	www.sign.ac.uk/what-we-do/methodology/checklists/
	NHLBI cohort and cross-sectional checklist	www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
Nonrandomised studies – case-control studies	SIGN checklist for case-control studies (2014)	www.sign.ac.uk/what-we-do/methodology/checklists/
	NHLBI case-control studies checklist	www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
Test performance/ diagnostic accuracy	QUADAS-2 (2011)	www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2
Prognostic studies	QUIPS tool (2013) [PDF, 389KB]	https://bmjopen.bmj.com › content › embed › inline-supplementary-material-3
Prognostic and predictive prediction models	CHARMS checklist (2014) – prediction models	Moons K, de Groot J, Bouwmeester W, Vergouwe Y, Mallett S, Altman D, Reitsma J and Collins G (2014) 'Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist', *PLoS Medicine* 11(10):e1001744. doi.org/10.1371/journal.pmed.1001744
Case series	IHE quality appraisal checklist for case series	IHE (2016). Institute of Health Economics: quality appraisal checklist for case series studies.

Study type	Applicable risk of bias assessment tools	Link/reference
	studies (2016, by Moga et al.)	www.ihe.ca/publications/ihe-quality-appraisal-checklist-for-case-series-studies Moga C, Guo B, Schopflocher D and Harstall C (2012) *Development of a quality appraisal tool for case series studies using a modified Delphi technique*, Institute of Health Economics, Alberta, Canada.
	NHLBI case series checklist	www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools
	NHS-CRD case series quality assessment scale (2001)	Khan K (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*, Centre for Reviews and Dissemination, University of York, England.
Qualitative studies	CASP checklist (2018)	casp-uk.net/casp-tools-checklists
	JBI Checklist for Qualitative Research (2017)	jbi.global/critical-appraisal-tools
Ethical analysis	Q-SEA (2017)	Scott A, Hofmann B, Gutierrez-Ibarluzea I, Lysdahl K, Sandman L and Bombard Y (2017) 'Q-SEA – a tool for quality assessment of ethics analyses conducted as part of health technology assessments', *GMS Health Technology Assessment* 13(Doc02):1–9.

AMSTAR = A MeaSurement Tool to Assess systematic Reviews; CASP = Critical Appraisal Skills Program; CHARMS = CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; IHE = Institute of Health Economics; JBI = Joanna Briggs Institute; NHLBI = National Heart, Lung, and Blood Institute; NHS-CRD = National Health Service – Centre for Reviews and Dissemination; Q-SEA = Quality Standards for Ethics Analyses; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; QUIPS = Quality in Prognostic Studies; RCT = randomised controlled trial; ROBINS-I = Risk Of Bias In Non-Randomised Studies – of Interventions; ROBIS = Risk Of Bias In Systematic reviews; SIGN = Scottish Intercollegiate Guidelines Network

Regardless of the study design, the following aspects are important to consider when assessing risk of bias:

- Possible bias in selecting the study population – The population included in the study should be appropriate and consist of a representative spectrum of participants. If the selection of participants is inappropriate, this could lead to (among other things) spectrum bias or selection bias.
- Validity of methods for outcome measurement and reporting – If a measure does not produce consistent findings, you cannot rely on the results. Some outcome measures are consistent, yet do not provide accurate/valid results. When outcomes are measured subjectively, and/or by surveys or observations, these results can be susceptible to recall bias, response bias, detection bias, verification bias, clinical review bias, diagnostic review bias and/or test review bias. Furthermore, selective reporting of outcomes/results will also introduce bias to the body of evidence.

- The applicability of intervention and study setting – The intervention and the comparator used in the study should be representative of the target intervention and the way this intervention is proposed to be used in Australia. If the intervention is used in a different setting (e.g. primary care instead of secondary care), the generalisability of the results of the study is questionable.

The best approach to assessing the risk of bias in individual studies will depend on the design of the study. Justify the approach taken (or modifications to the approaches below) to capture the key limitations of the study design.

Summarise the risk of bias across studies in a graphical or tabular form (e.g. a heat map or traffic light table that highlights the deficiencies of the evidence; see example in Table 30). Tools are available to assist with this, such as *robvis* (visualisation tool for risk of bias)[a], which can be used with Cochrane RoB 2, ROBINS-I and QUADAS-2 risk-of-bias tools.

Table 30 Example presentation of completeness of reporting of systematic reviews (AMSTAR 2)

Author	Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Reviews with meta-analyses																	
Smith	2011	M	M	N	P	N	N	N	P	P	N	N	N	N	N	N	M
Johannes	2019	P	N	N	N	M	M	P	N	N	M	N	N	N	N	N	N
Wong	2015	P	P	N	P	M	M	M	M	M	M	M	M	M	M	M	M
Reviews without meta-analyses																	
Eberle-Sejari	2020	P	N	N	P	N	N	N	M	M	M	–	–	–	N	–	N
Mitra	2017	N	N	N	N	N	N	N	N	N	N	–	–	–	N	–	N
Liu	2018	P	P	N	P	P	N	M	M	M	N	–	–	–	N	–	M

Domains: 1 = PICO; 2 = protocol; 3 = study design; 4 = search strategy; 5 = study selection; 6 = data extraction; 7 = excluded studies; 8 = included studies; 9 = risk of bias; 10 = funding source; 11 = meta-analysis 12 = impact of risk of bias; 13 = discussing risk of bias; 14 = heterogeneity; 15 = publication bias; 16 = conflicts of interest

Colour and letter coding: white = methodological requirements met (M); grey = methodological requirements partially met (P); black = methodological requirements not met (N); – = not applicable (no meta-analysis)

Systematic reviews and meta-analyses

The approach for assessing risk of bias in systematic reviews depends on how the included systematic review is used in the assessment report:

- If the published systematic review is included in its entirety, and the assessment report relies on a pooled result from the review, assess the quality of the review using a validated tool for systematic reviews. Report the methods used by the authors of the systematic review to assess risk of bias for the included studies.
- If individual studies in the systematic review are retrieved and used, or the systematic review is 'broken up' such that some studies are excluded, it is preferable to assess the risk of bias for

a mcguinlu.shinyapps.io/robvis

individual studies using a tool relevant to the study design. Using the risk of bias tables provided with a published systematic review is acceptable. Report when this approach is taken.

Only include systematic reviews (rather than individual studies within systematic reviews) if the review is of adequate quality and applicability to the assessment question. Justify the judgements regarding the quality and applicability.

Specific aspects of assessing risk of bias in systematic reviews are described in a publication by Shea et al.[101] and include whether the systematic review authors 1) have *a priori* agreed on review methods in a protocol, 2) have used a satisfactory technique for assessing the risk of bias of individual studies in the systematic review, 3) have reported any funding sources and potential conflicts of interest, 4) have investigated possible causes of heterogeneity, and 5) have carried out an adequate investigation of publication bias and discussed its impact.

Randomised controlled trials

The assessment of the risk of bias of randomised controlled trials (RCTs) is based on factual information about the design and conduct of the study, such as if and how the participants were allocated to groups, or whether or not participants or assessors were blinded. The 5 domains included in the Cochrane risk of bias tool are 1) bias arising from the randomisation process, 2) bias due to deviations from intended interventions, 3) bias due to missing outcome data, 4) bias in measurement of the outcome, and 5) bias in selection of the reported result.

It is important to consider the flow of participants through the included RCT. Consider the impact on the observed (treatment) effect of patients who are lost or discontinued at any point in the study. Tabulating the points at which patients discontinued or were lost to follow-up may assist in identifying potential biases associated with attrition.

As a minimum, data extraction should include the analysed patients as a proportion of the patients enrolled into the study *by study arm.* Differential losses to follow-up should be noted and incorporated in the interpretation of the synthesis of the data.

If a randomised trial is available for assessing an investigative technology, the effectiveness of a test usually depends on the patient and/or clinician knowing the result of the test. In many cases, blinding of allocation to different test arms is not possible. This may be acceptable, and in some circumstances preferable, as subsequent management decisions will be made in clinical practice with knowledge of the test that has been used to derive the results. The study should therefore not be rated down for risk of bias due to lack of blinding.

Nonrandomised studies

Nonrandomised studies have a higher risk of bias than randomised studies. Nonrandomised studies would usually be included in several steps of the linked evidence approach for assessment of an investigative technology (e.g. to determine the change in management due to an intervention). Methods for mitigating the risks associated with the differential distribution of known confounders because of nonrandom assignment (such as matching and controlling for confounders in the analysis) cannot adjust for the differential distribution of unknown confounders. If high-quality randomised studies are available and form the basis of the assessment report, it may not be necessary to consider the risk of bias of the identified nonrandomised studies.

The internal validity of a nonrandomised study can be elicited by reference to how the study design or conduct differs from that of a well-designed, double-blind RCT. Bias in a nonrandomised study is

defined as the systematic difference between the results of the nonrandomised study and the results expected from the ideal double-blind RCT. Potential sources of bias that are considered in the ROBINS-I checklist include:[102] 1) bias due to confounding, 2) bias in selecting the study population, 3) bias in the classification of interventions, 4) bias due to deviations from intended interventions, 5) bias due to missing data, 6) bias in measurement of outcomes, and 7) bias in selection of the reported results.

Studies on test accuracy/diagnostic accuracy

Some quality assessment tools have been specifically designed to assess the risk of bias in test accuracy studies (e.g. QUADAS-2, see Table 29). These studies would be included in the diagnostic accuracy step of the linked evidence approach in an assessment of an investigative technology. Important aspects for assessing risk of bias of accuracy studies as per QUADAS-2 are:[23] 1) participant characteristics and recruitment (including spectrum bias), 2) applicability of the index test, 3) validity of the reference standard (e.g. misclassification of the target condition), 4) blinding (includes test review bias, diagnostic review bias and clinical review bias), and 5) patient flow (includes verification and attrition bias).

Studies on prognosis

At least one risk of bias tool is available specifically for prognostic studies (the QUIPS tool). For the assessment of risk of bias in prognostic studies, 6 domains are included in the QUIPS tool:[103] 1) study participation, 2) study attrition, 3) prognostic factor measurement, 4) outcome measurement, 5) study confounding, and 6) statistical analysis and reporting.

Studies without a control group/case series

Case series are uncontrolled studies, and are therefore considered one of the weaker study designs from which to obtain evidence on the effectiveness of an intervention. However, case series have been increasingly included in HTAs due to absence of higher quality evidence, especially in assessments of investigative technologies. Studies that do not have a separate control group may provide relevant information (e.g. about how patients benefit from testing or treatment, or about intervention safety) through the use of before-and-after data.

When assessing therapeutic interventions, if patients are selected for inclusion in the study based on the severity of symptoms, there is the risk of regression to the mean. The second measurement will be closer to the population mean than the first measurement, and could be misinterpreted as being attributable to the intervention.[40] Outcomes that have a high degree of random variability (such as blood pressure) are most susceptible to regression to the mean phenomenon. The problem is often made worse when there are substantial ceiling and floor effects, which is the case in many common quality of life scoring instruments.[40]

Case series may not perform a before-and-after comparison. Evidence of a comparative effect that is derived through a naive comparison with another intervention is considered very low quality. Much of the uncertainty of this approach relates to the potential confounding in the naive comparison. However, there are some key considerations regarding the methodological quality of a case series that may influence the confidence of the findings: 1) patient selection, 2) (in)adequate ascertainment of exposure/outcome, 3) causality, and 4) reporting.

While the 'internal validity' of a case series may be reasonable, an estimate of the incremental treatment effect of a therapeutic intervention in a case series is only possible using a naive

comparison with a study of the main comparator(s), or of the natural history of the disease. Comparisons of this type are highly susceptible to confounding.

In randomised studies, confounders are usually balanced across arms. However, known and unknown confounders of intervention performance are likely to be imbalanced across separate case series. Potential confounders are discussed in Appendix 6.

A clear discussion of the potential confounding associated with a naive comparison should be provided during the interpretation of any results that are derived from a naive comparison of case series, or a comparison of a case series with the natural history of the disease.

Other study designs: qualitative studies, prognostic and predictive prediction models, and ethical analyses

Tools are available to assess the risk of bias for other study designs that do not provide a quantitative estimate of the clinical utility of the health technology; for example, qualitative studies and ethical analyses. Assessment groups should determine whether a bias assessment for these studies will assist MSAC in their decision-making (based on perceived importance of this evidence).

For quality assessment tools specifically designed to assess risk of bias for specific study designs, see Table 29.

Appendix 4 Certainty of the evidence (GRADE)

Each assessment should consider the overall quality of the evidence base (in addition to a separate assessment for each included study, as discussed in Appendix 3). This enables the conclusions for each efficacy and safety outcome to be weighed in terms of the strength of evidence across all the studies that reported the outcome in question. This feeds directly into the evidence synthesis, such that it is clear that the given outcome having a risk ratio of 'x' was based on k number of included studies, with N number of patients and characterised by (for example) reasonable consistency between trials or other key features. An overall assessment of the quality of the evidence for each outcome is necessary to indicate the certainty that MSAC may have that the evidence represents the 'true' effect of the health technology.

The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach is one option that authors may choose to assess the quality of evidence.

Overview of the GRADE approach for therapeutic technologies

The GRADE approach gives an overall measure of confidence in a study result (high, moderate, low or very low, indicated by a traffic light value), representing the quality of the evidence, for each critical outcome. GRADE requires an assessment of the following domains to rate the *quality* or *certainty* of the body of evidence:

- study design[104]
- risk of bias or study limitations[105]
- imprecision[106]
- inconsistency of results[107]
- indirectness of evidence (applicability of the population, intervention, comparator and outcomes)[108]
- publication bias.[109]

Assessment authors initially describe the study design (randomised controlled trials start with a high rating, observational studies with a low rating), and then rate the evidence down for weakness in any of the above domains. Alternatively, the evidence may be rated up when:[12]

- there is a large magnitude of effect
- there is a dose-response gradient
- all plausible confounders or other biases increases confidence in the estimated effect.

Following the assessment of individual study results and meta-analysis (if appropriate), an evaluation of the imprecision, inconsistency, indirectness and risk of publication bias across the evidence base (per outcome) should be performed. For each critical efficacy and safety outcome identified in the PICO confirmation, discuss the overall strength of the evidence base, noting the number of trials (k) that provide direct from test to health outcomes evidence and the corresponding number of participants. An example of a simplified GRADE table is shown in Table 31.

Table 31 GRADE table for critical and important outcomes

Quality assessment for [patient relevant outcome #1]						
No. of studies (Design)	**Limitations (ROB)**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Certainty**
k [design] k2 [design]						High ⨁⨁⨁⨁ Moderate ⨁⨁⨁◯ Low ⨁⨁◯◯ Very low ⨁◯◯◯

k = number of studies; ROB = risk of bias

⨁⨁⨁⨁ Very confident that the true effect lies close to that of the estimate of the effect.
⨁⨁⨁◯ Moderate confidence in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

⨁⨁◯◯ Confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.

⨁◯◯◯ Very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

Indirect or surrogate outcomes (not patient relevant) should only be included in the GRADE tables where direct from test to health outcomes evidence is inadequate and the supporting evidence from a surrogate endpoint (e.g. antibody titre) provides valuable context for interpreting the more limited patient-relevant outcome data (e.g. pathology-confirmed cases). When a surrogate outcome is included for a therapeutic technology, it should be rated down for indirectness of the outcome measure.

According to the GRADE approach, the directness domain includes applicability.

For certain outcomes, evidence from the included trials may be minimal or absent (e.g. long-term outcomes such as overall survival, certain low frequency severe adverse events). If such outcomes have been identified as critical, these should be included within the limits of the evidence. It may not be possible to provide an effect estimate, and the traffic light value may be very low or nil (if no estimate or evidence was available). This applies to the outcomes that are relevant, not just those with available evidence.

Multiple GRADE tables may be needed for an evaluation that addresses more than one population or different applications of an intervention (e.g. monitoring as well as diagnosis). More detail on how to apply the GRADE approach, as well as alternative formats for different types of evidence, can be found in the GRADE series of papers introduced in *Journal of Clinical Epidemiology*, volume 64 (2011) and references therein, and in the GRADE Handbook.[a]

Table 32 is an example of a basic summary of findings table. Examples from published systematic reviews can be found in the GRADE Handbook. Additional information for investigative technologies is in the next section.

a gradepro.org/resources/#handbook

Table 32 Basic summary of findings table

Summary of findings Comparison of intervention X with intervention Y in patients with Z condition and ABC clinical features						
Outcome (duration of follow-up)	**Number of studies (k), study design; patients (n)**	**Intervention absolute effect (RD) [95% CI]**	**Comparator absolute effect (RD) [95% CI]**	**Relative effect (RR) [95% CI]**	**Certainty**	**Comments**
Effectiveness outcome 1	…	…	…	…	⨁⨁⨁⨁ ⨁⨁⨁◯ ⨁⨁◯◯ ⨁◯◯◯ ◯◯◯◯	…
Effectiveness outcome 2						
Safety outcome 1 etc						

RD = risk difference; RR = relative risk

Specific issues for investigative technologies

Direct from test to health outcomes evidence

If direct from test to health outcomes evidence is available to assess an investigative technology, the standard GRADE approach (developed for therapeutic technologies) can be used, with consideration of how additional applicability issues are addressed (see Technical Guidance 10). Randomised trials, or systematic reviews of randomised trials, are the highest level of evidence, and any outcome measures that are not directly patient relevant are rated down for indirectness.

Any harms associated with testing or downstream consequences may be assessed as per the standard GRADE approach.

Linked evidence of clinical utility

Singh et al. (2012) recommend that, since most evidence for efficacy of investigative technologies is indirect, authors should consider the quality for each link in the evidence framework, or justify why this has not been considered.[110] For a linked evidence approach, adaptations of the GRADE process vary depending on the component of the linkage.

Elements that could affect the quality of the evidence could include variability in characteristics of the test population, differing test result definitions that lead to a change in management and differing options for treatment based on the same test result.

Change in management and health outcomes

For evidence on change in management and therapeutic efficacy, the standard GRADE approach for therapeutic technologies can be used, with the amendment that the outcomes are not rated down for indirectness, despite not directly informing the question of the clinical utility of the test.

The outcomes may be rated down for indirectness due to other reasons (e.g. indirectness of outcome if a change in diagnosis is reported, rather than a change in management; if outcomes reported are management recommendations, not management received; or indirectness of population for therapeutic efficacy, if the spectrum of patients treated does not match those who are likely to be treated based on test results).

Cross-sectional accuracy

Using the GRADE approach for questions regarding the diagnostic accuracy of a test are well established, with GRADEpro software having an option for diagnostic questions. Evidence for diagnostics usually includes studies of widely variable design, typically with few or no randomised controlled trials that adequately address *both* the tested and the treated populations. The study designs considered 'high' quality are cross-sectional studies (cohort type accuracy studies), whereas case-control accuracy studies are considered 'low' quality.

Assessment should consider both the quality of evidence as it relates to test accuracy, as well as the proportion of patients with unevaluable results or inconclusive results (who may need to have a second sample retrieved). The GRADE approach has been developed for the purpose of clinical practice guideline developers, and has therefore been targeted towards assisting clinicians to understand the strength of evidence for guiding treatment of individual patients. The emphasis has therefore been on the outcomes of true positives, true negatives, false positives and false negatives. For the purpose of population-based funding decisions, it is suggested that the more relevant outcome measures (which are less susceptible to pre-test probability differences) are sensitivity and specificity.

A summary of findings table adapted for diagnostic interventions is described in the GRADE Handbook[a] and adapted in Table 33.

Table 33 Summary of findings for diagnostic accuracy

Summary of findings – test important outcomes **Comparison of test X with test Y in patients suspected to have Z condition**						
Test outcome	**Number of studies (k); patients (n)**	**Intervention result**	**Comparator result**	**Absolute difference**	**Quality**	**Comments**
Sensitivity	…	% (range)	% (range)	…	⨁⨁⨁⨁ ⨁⨁⨁⊙ ⨁⨁⊙⊙ ⨁⊙⊙⊙ ⊙⊙⊙⊙	…
Specificity	…					
Unevaluable	…					
Inconclusive results	…					

a gradepro.org/resources/#handbook

Longitudinal accuracy

Tests that are performed to determine future patient outcomes, rather than current status, are considered to provide 'predictive accuracy' (see TG 11.5). This type of evidence requires adapting the standard GRADE process. Instead of considering RCTs as the ideal study design (as per therapeutic technologies) or cross-sectional data (as considered best for diagnostic accuracy studies), predictive accuracy is best assessed using prospective longitudinal data to confirm actual versus predicted patient outcomes. Prospective cohort studies should therefore be given a 'high' rating, whereas other study designs should be given a 'low' rating. If the GRADEpro software is used, select the study design 'randomised trial' if the evidence identified is prospective cohort studies, and 'observational study' for all other study designs. The results columns will need to be adapted to suit the information received.

Qualitative evidence

The standard GRADE approach cannot be applied to qualitative evidence. If evidence synthesis is based on qualitative research (e.g. for demonstrating value of knowing), quality assessment may be done using the GRADE-CERQual[a] approach,[111, 112] and as recommended by the Cochrane Collaboration. Instead of considering the confidence in an overall effect estimate, authors are asked to consider whether the review finding is representative of the phenomenon of interest. This is based on the following elements:

- the methodological limitations of the qualitative studies contributing to a review finding
- the relevance to the review question of the studies contributing to a review finding
- the coherence of the review finding
- the adequacy of data supporting a review finding.

A fifth element to address publication (dissemination) bias is in development.[113] Note that this approach does not address the limits of individual qualitative studies nor the methodology used for the evidence synthesis. Further information is on the GRADE-CERQual website.[b]

a CERQual = Confidence in the Evidence from Reviews of Qualitative research

b www.cerqual.org

Appendix 5 Study characteristics

This appendix discusses the consideration of the characteristics of the included studies. These characteristics include:

- the populations enrolled in the study
- the health technologies and circumstances of use
- the definitions and timing of outcome measures
- the statistical approaches used for reporting outcomes.

There are 2 key purposes for considering the characteristics of the included studies. Firstly, a comparison across arms and across studies will identify potential confounders or sources of heterogeneity, and inform the interpretation of the synthesis of the evidence. Secondly, the characteristics of the included evidence can be compared against the proposed use of the health technology in the Australian setting to assess the applicability.

The presentation of characteristics of the included studies will vary depending on the type of evidence and the volume of studies. For assessments that require multiple steps (such as a linked evidence approach for investigative technologies), a separate comparison of characteristics is required for each step.

For evidence presenting a direct comparison of a health technology with an appropriate comparator (such as a randomised controlled trial), it is preferable to present the characteristics of the studies in a tabulated format to enable comparison across arms and across studies. A comparison of the following characteristics may assist in identifying possible sources of heterogeneity and/or confounding:

- study design or conduct
- patient and disease characteristics
- eligibility criteria
- description of the interventions
- outcome definitions.

The level of detail that is appropriate and achievable may depend on the number of studies included in the assessment report. For assessment reports based on large numbers of included studies, this approach may not be feasible to undertake for a large number of characteristics. Focus on key participant, treatment and setting characteristics that are informative to decision-making. In cases where a large number of studies is included, the following approach may be appropriate:

1. Identify key characteristics (study, patient baseline or disease characteristics) that may have an impact on the comparative performance of the health technology (see Appendix 6).
2. Report key characteristics for each study in the study data extraction tables.
3. Tabulate the key characteristics of the included studies to enable a comparison across studies.
4. Highlight important differences in the key characteristics (where the differences have a likely or possible impact on the treatment effect).
5. Focus on key characteristics that will are informative to decision-making.

For assessment reports that have included a systematic review, it may not be necessary to present the study characteristics for individual studies. Present key characteristics of the populations, treatments and outcomes included in the systematic review. If individual studies are extracted from the systematic review to answer one or more assessment questions, the characteristics of the individual studies are required.

Participants

For each of the included studies, consider the following details about the study participants:

- eligibility criteria for participants considered for recruitment into the study
- baseline demographic and clinical characteristics for each group or study arm
- median duration (and range) of follow-up for each group and for the entire study (also indicate whether the study is ongoing).

Identify characteristics that may have an impact on the target outcomes. These may be identified from a background search of the literature, or by reviewing subgroup analyses of the included studies.

Important characteristics should be extracted from included studies and reported in study data extraction tables. Do not report extensive characteristics of included studies that are unlikely to affect the interpretation of the evidence.

For studies with high losses to follow-up (discontinuations, withdrawals or other causes of censoring), or differences across arms in terms of the extent or timing of losses to follow-up, compare the characteristics of the patients who were censored from the analysis with those who remained in the study. Whether information about this subgroup is or is not presented, consider the impact of censoring or loss to follow-up, particularly when it is differential across study arms.

The report results of comparisons of participant and disease characteristics should include:

- a summary of the key characteristics that may impact the treatment effect (regardless of whether there are differences between studies)
- key differences in these characteristics across arms within studies
- key difference in these characteristics across studies
- key differences in the characteristics presented in the included studies and the target population (particularly if there is an important subgroup that will access the proposed health technology in the Australian setting that is not represented in the included studies).

Where the assessment report is based on a subgroup of an included study, it is important to compare the baseline characteristics for the relevant subgroup as well as the whole study population. Discuss whether the selection of the subgroup has increased the risk of bias associated with the comparison of the proposed health technology and the main comparator.

Health technology details

Differences in the use of the proposed health technology or comparator in different studies may result in heterogeneity of the observed results across studies. Consider the following details about the health technologies provided in each study:

- how the health technology and the main comparator were defined and delivered – Important characteristics for interventions may involve dose, frequency/episodicity, duration, need for subsequent treatments, the line of therapy and concomitant treatments. Important characteristics for tests may involve the timing of the test, sample details and thresholds applied.
- criteria for concomitant or subsequent intervention or confirmatory testing
- settings in which the health technology and main comparator are used.

Outcomes

Outcome definitions can differ across studies and may result in heterogeneity of the observed results. The following aspects relating to outcomes should be considered, and extracted, for each of the included studies:

- the primary outcome (or state if no primary outcome has been nominated)
- secondary outcomes that were identified in the PICO confirmation.

For each outcome:

- identify the definition of the outcome
- state the units of measurement and the method of statistical analysis
- describe the population in which the analysis is performed (i.e. intention to treat, per protocol)
- describe the timing of the outcome assessment and who performed the assessment
- describe the instrument used to measure the outcome (e.g. questionnaire, response evaluation criteria, blood test), and state whether it has been validated
- state how missing data were dealt with (it is important to address both patients who remain in follow-up who have not yet experienced an event, as well as those that were removed from the analysis).

Outcome measures may appear similar across studies; however, they may be influenced by covariates used in statistical approaches, and by censoring rules. State whether censoring applied in the study is appropriate or may be informative. When recording the method of statistical analysis, include the name of the statistical test and sufficient details to allow MSAC to ascertain how the analysis was performed.

Composite outcomes

A composite outcome is one in which multiple endpoints are combined. It is usually defined as having been experienced when the first of any of the component endpoints is experienced, even though subsequent component endpoints may occur.

For assessment reports that include composite outcomes, additional details relating to the definition of each composite outcome should be considered and reported. The assessment report should discuss:

- the definitions of the individual components in the composite outcome
- the clinical importance of each of the components
- whether the composite outcome was explicitly prespecified
- whether the composite outcome can be disaggregated, or whether disaggregation is not possible due to censoring that occurs following the first event in a composite outcome.

The interpretation of a composite outcome should consider which of the components is driving the composite outcome, and whether this is similar across arms.

Patient-reported outcome measures

Patient-reported outcome measures include generic ('global') or condition-specific (e.g. for respiratory conditions, depression, arthritis) measures of quality of life, symptoms or function.

Patient-reported outcome measures may also include multiattribute utility instruments (MAUIs), in which the scoring method for the instrument is anchored on a quality-adjusted life year scale of 0 (death) to 1 (full health). Several commonly used MAUIs for which a detailed discussion of the validity or reliability is not required are the Health Utilities Index (HUI2 or HUI3), the EQ5D-3L or -5L ('EuroQol'), the SF-6D (a subset of the Short Form 36, or SF-36), the Assessment of Quality of Life (AQoL) instruments, and the Child Health Utility 9D (CHU9D) index for children and adolescents.

An assessment report should describe the patient-reported outcome measurement, and state whether it is validated for use in the proposed population, condition and interventions. Describe the timing of and the personnel who administered the assessment.

Missing data is an important consideration for all outcomes; however, it is common in patient-reported outcome measures. The assessment report should consider compliance with the patient-reported outcome measures, and whether compliance (particularly when different across study arms) may have affected the comparison across arms. Describe any methods used to adjust for response bias (or methods for adjusting for missing data).

Minimal clinically important difference

The definition of a minimal clinically important difference (MCID) varies across the literature. The central concept of an MCID is that it represents the smallest amount of difference in a score that would, in some way, be considered important. MCIDs may be reported in either relative or absolute measures.

When selecting an MCID, it is important that the source of the MCID is relevant for the population, disease and interventions included in the assessment report.

Likely sources for an MCID may be:

- study protocols (often for the purposes of powering the study)
- an MCID previously accepted by MSAC that is relevant to the study population and the proposed indication

- a commonly accepted MCID in the literature, relevant to the study population and the proposed indication
- a commonly accepted MCID in the literature for a similar indication that can reasonably be expected to be generalisable to the proposed indication.

The derivation of an MCID for a dichotomous outcome (e.g. haemorrhage or no haemorrhage) or time-to-event outcome (e.g. overall survival) is not straightforward and an MCID may not be available. The most common approach for determining a meaningful benefit to patients involves a consensus of clinical experts in the relevant fields.

The application of an MCID to a surrogate outcome requires a rigorous explanation. The MCID for the surrogate should reflect a minimal important difference in the target patient-relevant outcome. The application of an MCID to a test accuracy outcome (e.g. sensitivity or specificity) is not appropriate. The translation of a test result to change in management and the eventual impact on health outcomes (incorporating both test-negative and test-positive patients) is complex and not commonly quantified outside of decision analytics.

The interpretation of the results in the context of a nominated MCID can be difficult if only aggregated data are provided. Typically, an MCID reflects the average minimal difference in a score that is considered to be clinically important. Study results are most commonly aggregated to reflect the average estimate of change in a score for each study arm. If the average change experienced by a study arm is lower than the MCID, it does not mean that, for some patients, a clinically important change has not occurred. Equally, if one arm reports an average change above the MCID and the other arm reports an average change below the MCID, it may not be possible to infer that the difference between the arms is clinically meaningful.

A more meaningful method of examining response is to report the proportion of participants who experienced a change in a score that was greater than the MCID. A responder analysis can also be represented by a cumulative distribution function such that the proportion of responders can be viewed across multiple thresholds for an MCID.

Noninferiority margin

A claim of noninferiority means that, in terms of safety and effectiveness, the proposed therapeutic technology is no worse than the main comparator. However, a lack of a statistically significant difference between the proposed intervention and the comparator does not adequately establish noninferiority. It is common practice to require that the confidence limits of the difference in treatment effect do not include an *a priori* stated clinically meaningful difference favouring the comparator.

If the proposed intervention is claimed to be noninferior to the main comparator, state whether an acceptable noninferiority margin has been identified in the PICO confirmation, and for what outcomes. If a noninferiority margin is not available, describe any noninferiority margins identified in the literature, and state how they were derived.

The application of a noninferiority margin that was not prespecified in a study is difficult to justify. If a noninferiority margin is required to establish whether the proposed intervention is noninferior to a comparator, and studies have not prespecified a noninferiority margin, the selection of a conservative margin (i.e. narrow margin) is more appropriate.

A noninferiority margin is not necessary for all outcomes, and is typically only applied to the primary outcomes of noninferiority studies. Studies may be underpowered to support the use of

noninferiority margins for less common outcomes or outcomes that were not used to power the study.

Appendix 6 Sources of heterogeneity

This appendix describes possible sources of heterogeneity between studies, or when comparing one jurisdiction with another. It is a useful reference for describing potential confounders when combining studies in a meta-analysis, performing indirect comparisons of randomised trials or network meta-analyses, or comparing variables from the clinical study setting with the target population.

Make comparisons across studies or jurisdictions based on the distributions or proportions of each characteristic rather than simply identifying whether there is a representation of each characteristic in each study or jurisdiction. For example, 2 trials may include patients aged 20–60 years, thus, the population may appear homogeneous. However, if one trial has a much lower mean age, or the proportion of patients younger than 40 is far higher in one trial than another, this may be a source of heterogeneity and violate the assumption of transitivity.

Table 34 provides a list of important factors to consider when exploring heterogeneity in the evidence or the applicability of the evidence to the target population.

Table 34 Factors that might cause heterogeneity across studies or jurisdictions

Category	Factor
Study quality	Adequate concealment of randomisation
	Blinding
	Duration of follow-up
	Loss to follow-up, methods for censoring or imputation of missing data
	Crossover or treatment switching
Participant characteristics	Age, sex, performance status, comorbidities, physiological reserve
	Severity of disease, stage or duration of disease, previous therapy, genetic variation
	Intensity of surveillance, diagnostic workup
	Background therapy, advances in standard of care
	Values, expectations and adherence
Circumstances of use	Health systems, setting in hospital or ambulatory care
	Geography, urban or rural
	Date of studies (change in standard of care)
Management decisions	Regional/country variations in practice
	Different treatments available, accessible, reimbursed
Treatment characteristics	Dose, duration, timing
	Stopping or continuation criteria
Test characteristics	Assay platform, enzymes, reagents, protocols, primers
	Test thresholds
	Resolution of imaging
	Sampling method and handling

Category	Factor
	Interpretation of results, inter-rater variation
Outcome measures	Definition of outcome(s)
	Rating instrument
	Frequency of measurement
	Start point of measurement against duration or progression of disease or treatment, especially in time-to-event analyses
	Statistical approach and covariates

Appendix 7 Test accuracy measures

		True diagnosis Reference standard ('gold standard')		
		Biomarker or disease present	Biomarker or disease absent	
Index test	Positive	**a** **True positive (TP)**	**b** **False positive (FP)** **(type I error)**	**Positive predictive value** TP/test positive a/(a + b)
	Negative	**c** **False negative (FN)** **(type II error)**	**d** **True negative (TN)**	**Negative predictive value** TN/test negative d/(c + d)
		Sensitivity TP/disease positive a/(a + c) **Positive likelihood ratio (LR+)** TP rate/FP rate [a/(a + c)]/[b/(b + d)] sensitivity/(1 – specificity)	**Specificity** TN/disease negative d/(b + d) **Negative likelihood ratio (LR-)** FN rate/TN rate [c/(a + c)]/[d/(b + d)] (1 – sensitivity)/specificity	**Diagnostic odds ratio** (TP/FP)/(FN/TN) (a/b)/(c/d) LR+/LR-

Figure 35 The 2-by-2 table when a reference standard is available

		Test comparator or clinical utility standard		
		Positive	**Negative**	
Index test	Positive	**a** **Agreement, no change in management**	**b** **Apparent cases detected only by the index test**	
	Negative	**c** **Apparent cases detected only by the comparator test**	**d** **Agreement, no change in management**	
		Positive percent agreement 100% x a/(a + c)	**Negative percent agreement** 100% x d/(b + d)	**Overall percent agreement** 100% x (a + d)/ (a + b + c + d)

Figure 36 The 2-by-2 table for concordance

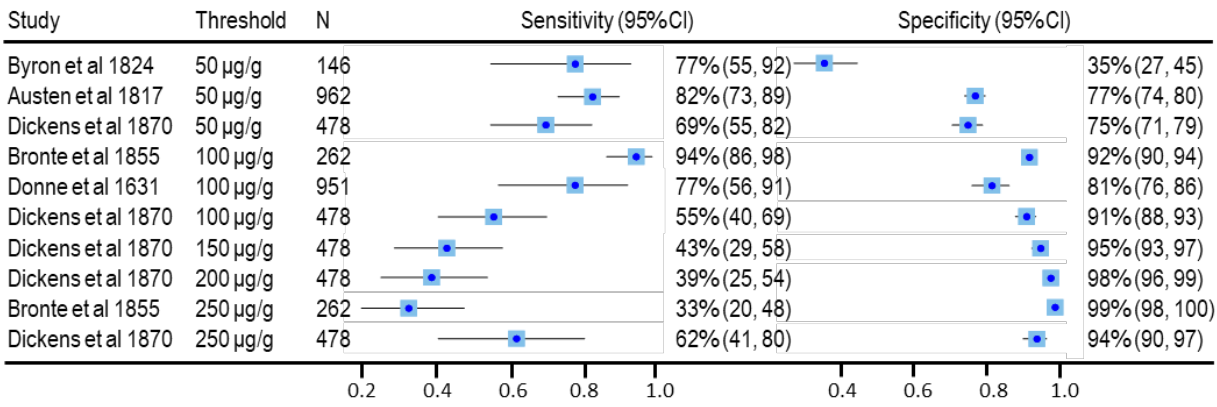
Table 35 Formulas for calculating test accuracy measures

Test accuracy measures	Calculations
Sensitivity	TP / disease positive = a / (a + c)
Specificity	TN / disease negative = d / (b + d)
Positive likelihood ratio (LR+)	TP rate / FP rate or sensitivity / (1 – specificity) = [a / (a + c)] / [b / (b + d)]
Negative likelihood ratio (LR–)	FN rate / TN rate or (1 – sensitivity) / specificity = [c / (a + c)] / [d / (b + d)]
Diagnostic odds ratio (DOR)	(TP / FP) / (FN / TN) = LR+ / LR– = (a / b) / (c / d)
Positive predictive value (PPV)a Study-specific For disease or biomarker prevalence rate in PICO population	TP / test positive = a / (a + b) $\frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$
Negative predictive value (NPV)* Study-specific For disease or biomarker prevalence rate in PICO population	TN / test negative = d / (c + d) $\frac{\text{specificity} \times (1 - \text{prevalence})}{(1 - \text{sensitivity}) \times \text{prevalence} + \text{specificity} \times (1 - \text{prevalence})}$
Number needed to diagnose (NND)	1 / Youden's index = 1 / (sensitivity + specificity –1) 1 / ([a / (a + c)] + [d / (b + d)] – 1)
Number needed to misdiagnose (NNM)	1 / (1 – accuracy) = 1 / 1 – [(a + d) / (a + b + c + d)] 1 / 1 – specificity – [prevalence × (sensitivity – specificity)]
Concordance measures	**Calculations**
Positive percent agreement	100% × [a / (a + c)]
Negative percent agreement	100% × [d / (b + d)]
Overall percent agreement	100% × [(a + d) / (a + b + c + d)]

a PPV and NPV can be calculated from the 2-by-2 table for individual studies. However, this value is only valid for the prevalence of the disease or the biomarker in that study. For the PPV and NPV values to be applicable to the Australian population defined in the PICO, these values should be calculated using the pooled sensitivity and specificity of the test and the applicable prevalence rate for the PICO population.

Meta-analytical methods

A key difference between pooling interventional/therapeutic study data and pooling test accuracy data is that test metrics (sensitivity and specificity) tend to be correlated. The interpretation of calculated test metrics for each study can be assisted by presenting confidence interval plots (forest plots) of the sensitivity and specificity side by side. Ordering the results by ascending sensitivity or specificity can help with visualising the relationship between the two (Figure 37).



CI = confidence interval; N = number of patients

The forest plot shows the sensitivity and specificity for a test compared with the reference standard for each threshold reported for each study. Overall, it looks like there is a trend that, as the threshold increases, the sensitivity decreases and the specificity increases.

Note: No overall pooled sensitivity or specificity values were calculated, as this was not appropriate. The forest plot includes duplicated data values: the population was the same for each threshold reported in the same study.

Figure 37 Forest plots showing the relationship between sensitivity and specificity of a test compared with the reference standard for different thresholds

Typically, there is an inverse correlation between sensitivity and specificity.[114] This may be due to different thresholds used to determine a positive sample, as in Figure 37. If the threshold to determine a positive test is decreased, this will permit more test positives, but will also increase the number of false positives. In this circumstance, sensitivity would increase and specificity would decrease. This inverse relationship between sensitivity and specificity, when related to differences in thresholds, is called a threshold effect.

For this reason, pooling of sensitivity and specificity using a bivariate meta-analysis method is preferred.

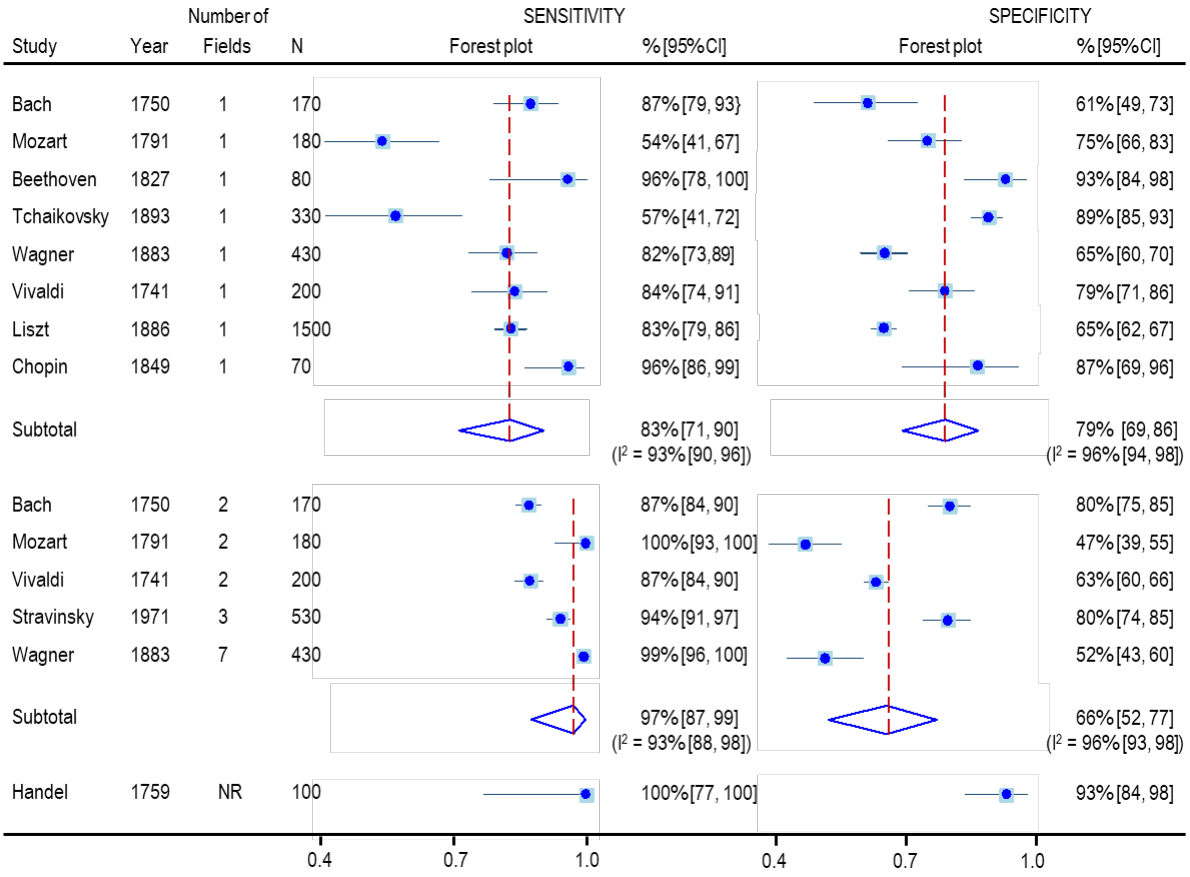
A bivariate model accounts for the correlation between sensitivity and specificity and is preferred when summary point estimates are sought. However, a minimum of 4 studies is required for this type of meta-analysis.

Bivariate meta-analysis

Bivariate models assume that the logit of sensitivity and specificity have a bivariate normal distribution between studies.[115] There are other options for performing a random effects bivariate meta-analysis that assume beta-binomial distributions,[116] and extensions of the bivariate model that may permit the inclusion of thresholds as a covariate.[117] These other methods are more complex and should be clearly described when applied.

A minimum of 4 studies is required for bivariate meta-analysis to obtain summary point estimates of sensitivity and specificity. In some cases, the bivariate models do not converge, especially if there are few studies or several zero cells in the 2-by-2 table.[118] If this occurs, separate univariate binomial meta-analyses for sensitivity and specificity can be used with justification and a discussion of the uncertainties in the approach. In some cases, it may be appropriate to use univariate models to pool the diagnostic odds ratio, which is an estimate that incorporates both sensitivity and specificity. The decision to estimate a summary point should be based on the characteristics of the included study data. If all studies report the same or very similar test thresholds and population characteristics, it is probably reasonable to use the bivariate model approach for overall estimates. Where the

population subgroups and/or test thresholds differ meaningfully, or the sensitivity and specificity vary over a large range, an overall pooled estimate is difficult to interpret and is unlikely to be appropriate under these circumstances.[119] Applying the bivariate model to each of the distinct subgroups is likely to be more appropriate, as shown in Figure 38.



CI = confidence interval; N = number of patients; I^2 = statistic describing the percentage of variation across studies that is due to heterogeneity rather than chance

Subgroup analysis showed that the sensitivity increased and the specificity decreased when more than one field was included in the imaging test, although the 95% CIs were overlapping suggesting the difference may not be statistically significant.

Figure 38 Forest plot of sensitivity and specificity for an imaging test compared with a reference standard according to the number of fields included

For some tests, there is no universally agreed threshold (or cutpoint) for determining a positive result, and some studies may use several thresholds. Threshold effects should be further analysed by including covariates in bivariate models. If a mixture of thresholds is used across and/or within studies, and there is no clear reason to limit the analysis to a single threshold, it may be appropriate to present a hierarchical summary receiver operating characteristic (HSROC) curve. HSROC curves characterise the relationship between sensitivity and specificity across the included thresholds, and this graphical representation of the included studies provides an easy way to examine both the threshold effect and between-study heterogeneity.

Common statistical software packages that can be used to perform meta-analysis of test accuracy studies using bivariate or hierarchical models include:

- STATA using the midas or metandi commands
- SAS using the metaDAS macro

- R using the mada package.

Univariate meta-analysis

As a general rule, a bivariate model approach or a hierarchical model approach is preferred; however, if the model does not converge, separate univariate binomial meta-analyses can be used. Justification for this approach and a discussion of the uncertainties in the approach should be provided. Non-convergence of models may occur when there are few studies or sparse data, particularly if there are several zero cells in the 2-by-2 table.[118]

Random effects univariate meta-analysis will produce 'average' estimates of sensitivity and specificity. This approach does not account for heterogeneity in sensitivity and specificity related to the threshold effect, and statistical measures of heterogeneity may be difficult to interpret. In the presence of a threshold effect, the individual pooled estimates of sensitivity and specificity may be incompatible.[115]

In the absence of visual heterogeneity across studies, separate random-effects univariate meta-analyses of sensitivity and specificity will approximate the use of more complex model-fitting methods. For univariate models, the most appropriate method for pooling sensitivity and specificity is to perform a binomial meta-analysis.[120]

The diagnostic odds ratio (DOR), defined as the ratio of the odds of positivity in those with the biomarker or condition relative to the odds of positivity of those without the biomarker or condition,[121] is a single parameter of test accuracy. Hence, DOR can be pooled using univariate models. However, DOR summary measures do not distinguish between the ability to detect true positive cases (sensitivity) and the ability to detect true negative cases (specificity) and are, therefore, more difficult to interpret in a clinically relevant way.[122] In other words, the same DOR may be achieved with different sensitivity and specificity values. The use of DORs can overcome the issues with the negative correlation between sensitivity and specificity,[114] and may help to ascertain the 'best performing' test. However, if a clinical situation requires greater test sensitivity and the trade-off in specificity is permissible (but not the other way around), the highest DOR might not be the best performing test. DORs can be applied in meta-regression to explore heterogeneity.

Multiple thresholds from single studies

Test accuracy data for any one patient should only be included in a meta-analysis once. If 2 studies have the same or overlapping patient cohorts, only one can be included in the meta-analysis. Similarly, if individual studies report test accuracy data for the same patients or samples for different thresholds, only one threshold can be included in a single meta-analysis if summary estimates are to be reported. If a meta-analysis of different thresholds were undertaken, it would be appropriate to include the same study in the separate meta-analyses for the different thresholds, as shown in Figure 38.

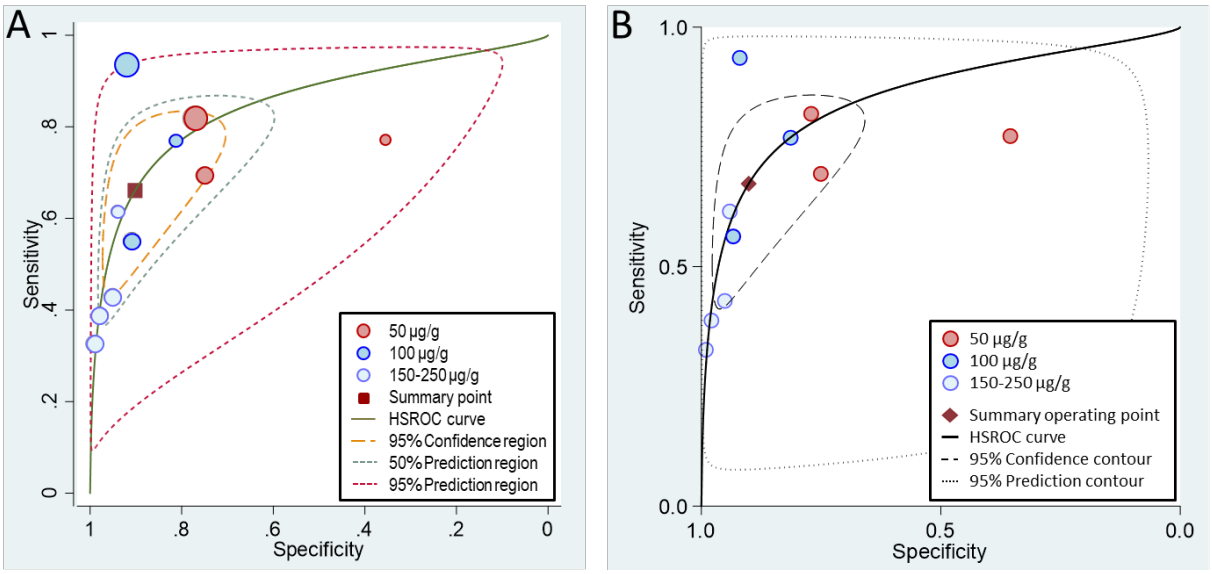
Hierarchical summary receiver operating characteristic (HSROC) curve

For some tests, there is no universally agreed threshold for determining a positive result and some studies may use several different thresholds. If a mixture of thresholds is used across and/or within studies, and there is no clear reason to limit the analysis to a single threshold, it may be appropriate to present a HSROC curve. HSROC curves can be generated either by using HSROC models that directly estimate HSROC parameters, or by transforming the estimated parameters of a bivariate model so that an HSROC curve can be fitted. The 2 models are mathematically equivalent and provide equivalent estimates of expected sensitivity and specificity.[122]

HSROC curves characterise the relationship between sensitivity and specificity across the included thresholds and account for within- and between-study heterogeneity. The HSROC curve plots the true positive rate (or sensitivity) against the false positive rate (or 1 – specificity), and this graphical representation of the included studies provides an easy way to examine both the threshold effect and between-study heterogeneity. The 95% confidence region is a measure of the precision of the test accuracy estimate and the 95% prediction region is a measure of between-study variability or heterogeneity, defining the area in the HSROC space where a future study would lie.[123] However, as test accuracy studies tend to be highly variable, the 95% prediction regions often cover large areas of the HSROC space. A 50% prediction region is equivalent to the interquartile range.

Figure 39 shows HSROC curves generated in STATA using the metandi command (HSROC model) and the midas command (bivariate model), using the same studies shown in the forest plot in Figure 37, with respecified thresholds. Although the point estimates derived for each HSROC curve are the same, there are some differences between the 2 HSROC curves with respect to the 95% confidence and predictive regions. Both curves enable visualisation of the threshold effect, where the studies with thresholds of 50 µg/g were more sensitive but less specific than those with thresholds at or above 100 µg/g, but this was more difficult to discern in the forest plot. The large 95% prediction region suggests that there is some heterogeneity between the studies in Figure 39. In contrast, in Figure 40, the 95% prediction region fits much more tightly to the HSROC curve, suggesting there is less heterogeneity between the included studies.

When HSROC curves include multiple thresholds from the same study, reporting of summary measures such as the summary point sensitivity and specificity values, as well as the area under the ROC curve (AUROC), are not appropriate.

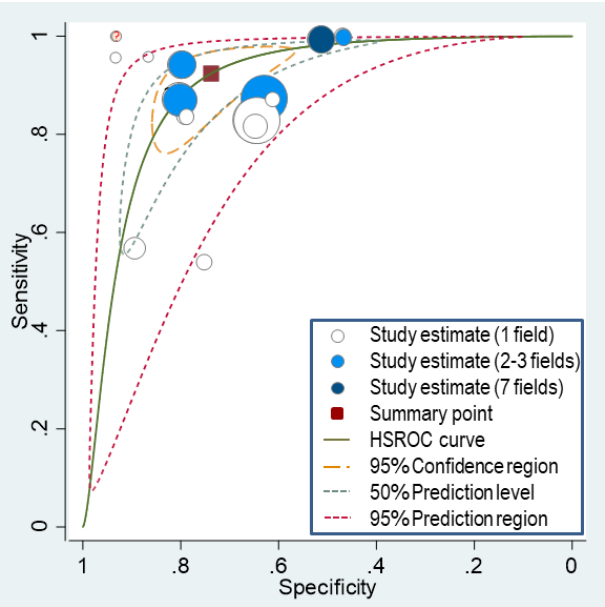


HSROC = hierarchical summary receiver operating characteristic

The HSROC curves were generated in STATA using the metandi command (A) and the midas command (B), using the same studies shown in the forest plot in Figure 37. The curves show a trend where the studies with thresholds at or below

50 µg/g were more sensitive but less specific than those with thresholds at or above 100 µg/g. The 95% prediction region, defining where a future study would lie, is much larger when using the midas command.

Figure 39 HSROC curves summarising the accuracy of a test compared with the reference standard for different test thresholds

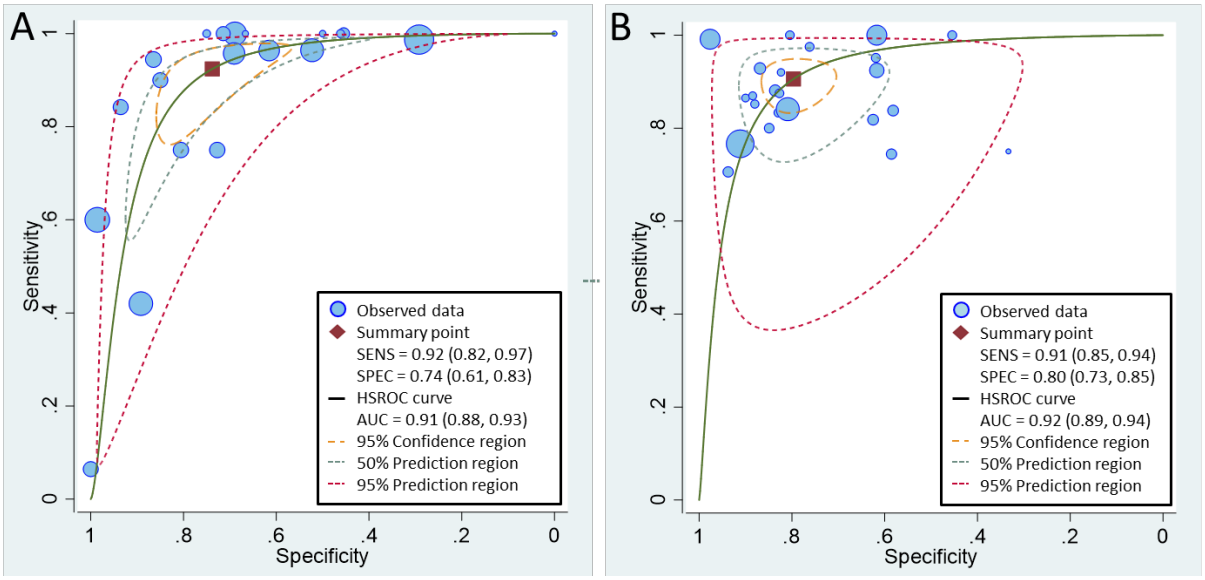


HSROC = hierarchical summary receiver operating characteristic

The HSROC curve was generated in STATA using the metandi command and shows a threshold effect; the sensitivity increases and the specificity decreases as the number of fields included in the test increases.

Figure 40 HSROC curve summarising the accuracy of a test compared with the reference standard for different test thresholds where there is no reporting of multiple thresholds in the same study

In some cases, an HSROC curve may show a threshold effect even if there are no apparent differences in the test or population characteristics between studies (Figure 41). This may be due to differences in test characteristics between studies (or laboratories) that may or may not have been reported. For example, differences in the test protocol (e.g. timing of processing steps, concentration of the solutes used) or laboratory equipment used, or variations in the interpretation of the results (e.g. scoring algorithms, inter-rater variability) may indicate systematic detection differences between studies, resulting in a 'threshold effect'. However, care should be taken in concluding a threshold effect in the absence of evidence that different thresholds apply across studies.



AUC = area under the ROC curve; HSROC = hierarchical summary receiver operating characteristic; SENS = sensitivity; SPEC = specificity

The HSROC curves were generated in STATA using the metandi command. The HSROC curve for the test in panel A shows a possible threshold effect, whereas the HSROC curve for the test in panel B does not.

Figure 41 HSROC curves summarising the accuracy of 2 different tests compared with the relevant reference standard where there were no obvious key differences between studies

The AUROC is the average of the true-positive rate over the entire range of false-positive rate values. The AUROC value serves as a global measure of test accuracy that can be interpreted as follows: given a randomly selected patient with the condition, and a randomly selected patient without the condition, it is the probability that the patient with the condition would be ranked more highly than the patient without the condition.[124] The following guidelines have been suggested for interpretation of AUROC values: for values above 0.9, test accuracy is high; for values between 0.7 and 0.9, test accuracy is moderate; and for values below 0.7, test accuracy low.[125] An AUROC value of ≤0.50 indicates that the test cannot discriminate between true positives and true negatives, with the curve lying on or below the major diagonal.

Assessing heterogeneity between studies included in a meta-analysis

A test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect. As test accuracy is usually calculated with 2 correlated estimates (sensitivity and specificity) from the same study, analysing the variability or heterogeneity in these estimates between studies is challenging.

Heterogeneity is usually measured using 2 different measures: the Cochran's Q statistic, or the inconsistency measure I-squared (I^2). The Cochran's Q statistic is the sum of the squared deviations of each study's estimate from the overall pooled estimate, according to the study weighting in the meta-analysis.[126] P values are obtained by comparing the statistic with a chi-squared (χ^2) distribution with k–1 degrees of freedom, where k is the number of studies. However, the power of the test is low when only a small number of studies is included in the meta-analysis, and consequently the test is poor at detecting true heterogeneity as significant.

The I^2 statistic, which does not depend on the number of studies, measures the degree of inconsistency, or the percentage of total variation across studies, that is due to heterogeneity rather

than chance. I^2 can be readily calculated from the Cochran's Q statistic as: 100% × (Q – df)/Q, where df is the degrees of freedom.[127] Negative values of I^2 are considered to be equal to zero so that I^2 lies between 0% (no heterogeneity) and 100%. It should be noted that calculating separate I^2 statistics for sensitivity and specificity fails to account for variation explained by the correlation between sensitivity and specificity, as well as for threshold effects, and will overestimate the degree of heterogeneity observed.

Publication bias

Publication bias is usually evaluated by visual inspection of a funnel plot. If there is no publication bias, studies are evenly distributed within the inverted funnel. In the presence of publication bias, the distribution of studies in the funnel plot will be asymmetric. The standard methods for generating a funnel plot to determine publication bias were developed for therapeutic intervention studies by Egger et al.[128] and Begg and Mazumdar[129] and can be inaccurate for test accuracy studies.[130]

The method by Deeks, Macaskill and Irwig[131] has been developed for use with test accuracy studies. It plots the diagnostic log odds ratio against the effective sample size ($1/ESS^{1/2}$), where the effective sample size is a simple function of the number of diseased and nondiseased individuals. This method is recommended for test accuracy studies in the *Cochrane handbook for systematic reviews of diagnostic test accuracy*.[132]

Additional test accuracy measures

When describing the accuracy of a test, there may be additional measures that will provide relevant information for decision-makers. When presenting any test accuracy measure, provide an explanation of the measure, and an interpretation of the results. If additional test accuracy measures are not likely to influence decision-making, do not present them.

The number needed to diagnose or misdiagnose

The number needed to diagnose (NND) is the number of patients who need to be examined to correctly identify one person with the disease in the PICO population. The number needed to misdiagnose (NNM) is the number of patients who need to be tested for one to be misdiagnosed by the test.

The number needed to diagnose or misdiagnose provides some information about the usefulness of the test in the clinical setting. The fewer patients needed to test to identify someone with the disease and the more patients who are tested before one is misdiagnosed, the more useful the test is to clinicians.

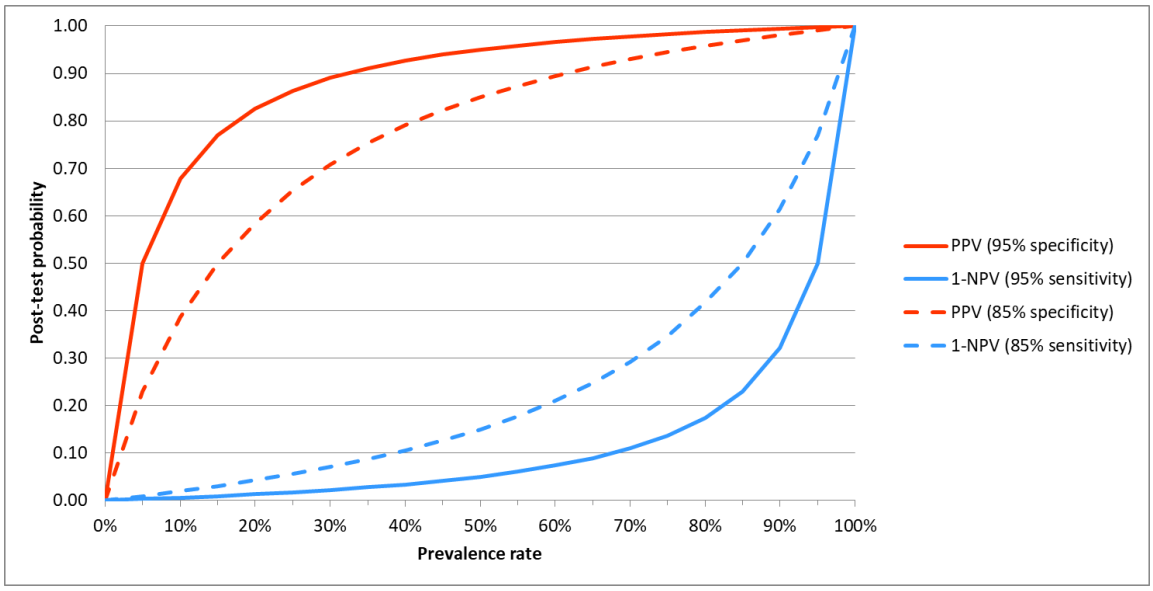
The post-test probability of having the disease with a positive or negative test result

The post-test probability of a test correctly identifying patients with and without disease provides a measure of the usefulness of the test in a clinical setting. The post-test probability can be calculated using either the positive and negative likelihood ratios (LR+ and LR– values) or the positive and negative predictive values (PPV and NPV).

The pre-test probability of having the biomarker or condition is equivalent to its prevalence rate in the target population.

Meta-analysis of PPV and NPV from individual studies is not recommended because these values are affected by the prevalence of the biomarker or condition in the target population and are not

directly comparable when the prevalence rate varies between studies. However, PPV and NPV can be calculated from the pooled sensitivity and specificity estimates by applying the estimated prevalence of the disease or biomarker in the target population (the formula for this calculation is provided in Table 35). Both PPV and NPV are valuable metrics for the interpretation of test accuracy in the clinical setting. PPV is the percentage of patients with a positive test who actually have the biomarker or condition, and is equivalent to the post-test probability of a positive test result being true. NPV is the percentage of patients with a negative test who do not have the biomarker or condition and its inverse (1–NPV) is equivalent to the post-test probability of a negative test being false. As the prevalence rate increases, for any given pair of sensitivity and specificity values, the PPV will increase and the 1–NPV will decrease, as shown in Figure 42.



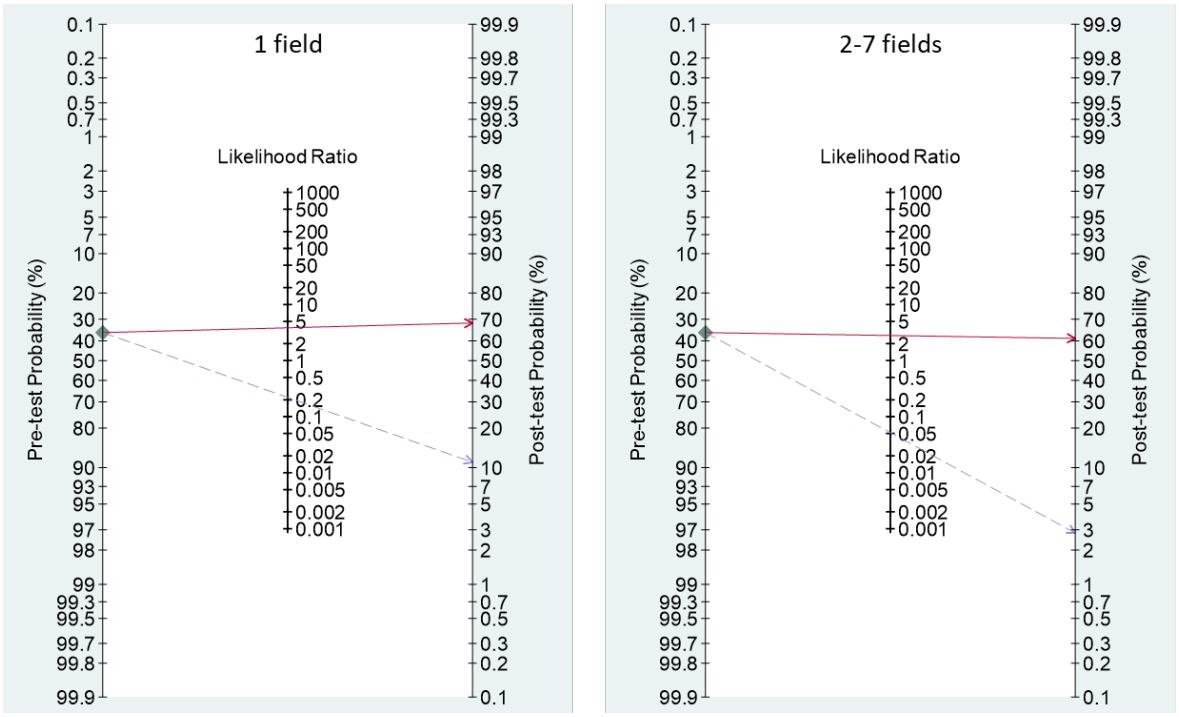
NPV = negative predictive value; PPV = positive predictive value

The solid lines show the post-test probability of being truly positive for a sensitivity and specificity of 95%; the dotted lines show how the post-test probability changes when the sensitivity and specificity are reduced to 85%.

Figure 42 Graph showing the relationship between the prevalence rate and the post-test probability of a positive test being truly positive (PPV) and a negative test being falsely negative (1–NPV)

The summary likelihood ratios (LRs) of a test plus the prevalence rate (pre-test probability) of the biomarker or condition also enable the post-test probability of having the biomarker or condition to be estimated by plotting these values on a Fagan's nomogram.

The red line plots the pre-test prevalence rate and the LR+ to show the post-test probability of having the condition if the test result is positive (Figure 43). The blue line plots the pre-test prevalence rate and the LR– to show the post-test probability of having the condition if the test result is negative.



CI = confidence interval; LR+ = positive likelihood ratio; LR– = negative likelihood ratio

The Fagan's nomogram was generated in STATA using the midas command.

The prevalence rate (or pre-test probability of having the condition) is 36%. LR+ (red solid line) and LR– (blue dashed line) values were plotted to obtain the post-test probability of having the condition. The post-test probability of having the condition is almost double the pre-test probability with a positive test result, although the number of fields photographed has little effect on the diagnosis. However, the post-test probability of having the condition decreases from 11% to 3% if >1 field is photographed. This increases the usefulness of the test as a triage test, where only patients with a positive test result are retested with other tests for confirmation of the presence of the condition.

Figure 43 Fagan's nomogram showing the post-test probability of having the condition with a positive or negative imaging test result, depending on the number of fields photographed

Appendix 8 Codependent technologies

Introduction

Health technologies are codependent when the patient health outcomes related to the use of one technology are contingent on, or modified by, another technology. A clear example is the use of an investigative technology to determine eligibility or appropriate treatment with a therapeutic technology. Although most treatment is guided by one or more tests, such that the benefit of the treatment relies upon the test information, very few technology pairs require a codependent approach.

Material codependency

A codependent assessment is necessary when the Minister for Health requires advice from 2 different advisory committees. This is called a 'material codependency'.

Most examples of material codependency involve the sequential use of an investigative technology (test) followed by treatment with a medicine based on the results of the test. While there may be other examples of codependency that include tests and other types of therapeutic interventions (e.g. surgery, radiotherapy, implantation of a device), most of these examples will be considered by a single advisory committee (MSAC) and the relationship is not a material codependency.

As most examples of material codependency are test–medicine pairs, guidance for assessing codependent technologies has been included in the Pharmaceutical Benefits Advisory Committee (PBAC) guidelines (Product Type 4).

Integrated versus streamlined approach

An integrated codependent assessment is a combined assessment report required if the 2 technologies are lodged at the same time and are considered jointly by 2 advisory committees.

A streamlined codependent assessment represents individual assessment reports for each of the technologies, lodged at the same time and considered in parallel by 2 advisory committees.

Further information regarding when an integrated or streamlined assessment is required is provided in Product Type 4 of the PBAC guidelines.

Additional information requests

The codependent product type (P4) of the PBAC guidelines describes 62 additional information requests required for the assessment of codependent technologies. In many cases, the information requests describe similar considerations to those outlined in the investigative technologies guidance presented in Section 2B of these guidelines, and the 2 guidelines should be read in conjunction.

Presentation of a codependent assessment report

An assessment report (or submission) for consideration of a codependent technology by MSAC should provide relevant evidence for Section 1 through Section 5, modified by guidance for product

type 4 in the PBAC guidelines. Responses to additional information requests relevant to each section should be included in the assessment report.

In general, an assessment report for a codependent technology would resemble an assessment report for an investigative technology.

Where the key study reporting on health outcomes enrols both test-positive and test-negative patients, there may be adequate evidence to adopt a direct from test to health outcomes approach. It is more common for health outcomes evidence to report only on biomarker-positive patients, and therefore a linked evidence approach is usually required.

Whether the assessment involves a direct from test to health outcomes approach or a linked evidence approach, the therapeutic component of the assessment report would resemble Section 2 of a submission to the PBAC (as described in the PBAC guidelines).

Appendix 9 Expert opinion

Uses of expert opinion

Consider providing expert opinion to supplement or support the observed data from randomised trials or nonrandomised studies.

Determining an appropriate body of experts will depend on the nature of the information gap that requires filling. Experts may be panels of medical practitioners, a medical specialty group or consumers. Consumers may provide advice on factors such as the patient relevance of outcomes (particularly if elicited at the time of trial design) or how health technologies might be used. Expert opinion can be useful in several aspects of preparing a PICO confirmation or an assessment report for consideration by MSAC. For example, expert opinion can be used to help:

- define the clinical need for the proposed health technology and inform the main indication (discussed in Technical Guidance 1)
- determine how the health technology is most likely to alter the clinical management algorithm (TG 2.6) and to support the choice of the main comparator (TG 2.3), noting that a comparator should not be determined by expert opinion alone
- interpret the clinical importance and patient relevance of the outcome measures reported in studies (Appendix 5)
- modify the patterns of health care resource use measured in studies conducted in different settings, such as in other countries (Technical Guidance 22 and Section 4)
- predict which health care resources would be used and how often each would be used to manage outcomes reported in the included studies (Technical Guidance 22 and Section 4)
- estimate the proportion of patients with the medical condition that would be eligible according to the requested listing, and predict uptake rates (Technical Guidance 19 and Section 4)
- predict the impact on the utilisation of other health technologies (TG 27.3).

In several of the examples listed above, trial data, registry data or analyses of data from other countries, where available, would be used in preference to expert opinion, and it would be expected that the expert opinion supports the applicability of the observed data. An example would be using expert opinion to support the representativeness of a utilisation evaluation conducted in another country. In this case, expert opinion reduces uncertainty.

Presenting expert opinion

Justify the use of expert opinion in the introduction of the appropriate section. Include a clear rationale for, and the aims of, eliciting the expert opinion. Where expert opinion is used to fill a gap in information, clearly describe the nature of this gap and indicate the other steps that have been taken to address the gap, such as a literature search.

Describing the collection and collation of expert opinion

Using a well-designed methodology to elicit expert opinion helps to reduce uncertainty. The methods used may vary from large, published questionnaires and surveys with statistical analysis to

a summary of interviews with a panel of clinical experts. Present expert opinion as qualitative or quantitative (but not statistically analysed) information.

Include copies of administered surveys or hypothetical scenarios that were presented to experts.

When summarising expert opinions and their variability, interpret the findings, and discuss the limitations and biases of the method chosen. Qualitative studies and interviews should follow best practice for reporting and analysis. Indicate how the opinions have been used in the PICO confirmation or the assessment report.

Where multiple sources of expert opinion are available to address a single assumption or estimate, compare the results, and assess their concordance or lack of it. Present a summary table that compares multiple sources or multiple variables. Table 36 provides guidance on the details that should be included. Where multiple estimates (or data) are generated to fill a gap in the information – either from multiple sources of expert opinion, or a combination of expert opinion and observed data – compare the estimates (or data) and justify the choice of data used in the submission.

If expert opinion is used for an assessment in place of observed data, as may occur when observed data are generated from other health care systems or are historical, present both and clearly justify the use of expert opinion. State whether expert opinion (compared with alternative sources of data) is likely to lead to a more favourable clinical, economic or financial assessment of the proposed health technology.

Use of information that is uncertain in the clinical, economic or financial analyses of the proposed health technology will raise concerns for MSAC. Where expert opinion is sought for a disease or condition for which the number of medical practitioners is likely to be large, do not rely on surveys of small numbers of practitioners because this leads to highly uncertain results. In all cases where expert opinion is used to derive estimates for the assessment, use the final estimate to minimise the risk to MSAC of relying on an overestimation of the effectiveness or cost-effectiveness, or underestimation of the financial implications to the Australian Government or other funding body. To reduce uncertainty associated with expert opinion, provide sensitivity analyses around the derived estimates, or clearly state where the results in the assessment are not sensitive to different estimates.

Table 36 Methods to collect and collate expert opinion

Information to be provided	Notes
Criteria for selecting experts	Prefer a random or comprehensive set of health practitioners likely to prescribe the proposed health technology, or the appropriate medical specialty group. In general, an advisory board or group of practitioners associated with the manufacturer or sponsor may not be representative of experts in Australian clinical practice. The generalisability of expert opinion derived from such boards is difficult to assess.
Number of experts approached[a]	Where the likely number of health practitioners is large, it is less acceptable to provide expert opinion derived from a small number of practitioners.
Number of experts who participated[a]	Assess whether the extent and characteristics of the nonresponders are likely to diminish the representativeness of the opinions provided, compared with the intended sample approached.
Declaration of potential conflicts of interest from each expert or medical specialty group whose opinion was sought	Provide a signed statement from each expert and specialty group specifying any potential conflict of interest and stating the nature of any contractual arrangement, including how much payment was offered and accepted. Where the collection of expert opinion has been contracted out, the contractor should provide this statement, reporting on both the arrangements made between the applicant or evaluator and the contractor, and the arrangements made between the contractor and those whose opinions were sought.
Background information provided and its consistency with the totality of the evidence provided in the assessment report	Include a copy of any background information provided in the technical document or attachment. If background information has been provided, ask the experts to define the comparative clinical place of the proposed health technology and the main comparator based on this background information. Including the experts' definitions in the technical document or attachment allows an assessment of the consistency of the background information with the evidence provided in the assessment.
Method used to collect opinions	For example, were the experts approached individually or was a meeting convened? Was any incentive used to maximise responses?
Medium used to collect opinions	For example, was information gathered by direct interview, telephone interview or self-administered questionnaire?
Questions asked[b]	Explain the design of the tool (quantitative or qualitative). Describe its development. Indicate whether it was pilot tested and, if so, provide the results of that testing and explain how the results were used to improve the questions. On a question-by-question basis, assess the extent to which each question is neutral or biased, and the extent to which each question is open or closed. To allow an independent assessment, include the questionnaire or an outline of the interview questions in the technical document (or attach a copy).
Whether iteration was used in the collation of opinions and, if so, how it was used	The Delphi technique, for example, uses an iterative approach.
Number of responses received for each question[a]	Assess whether the extent of any nonresponse is likely to diminish the representativeness of the opinions provided to particular questions, compared with the intended sample approached.

Information to be provided	Notes
Whether all experts agreed with each response	If not, specify (i) the approach used to finalise the estimates (e.g. the majority opinion or a Delphi technique could be applied; for quantitative results, point estimates [such as the mean, median or mode] could be presented), and (ii) the approach used to present the variability in the opinions (e.g. present the range of opinions expressed, including common and outlying views; for quantitative results, measures of variance [such as confidence intervals, range, centiles] could be presented).

a Tabulate these information items.

b The way the questions are asked is an important source of potential bias in obtaining expert opinion. A particularly influential extension question extends the respondent beyond 'what' the opinion is (e.g. what would be done, what extent of benefit would be clinically important) to ask 'why' (e.g. explain why would you do this, explain why this is important). Conveying these reasons alongside expert opinion–based estimates might help improve their acceptability. Including these explanations in the technical document or attachment would allow the opinions to be assessed based on the underlying reasoning rather than only depending on the authority of the experts.

Appendix 10 Including nonhealth outcomes in a supplementary analysis

Presenting nonhealth outcomes

Occasionally, listing a proposed health technology might generate worthwhile impacts that are not captured as health outcomes, such as the value of information to the patient generated by an additional diagnostic test that does not change management of a medical condition.

Supplementary methods to estimate the monetary (or other) value of the nonhealth benefit may include a conjoint analysis or a discrete choice experiment that includes a monetary attribute, an attribute reflecting a range of options for each of the nonhealth outcomes of interest, and/or other attributes.

Where there are no other substantive changes in health outcomes between the proposed health technology and its main comparator, this estimate (e.g. willingness to pay) can be included in a supplementary cost-benefit analysis. Where this cost-benefit analysis results in a consumer surplus, nominate a suitable basis for sharing this consumer surplus between the sponsor and the taxpayer.

Production changes

In the context of health economics analyses, a production change is a change in total output value across society of productive work in the economy. Productivity is a function of output units (e.g. days of work) multiplied by their value (e.g. an appropriate daily wage as a proxy for the value of each day of work).

Health interventions may claim to result in a change in production across society associated with patients gaining or losing working time as a result of changes in their health and consequent capacity to work. Less commonly, a health intervention may claim that worker efficiency will be affected, such that the value of work output is changed on a per-unit basis (i.e. it can be represented by a higher or lower wage).

Changes in production as an outcome of therapy may be included in supplementary analyses in submissions to MSAC, but do not include them in the base-case analysis. This separation allows MSAC to consider the impact of including production changes on the direction and extent of change in the base case. Including production gains favours interventions that improve the health of people who are able, and choose, to return to contributing to societal production and, hence, there are equity implications of including productivity changes in the base case.

If presenting productivity claims associated with a proposed health technology, there are several difficulties in estimating the net present value of production changes. From a societal perspective, the productivity of an individual worker cannot be considered in isolation, but should be considered in the context of a workplace, a workforce and society. The following 3 underpinning assumptions should be incorporated into all productivity analyses:

- For short-term absence, production will be made up on return to work.
- Employers usually have excess capacity in the labour force to cover absenteeism.
- For long-term absence, production will be made up by a replacement worker who would otherwise be unemployed.

When presenting estimates of the marginal increase in society's production because of the return of healthy workers:

- provide details of the method used and its assumptions
- discount appropriately any productivity changes anticipated beyond one year
- address each of the assumptions listed above when estimating production changes from the potential working time gained or lost (reported in time units).

Address each of these 3 factors to provide robust evidence in support of estimates.

For example, a claim that returning to health from an episode of illness leads to a recovery of production lost depends on demonstrating the following 3 factors:

- The worker returns to work and the worker is productive.
- The production lost is not made up elsewhere by others in the company or the same worker following return to work.
- No temporary replacement has been employed.

Ensure that estimates of the proportion of people who choose to return to work account for those who would choose not to return (and instead use their time gain on other activities that will have been captured by a gain in utility weights), as well as the influence of incentives provided through sickness benefits, which may operate differently across jurisdictions.

The approach above may be adapted to other contexts, such as a health technology that prevents future episodes of illness, or one that might improve production capacity in individuals who, without the proposed health technology, would otherwise stay at work, although unwell, and therefore function at less than full production capacity.

When the economic approach is a cost-utility analysis, discuss how the method of estimating productivity changes avoids double-counting the estimates of health-related quality-of-life changes. The utility weights in this analysis already capture these health-related changes because they incorporate the utility impacts of productive capacity for the individual receiving the proposed health technology. These health-related changes are therefore already appropriately included in the denominator of the cost-utility ratio.

Strongly justify any production changes that are combined with surrogate outcome indicators in an economic evaluation, because this combination is generally associated with inappropriately high levels of uncertainty.

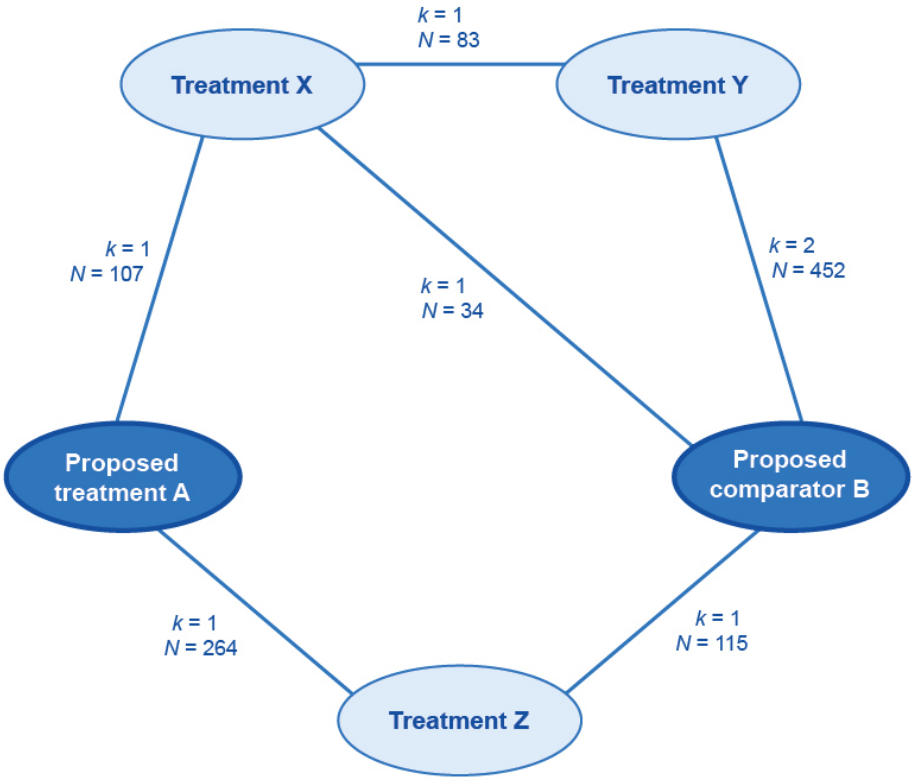
Appendix 11 Selection of studies for indirect comparison

Where the approach taken in the assessment report includes an indirect comparison, the studies included will typically be randomised trials of the intervention and randomised trials of the main comparator(s), which share a common reference. In some circumstances, included trials will have transitivity issues that would reduce the certainty of the results of the indirect comparison. If this is the case, justify the exclusion of trials that are unsuitable for use in the indirect comparison.

In general, a simple indirect comparison (e.g. a pairwise Bucher method) is adequate to estimate the treatment effect of a proposed health technology compared with a main comparator. In some circumstances, more complex methods (such as network meta-analysis) may be appropriate.

A general approach to identifying suitable trials for an indirect comparison is summarised here:

1. Perform appropriate searches to identify all studies of the intervention and of the comparator (Appendix 2).
2. Draw a network diagram to show all the possible links (Figure 44).



k = number of trials; *N* = number of patients enrolled

Figure 44 Example network diagram of the trials included to inform an indirect comparison of the proposed health technology with the main comparator

3. Where pairwise comparisons are possible, the assessment report may seek to exclude linkages requiring multiple steps, or include these as a supplementary analysis. In the absence of pairwise

comparisons, presenting the smallest number of steps is usually preferred. Provide a justification for the choice of the base case.

4. Examine heterogeneity within trial sets and across trial sets, and justify the exclusion of trials with differences in factors that may affect the transitivity of the trials in the indirect comparison.
    - Do not exclude studies on the basis of differences in characteristics that are unlikely to influence the treatment effect for either the proposed health technology, or the proposed comparator. The exclusion of studies due to concerns regarding heterogeneity of characteristics is supported by a demonstrable impact on the treatment effect, and requires a clear explanation.
    - If studies are removed at this step, it is useful to include them in a sensitivity analysis.
    - Possible sources of heterogeneity are listed in Appendix 6.
5. Examine the event rates in the common reference arms. The indirect comparison will, by design, adjust for differences in the event rates in the common reference arms of the studies. However, the success of the indirect comparison is based on an assumption of a constant relative (or absolute) treatment effect in the studies across different baseline risks. This is not certain.
    - If there is evidence from subgroup analyses that there is a constant treatment effect across different risk groups, do not exclude studies on this basis. Select the most appropriate outcome (relative or absolute) that best fits the assumption of constant treatment effects.
    - If it is likely that there is not a constant treatment effect, consider and justify the exclusion of studies on the basis of differences in the event rates in the common reference arms.
6. Present a list of the studies included in the main analysis, the studies included in supplementary or sensitivity analyses, and the studies excluded from all analyses.

Appendix 12 Translating comparative treatment effects of proposed surrogate measures to target clinical outcomes

Introduction

MSAC prefers submissions that do not rely on proposed surrogate measures (PSMs) to inform effectiveness in terms of patient-relevant or clinically relevant outcomes. Where possible, present evidence from direct randomised trials of the treatment effect of the proposed health technology on clinically relevant outcomes.

If no such evidence is available, establish the likely comparative treatment effect on clinically relevant outcomes by transforming the comparative treatment effect of a surrogate measure.

A surrogate measure is a biomarker that is intended to substitute for one or more target clinical outcomes (TCOs). Although a surrogate measure may or may not have clinical relevance, it is not the key purpose for treatment, which is to affect the severity of, or the transition to, future TCOs.

In terms of relevance to MSAC, the relationship between a PSM and a TCO is one that quantifies the change in the TCO as a consequence of a change in the PSM. Throughout this appendix, the transformation of the PSM to the TCO should be interpreted as the transformation of the comparative treatment effect on the PSM to the comparative treatment effect on the TCO.

This appendix takes the following approach:

- Define the PSM and the TCO.
- Establish the biological reasoning for the link between the PSM and the TCO, including how pivotal the PSM is to the causation pathway of the TCO, and present epidemiological evidence to support this.
- Present randomised trial evidence to support the nature of the PSM–TCO comparative treatment effect relationship.
- Translate the comparative treatment effect on the PSM from the studies included in Section 2, to an estimate of the comparative treatment effect for the TCO.

When interpreting the evidence to identify the relationship between the PSM and the TCO, and the relationship between the comparative treatment effect on the PSM and the comparative treatment effect on the TCO, present indications of causality. That is, the PSM (and the comparative treatment effect on the PSM) always precedes the TCO (and the comparative treatment effect on the TCO), and their associations are strong, measured with high precision, and maintained after adjustment for confounders (if there are sufficient numbers of trials with sufficient information to enable such adjustment).

Use the following types of evidence to analyse a PSM–TCO relationship (listed from strongest to weakest):

1. multitrial meta-regression
2. single trial or small number of randomised trials where individual patient data are available (including multicentre analysis in which participants were randomised by centre)
3. one randomised trial – no individual patient data or not randomised by centre
4. no randomised trial data.

Given the uncertainty associated with transforming PSMs to TCOs, ensure that the treatment effect observed on the PSM is robust and unbiased. Bias may result from, for example, issues of study quality, imbalances in baseline characteristics, loss to follow-up, discontinuations, inappropriate dosing, subgroup analysis or adjustments for crossover. Where an unknown proportion of the comparative treatment effect on the PSM may be the result of bias, the estimate of the comparative treatment effect on the TCO will be uncertain. In the absence of a robust estimate of the comparative treatment effect on the PSM, transformation to a comparative treatment effect on the TCO is not informative.

The approach taken in this appendix has been informed by the Surrogate to Final Outcomes Working Group report,[a] and this remains a useful resource when additional explanation is required.

Definition, selection and measurement

Proposed surrogate measure

If an intervention may have multiple benefits (e.g. avoiding multiple strains of a virus or multiple forms of cardiovascular events), a PSM that captures the overall intended clinical outcome is more persuasive. Ensure that the PSM is responsive, and able to be measured with reliability and validity.

Define and describe the PSM, with reference to the epidemiological and randomised trial evidence identified in this appendix, by including the following:

- the units of measurement
- the measurement tool(s) or criteria used
- the evidence of reliability from test to test
- the variability across observers or different measurement tools
- the measurement of the comparative treatment effect (e.g. odds ratio, standardised mean difference).

Ensure that the definition and method of measurement of the PSM are consistent across the evidence. Report and discuss any discrepancies.

Target clinical outcome

Ensure that the choice of TCO is patient relevant and captures the key purposes for intervening in a disease process. The goal of treatment may be to improve quality of life, or prevent or slow a medical condition in the long term. Ensure that the TCO is consistent with the health states defined in the natural history of the disease or condition. In some cases, more than one TCO may be

a www.pbs.gov.au/info/industry/useful-resources/pbac-feedback

required to capture the effects of the proposed health technology on the disease or condition process. Justify use of the nominated TCO if it does not capture an outcome of the disease or condition, or an adverse outcome of the treatment. There may be evidence that the proposed health technology has a positive treatment effect for one TCO (e.g. myocardial infarction) and a negative treatment effect for another TCO (e.g. haemorrhagic stroke).

With reference to the epidemiological and randomised trial evidence identified in this appendix:

- justify the choice of the TCO and justify the exclusion of other potentially relevant TCOs (particularly those for which the proposed health technology may have a negative treatment effect)
- describe how the TCO is patient relevant and nominate, with evidence, the extent of change that would be considered meaningful (see Appendix 5, minimal clinically important difference)
- state whether the TCO is reversible
- state whether the TCO is itself a substitute for a more clinically relevant outcome (multistep transformation to a subsequent TCO is discouraged)
- provide the units of measurement
- list the measurement tools or criteria used
- provide evidence of reliability from test to test
- explore variability across observers or different measurement tools
- describe the measurement of the comparative treatment effect (e.g. odds ratio, standardised mean difference).

Ensure that the definition and method of measurement of the TCO are consistent across the evidence. Report and discuss any discrepancies.

Relationship between the proposed surrogate measure and the target clinical outcome

When exploring the nature of the PSM–TCO relationship described in subsequent parts of this appendix, comment on the following:

- Is the nature of the PSM–TCO relationship still current?
- Have there been changes to treatments or health care systems over time that may have affected the PSM–TCO relationship?
- Is there any evidence of resistance or tolerance to a health technology, or a waning treatment effect over time? Consider and explain any waning treatment effects, and any effects of having no long-term randomised trials that capture the PSM and the TCO.

Derive the PSM–TCO comparative treatment effect relationship from randomised trials that measure both the PSM and the TCO. If this type of evidence is unavailable, it is difficult to quantify the link between changes in the PSM and changes in the TCO.

Biological reasoning and epidemiological evidence

Biological reasoning

The information request for biological reasoning concerns the disease pathogenesis, and disease or condition pathways, and how the PSM and the TCO relate to them, independent of health

technology actions. To provide confidence that altering the PSM provides clinical benefit, clearly explain the biological relationship between the PSM and the TCO.

Present and discuss the disease or condition pathway, clearly linking the PSM to the TCO. State whether the PSM is a necessary step in the development of the TCO, and discuss how close the development of the PSM is, in both temporal and pathological terms, to the development of the TCO.

Epidemiological evidence

Epidemiological or observational studies support a claimed biological plausibility of the PSM–TCO relationship.

Describe in detail the epidemiological evidence identified, which may include in vitro studies, animal studies, case reports, cross-sectional observational studies, ecological association studies, retrospective observational cohort studies, non-population-based prospective observational cohort studies, or population-based prospective observational cohort studies.

Describe the limitations of the evidence with reference to the study design (e.g. individual-based associations from observational studies are more convincing than ecological associations).

Present the statistical associations, including the nature or shape of the association, the strength of the association and the precision (95% confidence interval [CI]). Report all relevant statistical outputs, such as regression coefficients and R-squared.

Describe and explain any contradictory findings, primarily where the direction of effect changes, or there is a large difference in the magnitude of effect.

Randomised trial data for all health technologies

Identifying relevant trials

Review the literature systematically to find randomised trials that explore the relationship between the PSM and the TCO, irrespective of the health technology examined. Present the search terms, inclusion criteria and the PRISMA flowchart, clearly showing the exclusion of trials. List the excluded trials and reasons for exclusion in an attachment.

From the list of included trials, compile a list of the health technologies, categorised by mechanism of action or class, that act on the PSM (see Table 37). Present the extension studies associated with the identified trials.

For each mechanism of action, discuss the biological reasoning for the effect of the health technology on the PSM. Discuss whether the mechanism of action of the health technology is the same as, or similar to, the pathological mechanism of the disease or condition. Rationalise any lag in onset of the treatment effect and the implications for the PSM or the TCO, or both.

Table 37 Biological reasoning for the effect of the health technology on the proposed surrogate measure

Class of health technology	Mechanism of action	Biological reasoning for the effect of the health technology on the proposed surrogate measure	Trials available, citations (health technology included in each trial)
[add]	[add]	[add]	[add]
[add]	[add]	[add]	[add]

Trial characteristics

For each of the included trials or meta-analyses, discuss the following factors that may affect the estimate of the relationship between the comparative treatment effect on the PSM and the comparative treatment effect on the TCO:

- the quality of the included trials or meta-analyses (present an assessment of the internal validity of the included trials according to the guidance provided in Appendix 3, in an attachment)
- whether relevant trials have been excluded from any meta-analyses or meta-regressions
- whether the analysis of the PSM was designed prospectively or retrospectively.

Present the characteristics of each of the trials as per Table 38.

Table 38 Characteristics of trials included in the assessment of the relationship between the proposed surrogate measure and the target clinical outcome

Trial and date	Patient characteristics	Disease or condition characteristics	Treatment settings	Measurement of proposed surrogate measure and target clinical outcome
[add]	[add]	[add]	[add]	[add]
[add]	[add]	[add]	[add]	[add]

Trial results

Present the results of the randomised trials and the proposed relationship between the comparative treatment effect on the PSM and the comparative treatment effect on the TCO (Table 39). If multiple trials exist for a class of health technologies, clearly show the results of a meta-analysis. Present the results of any meta-regressions, including the intercept and coefficient (and their 95% CIs), the R-squared for trials and for individuals (if individual patient data are available), and the surrogate threshold effect as determined by prediction bands. Justify where a meta-regression is not presented.

Discuss the PSM–TCO comparative treatment effect relationship. Include details of the shape of the relationship (e.g. linear, exponential) and whether there is any evidence of a floor or ceiling effect, below or above which the comparative treatment effect on the PSM no longer predicts a comparative treatment effect on the TCO.

Table 39 Results of randomised trials

Trials/meta-analyses (grouped and meta-analysed by class or mechanism of action)	Baseline value of PSM/final value of PSM[a]	Comparative treatment effect on PSM	Comparative treatment effect on TCO[b]	Proposed relationship (and measure of uncertainty)
[add]	[add]	[add]	[add]	[add]
[add]	[add]	[add]	[add]	[add]

PSM = proposed surrogate measure; TCO = target clinical outcome

a Where the PSM is a continuous variable, present the mean baseline and mean final value for the PSM, separated by treatment arm. Where the PSM is a dichotomous variable, such as progression-free survival, this column may be adapted to show the proportion in each arm achieving the PSM.

b Where the trial has included a placebo, no treatment or best supportive care arm, report the absolute number of TCO events in that arm to give an indication of the baseline risk. A long-standing comparator may also be used as an adequate reference for baseline risk.

Where available, present results of the relationship between the comparative treatment effect for the PSM and the TCO across different trial dates, disease or condition stages, treatment settings and patient populations. State which particular subpopulations do not have trial evidence available (or which subpopulations are not included in the overall study populations). If these subpopulations would have access to the health technology through the proposed funding arrangements (Technical Guidance 3), strongly justify the extrapolation of the PSM–TCO relationship to this subpopulation.

Discuss where the relationship of the comparative treatment effect for the PSM and the TCO differs across trials, health technologies or mechanisms of action. Discuss possible causes of the heterogeneity – for example:

- mechanism of action of the health technology
- population characteristics
- disease or condition characteristics, or severity
- treatment settings
- definition or measurement of the PSM
- definition or measurement of the TCO
- quality of the trial
- nature of the proposed relationship (e.g. linear, asymptotic, floor or ceiling effects).

Multiplicity of pathways

Although unexplained heterogeneity is difficult to interpret, heterogeneity that can be linked to a characteristic will require further consideration, particularly if the cause of the difference in the PSM–TCO relationship differs according to mechanism of action of a health technology, population characteristics, or disease or condition characteristics. Where the PSM–TCO relationship differs between trials, it is likely that the TCO can be affected by an alternative pathological pathway that is more or less prevalent across the included trials.

Where the PSM–TCO comparative treatment effect relationships differ according to the mechanism of action, explain why different health technologies with similar effects on the PSM may result in different effects on the TCO.

Where the PSM–TCO comparative treatment effect relationships differ according to patient characteristics, or disease or condition characteristics, explain why similar changes in the PSM in these subpopulations may result in different effects on the TCO.

Alternative pathological pathways that do not involve the PSM undermine the validity of the PSM. Therefore, where appropriate, exclude trials with health technologies or populations in which the alternative pathway is present if:

- there is compelling evidence of the existence of the alternative pathway (such evidence may be randomised trial evidence linking an alternative PSM with the TCO)

and

- the alternative pathway is not present for the proposed health technology (and the main comparator) or the population in which listing is being sought.

Present evidence to support these claims.

Where trials are removed that have health technologies of different mechanisms of action or populations that do not reflect the proposed listing, present the estimate of the PSM–TCO comparative treatment effect relationship with all trials included as the base case. Remove less-relevant trials through a sensitivity analysis.

Validity of results

For each of the trials, meta-analyses and meta-regressions, compare the observed TCO comparative treatment effect with the predicted effect on the TCO if calculated according to the epidemiological evidence previously identified (Table 40).

Table 40 Comparing randomised trial evidence and epidemiological evidence

Trial, meta-analysis or meta-regression	Comparative treatment effect on PSM	Observed comparative treatment effect on TCO	Predicted comparative treatment effect on TCO after applying the relationship observed in epidemiological studies
[add]	[add]	[add]	[add]
[add]	[add]	[add]	[add]

PSM = proposed surrogate measure; TCO = target clinical outcome

Discuss differences between the observed and predicted comparative treatment effect on the TCO.

Summarising the evidence

Several parameters of the evidence presented are critical to understanding and interpreting the translation of the PSM for the proposed health technology to an estimate of the TCO (Table 41). These are general conditions, outside of which the translation of the PSM to the TCO becomes less certain.

Table 41 Summary of conditions under which the relationship has been determined

Parameter of evidence	Results	Cross-reference
Median baseline value of PSM (IQR)	[add]	[add]
Median final value of PSM (IQR)	[add]	[add]
Median change in PSM (IQR)	[add]	[add]
Median change in PSM for the comparator (IQR)	[add]	[add]
Range of disease or condition severity	[add]	[add]
Range of patient characteristics (e.g. age, sex, race)	[add]	[add]
Range of trial dates	[add]	[add]
Range of TCO event rates (from placebo arms)[a]	[add]	[add]
Range of estimates of the PSM–TCO comparative treatment effect relationship	[add]	[add]

IQR = interquartile range; PSM = proposed surrogate measure; TCO = target clinical outcome

a Placebo, no treatment or best supportive care arms, or long-standing comparator.

Where more than one estimate of the relationship between the comparative treatment effect on the PSM and the comparative treatment effect on the TCO has been established, justify the selection of one estimate for the base case, and present the remainder as sensitivity analyses.

Applying the relationship between comparative treatment effects to the proposed technology

Mechanism of action

When applying the PSM–TCO comparative treatment effect relationship to the trial evidence for the proposed health technology, it is critical that both the proposed health technology and the main comparator have the same mechanism(s) of action as health technologies for which the PSM–TCO comparative treatment effect has been established. Where one or both of the proposed health technology and the main comparator do not share the mechanism of action with the technology used to establish the PSM–TCO relationship, the comparative treatment effect on the PSM may have a very different relationship to the comparative treatment effect on the TCO. Where this is the case, the transformation of the PSM to the TCO will be uncertain.

Explain the mechanism(s) of action and the biological reasoning for the mechanism(s) of action of the proposed health technology and the main comparator on the PSM and the TCO. Identify any differences in the mechanism(s) of action of the proposed health technology compared with that of the main comparator and the health technologies identified in the trial evidence for establishing the PSM–TCO relationship. Clearly explain how any differences will not result in a different measurement of the PSM–TCO comparative treatment effect relationship.

If the proposed health technology and the main comparator are within the same class of health technologies, it is still important to identify differences in physiological effects, and discuss whether different effects can alter the disease or condition process and, hence, the PSM–TCO comparative treatment effect relationship.

Applicability of the evidence

The applicability of the results of the relationship between the treatment effect on the PSM and the treatment effect on the TCO to different populations and stages of disease is not guaranteed. However, evidence of consistency across different populations and stages of disease is supportive. Compare the patient population, disease or condition stages, and circumstances of use for the proposed health technology and the included PSM–TCO studies. If there are differences, justify why the relationship between the treatment effect on the PSM and the treatment effect on the TCO is applicable to the clinical trial(s) of the proposed health technology.

The PSM–TCO comparative treatment effect relationship is uncertain beyond the observed ranges for the PSM in the PSM–TCO studies. Compare the baseline values of the PSM and the comparative treatment effect on the PSM identified in the PSM–TCO literature with that observed for the key trials of the proposed health technology, and discuss.

Estimate the comparative treatment effect for the proposed health technology

Present the proposed health technology's comparative treatment effect (with CIs) on the PSM for each trial and for a pooled analysis. Translate this using the relationship established from the PSM–TCO literature. The comparative treatment effect on the PSM and the estimate of the PSM–TCO relationship will have a degree of uncertainty; thus, capture this in the statistical approach and present as a 95% CI around the estimated comparative treatment effect on the TCO. Do not simply translate the upper and lower CIs of the comparative treatment effect for the PSM observed in the key trial by the point estimate of the PSM–TCO relationship established from the literature, because this does not adequately capture the uncertainty in the estimate of the comparative treatment effect on the TCO.

Discuss the implications of any surrogate threshold effect.

State whether there are any concerns about the duration of the treatment effect.

References

1 Department of Health (2020). *National Strategic Action Plan for Rare Diseases*, Department of Health, Australian Government, Canberra, accessed 8 April 2021.

2 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL & ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 17(5):405-424, doi: 10.1038/gim.2015.30.

3 Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, Tsimberidou AM, Vnencak-Jones CL, Wolff DJ & Younes A (2017). Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *The Journal of Molecular Diagnostics* 19(1):4-23, doi: 10.1016/j.jmoldx.2016.10.002.

4 Yuste R, Goering S, Arcas BAY, Bi G, Carmena JM, Carter A, Fins JJ, Friesen P, Gallant J, Huggins JE, Illes J, Kellmeyer P, Klein E, Marblestone A, Mitchell C, Parens E, Pham M, Rubel A, Sadato N, Sullivan LS, Teicher M, Wasserman D, Wexler A, Whittaker M & Wolpaw J (2017). Four ethical priorities for neurotechnologies and AI. *Nature* 551(7679):159-163, doi: 10.1038/551159a.

5 Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbegovic B, Alonso-Coello P, Post PN, Busse JW, Glasziou P, Christensen R & Schunemann HJ (2013). GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 66(2):158-172, doi: 10.1016/j.jclinepi.2012.01.012.

6 Ciani O, Buyse M, Drummond M, Rasi G, Saad ED & Taylor RS (2017). Time to review the role of surrogate end points in health policy: state of the art and the way forward. *Value in Health* 20(3):487-495, doi: 10.1016/j.jval.2016.10.011.

7 Botkin JR, Teutsch SM, Kaye CI, Hayes M, Haddow JE, Bradley LA, Szegda K & Dotson WD (2010). Outcomes of interest in evidence-based evaluations of genetic tests. *Genetics in medicine* 12(4):228-235, doi: 10.1097/GIM.0b013e3181cdde04.

8 Wurcel V, Cicchetti A, Garrison L, Kip MMA, Koffijberg H, Kolbe A, Leeflang MMG, Merlin T, Mestre-Ferrandiz J, Oortwijn W, Oosterwijk C, Tunis S & Zamora B (2019). The value of diagnostic information in personalised healthcare: a comprehensive concept to facilitate bringing this technology into healthcare systems. *Public Health Genomics* 22(1-2):8-15, doi: 10.1159/000501832.

9 Committee for Medicinal Products for Human Use (2005). *Guidelines on the choice of the non-inferiority margin*. European Medicines Agency, London.

10 Merlin T, Weston A & Tooher R (2009). Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. *BMC Medical Research Methodology* 9:1-8, doi: 10.1186/1471-2288-9-34.

11 Glasziou P, Chalmers I, Rawlins M & McCulloch P (2007). When are randomised trials unnecessary? Picking signal from noise. *BMJ* 334(7589):349-351, doi: 10.1136/bmj.39070.527986.68.

12 Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, Jaeschke R, Rind D, Dahm P, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Murad MH, Schunemann HJ & GRADE Working Group (2011). GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 64(12):1311-1316, doi: 10.1016/j.jclinepi.2011.06.004.

13 Bucher H, Guyatt G, Griffith L & Walter S (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *50* 6:683-691.

14 Signorovitch J, Wu E, Andrew P, Gerrits C, Kantor E, Bao Y, Gupta S & Mulani P (2010). Comparative effectiveness without head-to-head trials. *Pharmacoeconomics* 28(10):935-945.

15 Caro J & Ishak K (2010). No head-to-head trial? Simulate the missing arms. *Pharmacoeconomics* 28(10):957-967.

16 Latimer NR, Abrams KR, Lambert PC, Crowther MJ, Wailoo AJ, Morden JP, Akehurst RL & Campbell MJ (2014). Adjusting survival time estimates to account for treatment switching in randomized controlled trials – an economic evaluation context: methods, limitations, and recommendations. *Medical Decision Making* 34(3):387-402, doi: 10.1177/0272989X13520192.

17 Pitrou I, Boutron I, Ahmad N & Ravaud P (2009). Reporting of safety results in published reports of randomized controlled trials. *Archives of Internal Medicine* 169(19):1756-1761, doi: 10.1001/archinternmed.2009.306.

18 Siontis KC, Siontis GC, Contopoulos-Ioannidis DG & Ioannidis JP (2014). Diagnostic tests often fail to lead to changes in patient outcomes. *J Clin Epidemiol* 67(6):612-621, doi: 10.1016/j.jclinepi.2013.12.008.

19 US Preventive Services Taskforce (2015). *Procedure manual*, accessed April 2021, https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual.

20 Merlin T, Lehman S, Hiller JE & Ryan P (2013). The "linked evidence approach" to assess medical tests: a critical analysis. *International Journal of Technology Assessment in Health Care* 29(3):343-350, doi: 10.1017/s0266462313000287.

21 Doust J (2010). Qualification versus validation of biomarkers. *Scandinavian Journal of Clinical and Laboratory Investigation* 70(Suppl 242):40-43, doi: 10.3109/00365513.2010.493380.

22 Pletcher MJ & Pignone M (2011). Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. *Circulation* 123(10):1116-1124, doi: 10.1161/CIRCULATIONAHA.110.943860.

23 Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC & Bossuyt PMM (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 155(8):529-536, doi: 10.7326/0003-4819-155-8-201110180-00009.

24 Glasziou P, Irwig L & Deeks JJ (2008). When should a new test become the current reference standard? *Annals of Internal Medicine* 149(11):816-821, doi: 10.7326/0003-4819-149-11-200812020-00009.

25 Trikalinos TA & Balion CM (2012). Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *Journal of General Internal Medicine* 27(Suppl 1):S67-S75, doi: 10.1007/s11606-012-2031-7.

26 Mathes T & Pieper D (2019). An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. *Systematic reviews* 8(1):1-8, doi: 10.1186/s13643-019-1131-4.

27 National Health and Medical Research Council (2009). *NHMRC additional levels of evidence and grades for recommendations for developers of guidelines*. NHMRC, Canberra.

28 Jonas DE, Wilt TJ, Taylor BC, Wilkins TM & Matchar DB (2012). Chapter 11: challenges in and principles for conducting systematic reviews of genetic tests used as predictive indicators. *Journal of General Internal Medicine* 27(Suppl 1):83-93, doi: 10.1007/s11606-011-1898-z.

29 Moons KGM (2010). Criteria for scientific evaluation of novel markers: a perspective. *Clinical Chemistry* 56(4):537-541, doi: 10.1373/clinchem.2009.134155.

30 Altman DG, McShane LM, Sauerbrei W & Taube SE (2012). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. *PLoS Medicine* 9(5):1-32, doi: 10.1371/journal.pmed.1001216.

31 Riley RD, Sauerbrei W & Altman DG (2009). Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. *British Journal of Cancer* 100(8):1219-1229, doi: 10.1038/sj.bjc.6604999.

32 Rotily M & Roze S (2013). What is the impact of disease prevalence upon health technology assessment? *Best Practice & Research: Clinical Gastroenterology* 27(6):853-865, doi: 10.1016/j.bpg.2013.08.018.

33 Staub LP, Lord SJ, Simes RJ, Dyer S, Houssami N, Chen RY & Irwig L (2012). Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation. *BMC Medical Research Methodology* 12:1-9, doi: 10.1186/1471-2288-12-12.

34 Australian Commission on Safety and Quality in Health Care (2013). *Medical practice variation: background paper*. ACSQHC, Sydney.

35 Wennberg J, Brownlee S, Fischer E, Skinner J & Weinstein J (2008). *Improving quality and curbing health care spending: opportunities for the Congress and the Obama Administration: an agenda for change*. Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire.

36 Hajjaj FM, Salek MS, Basra MK & Finlay AY (2010). Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. *Journal of the Royal Society of Medicine* 103(5):178-187, doi: 10.1258/jrsm.2010.100104.

37 Carter SM, Degeling C, Doust J & Barratt A (2016). A definition and ethical evaluation of overdiagnosis. *Journal of Medical Ethics* 42(11):705-714, doi: 10.1136/medethics-2015-102928.

38 Bell KJL, Doust J, Glasziou P, Cullen L, Harris IA, Smith L, Buchbinder R & Barratt A (2019). Recognizing the potential for overdiagnosis: are high-sensitivity cardiac troponin assays an example? *Annals of Internal Medicine* 170(4):259-261, doi: 10.7326/M18-2645.

39 Croft P, Altman DG, Deeks JJ, Dunn KM, Hay AD, Hemingway H, LeResche L, Peat G, Perel P, Petersen SE, Riley RD, Roberts I, Sharpe M, Stevens RJ, Van Der Windt DA, Von Korff M & Timmis A (2015). The science of clinical practice: disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice. *BMC Medicine* 13(1):1-8, doi: 10.1186/s12916-014-0265-4.

40 Morton V & Torgerson DJ (2005). Regression to the mean: treatment effect without the intervention. *Journal of Evaluation in Clinical Practice* 11(1):59-65, doi: 10.1111/j.1365-2753.2004.00505.x.

41 Doust JA, Bell KJL & Glasziou PP (2020). Potential consequences of changing disease classifications. *JAMA: Journal of the American Medical Association* 323(10):921-922, doi: 10.1001/jama.2019.22373.

42 Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG & PROGRESS Group (2013). Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Medicine* 10(2):1-9, doi: 10.1371/journal.pmed.1001381.

43 Pepe MS, Janes H, Longton G, Leisenring W & Newcomb P (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* 159(9):882-890, doi: 10.1093/aje/kwh101.

44 Janes H, Pepe MS, Bossuyt PM & Barlow WE (2011). Measuring the performance of markers for guiding treatment decisions. *Annals of Internal Medicine* 154(4):253-259, doi: 10.7326/0003-4819-154-4-201102150-00006.

45 Bonetti M & Gelber RD (2000). A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. *Statistics in Medicine* 19(19):2595-2609, doi: 10.1002/1097-0258(20001015)19:19<2595::aid-sim562>3.0.co;2-m.

46 Bell KJ, Glasziou PP, Hayen A & Irwig L (2014). Criteria for monitoring tests were described: validity, responsiveness, detectability of long-term change, and practicality. *J Clin Epidemiol* 67(2):152-159, doi: 10.1016/j.jclinepi.2013.07.015.

47 Doust J & Glasziou P (2013). Monitoring in clinical biochemistry. *Clinical Biochemist Reviews* 34(2):85-92.

48 Bell KJ, Irwig L, Craig JC & Macaskill P (2008). Use of randomised trials to decide when to monitor response to new treatment. *BMJ* 336(7640):361-365, doi: 10.1136/bmj.39476.623611.25.

49 Glasziou PP, Irwig L, Heritier S, Simes RJ & Tonkin A (2008). Monitoring cholesterol levels: measurement error or true change? *Annals of Internal Medicine* 148(9):656-661, doi: 10.7326/0003-4819-148-9-200805060-00005.

50 Cahan EM, Hernandez-Boussard T, Thadaney-Israni S & Rubin DL (2019). Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digital Medicine* 2:1-6, doi: 10.1038/s41746-019-0157-2.

51 Park SH, Kim Y-H, Lee JY, Yoo S & Kim CJ (2019). Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. *Science Editing* 6(2):91-98, doi: 10.6087/kcse.164.

52 Parikh RB, Obermeyer Z & Navathe AS (2019). Regulation of predictive analytics in medicine. *Science* 363(6429):810-812, doi: 10.1126/science.aaw0029.

53 Sendak M, Gao M, Nichols M, Lin A & Balu S (2019). Machine learning in health care: a critical appraisal of challenges and opportunities. *Journal for Electronic Health Data and Methods* 7(1):1-4, doi: 10.5334/egems.287.

54 Fazel S & Wolf A (2018). Selecting a risk assessment tool to use in practice: a 10-point guide. *Evidence-Based Mental Health* 21(2):41-43, doi: 10.1136/eb-2017-102861.

55 Gonzalez-McQuire S, Dimopoulos M-A, Weisel K, Bouwmeester W, Hájek R, Campioni M, Bennison C, Xu W, Pantiri K & Hensen M (2019). Development of an initial conceptual model of multiple myeloma to support clinical and health economics decision making. *MDM policy & Practice* 4(1):1-22, doi: 10.1177/2381468318814253.

56 Haji Ali Afzali H, Bojke L & Karnon J (2018). Model structuring for economic evaluations of new health technologies. *Pharmacoeconomics* 36(11):1309-1319, doi: 10.1007/s40273-018-0693-7.

57 Haji Ali Afzali H, Karnon J, Theou O, Beilby J, Cesari M & Visvanathan R (2019). Structuring a conceptual model for cost-effectiveness analysis of frailty interventions. *PLoS One* 14(9):1-13, doi: 10.1371/journal.pone.0222049.

58 Kaltenthaler E, Tappenden P, Paisley S & Squires H (2011). *NICE DSU technical support document 13: identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models*, Decision Report Unit, School of Health and Related Research, University of Sheffield, Sheffield, accessed 8 April 2021.

59 Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M & ISPOR-SMDM Modeling Good Research Practices Task Force (2012). Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 2. *Value in Health* 15(6):804-811, doi: 10.1016/j.jval.2012.06.016.

60 Tabberer M, Gonzalez-McQuire S, Muellerova H, Briggs AH, Rutten-van Molken MPMH, Chambers M & Lomas DA (2017). Development of a conceptual model of disease progression for use in economic modeling of chronic obstructive pulmonary disease. *Medical Decision Making* 37(4):440-452, doi: 10.1177/0272989X16662009.

61 Tappenden P & Chilcott JB (2014). Avoiding and identifying errors and other threats to the credibility of health economic models. *Pharmacoeconomics* 32(10):967-979, doi: 10.1007/s40273-014-0186-2.

62 Vemer P, Corro Ramos I, van Voorn GAK, Al MJ & Feenstra TL (2016). AdViSHE: A validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics* 34(4):349-361, doi: 10.1007/s40273-015-0327-2.

63 Barton P, Bryan S & Robinson S (2004). Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research & Policy* 9(2):110-118, doi: 10.1258/135581904322987535.

64 Detsky AS, Naglie G, Krahn MD, Redelmeier DA & Naimark D (1997). Primer on medical decision analysis: part 2 – building a tree. *Medical Decision Making* 17(2):126-135, doi: 10.1177/0272989X9701700202.

65 Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ & Kuntz KM (2012). State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 3. *Medical Decision Making* 32(5):690-700, doi: 10.1177/0272989X12455463.

66 Williams C, Lewsey JD, Mackay DF & Briggs AH (2017). Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. *Medical Decision Making* 37(4):427-439, doi: 10.1177/0272989X16670617.

67 Woods B, Sideris E, Palmer S, Latimer NR & Soares M (2017). *NICE DSU technical support document 19: partitioned survival analysis for decision modelling in health care: a critical review*, Decision Support Unit, School of Health and Related Research, University of Sheffield, Sheffield, accessed 8 April 2021.

68 Karnon J & Haji Ali Afzali H (2014). When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics* 32(6):547-558, doi: 10.1007/s40273-014-0147-9.

69 Karnon J, Stahl J, Brennan A, Caro JJ, Mar J & Moller J (2012). Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 4. *Medical Decision Making* 32(5):701-711, doi: 10.1177/0272989X12455462.

70 Tsoi B, O'Reilly D, Jegathisawaran J, Tarride J-E, Blackhouse G & Goeree R (2015). Systematic narrative review of decision frameworks to select the appropriate modelling approaches for health economic evaluations. *BMC Research Notes* 8:1-15, doi: 10.1186/s13104-015-1202-0.

71 Ghabri S, Stevenson M, Moller J & Caro JJ (2019). Trusting the results of model-based economic analyses: is there a pragmatic validation solution? *Pharmacoeconomics* 37(1):1-6, doi: 10.1007/s40273-018-0711-9.

72 Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD & ISPOR-SMDM Modeling Good Research Practices Task Force (2012). Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group – 6. *Medical Decision Making* 32(5):722-732, doi: 10.1177/0272989X12458348.

73 Fleurence RL & Hollenbeak CS (2007). Rates and probabilities in economic modelling: transformation, translation and appropriate application. *Pharmacoeconomics* 25(1):3-6, doi: 10.2165/00019053-200725010-00002.

74 Royston P & Lambert PC (2011). *Flexible parametric survival analysis using Stata: beyond the Cox model*. Stata Press.

75 Bagust A & Beale S (2014). Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Medical Decision Making* 34(3):343-351, doi: 10.1177/0272989X13497998.

76 Grieve R, Hawkins N & Pennington M (2013). Extrapolation of survival data in cost-effectiveness analyses: improving the current state of play. *Medical Decision Making* 33(6):740-742, doi: 10.1177/0272989X13492018.

77 Karnon J & Vanni T (2011). Calibrating models in economic evaluation: a comparison of alternative measures of goodness of fit, parameter search strategies and convergence criteria. *Pharmacoeconomics* 29(1):51-62, doi: 10.2165/11584610-000000000-00000.

78 Latimer NR (2013). Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Medical Decision Making* 33(6):743-754, doi: 10.1177/0272989X12472398.

79 Royston P (2001). Flexible parametric alternatives to the Cox model, and more. *Stata Journal* 1(1):1-28, doi: 10.1177/1536867X0100100101.

80 Whyte S, Walsh C & Chilcott J (2011). Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Medical Decision Making* 31(4):625-641, doi: 10.1177/0272989X10384738.

81 Barnett AG, van der Pols JC & Dobson AJ (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* 34(1):215-220, doi: 10.1093/ije/dyh299.

82 Brazier J, Ara R, Azzabi I, Busschbach J, Chevrou-Severac H, Crawford B, Cruz L, Karnon J, Lloyd A, Paisley S & Pickard AS (2019). Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force report. *Value in Health* 22(3):267-275, doi: 10.1016/j.jval.2019.01.004.

83 Becker F, Van El CG, Ibarreta D, Zika E, Hogarth S, Borry P, Cambon-Thomsen A, Cassiman JJ, Evers-Kiebooms G, Hodgson S, Janssens ACJW, Kaariainen H, Krawczak M, Kristoffersson U,

Lubinski J, Patch C, Penchaszadeh VB, Read A, Rogowski W, Sequeiros J, Tranebjaerg L, Van Langen IM, Wallace H, Zimmern R, Schmidtke J & Cornel MC (2011). Genetic testing and common disorders in a public health framework: how to assess relevance and possibilities. *European Journal of Human Genetics* 19(Suppl 1):S6-S44, doi: 10.1038/ejhg.2010.249.

84 Burke W, Laberge AM & Press N (2010). Debating clinical utility. *Public Health Genomics* 13(4):215-223, doi: 10.1159/000279623.

85 Korngiebel DM, Zech JM, Chappelle A, Burke W, Carline JD, Gallagher TH & Fullerton SM (2019). Practice implications of expanded genetic testing in oncology. *Cancer Investigation* 37(1):39-45, doi: 10.1080/07357907.2018.1564926.

86 Dixon-Woods M, Agarwal S, Jones D, Young B & Sutton A (2005). Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research & Policy* 10(1):45-53, doi: 10.1258/1355819052801804.

87 Harden A, Thomas J, Cargo M, Harris J, Pantoja T, Flemming K, Booth A, Garside R, Hannes K & Noyes J (2018). Cochrane Qualitative and Implementation Methods Group guidance series – paper 5: methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. *J Clin Epidemiol* 97:70-78, doi: 10.1016/j.jclinepi.2017.11.029.

88 Booth A (2020). *A methodological update on the use of qualitative evidence in health technology assessment*, Decision Support Unit, School of Health and Related Research, University of Sheffield, Sheffield, accessed 8 April 2021.

89 Saarni SI, Braunack-Mayer A, Hofmann B & van der Wilt GJ (2011). Different methods for ethical analysis in health technology assessment: an empirical study. *International Journal of Technology Assessment in Health Care* 27(4):305-312, doi: 10.1017/S0266462311000444.

90 Assasi N, Tarride JE, O'Reilly D & Schwartz L (2016). Steps toward improving ethical evaluation in health technology assessment: a proposed framework. *BMC Medical Ethics* 17(1):1-16, doi: 10.1186/s12910-016-0118-0.

91 Hofmann B (2005). Toward a procedure for integrating moral issues in health technology assessment. *International Journal of Technology Assessment in Health Care* 21(3):312-318, doi: 10.1017/s0266462305050415.

92 Norheim OF, Baltussen R, Johri M, Chisholm D, Nord E, Brock D, Carlsson P, Cookson R, Daniels N, Danis M, Fleurbaey M, Johansson KA, Kapiriri L, Littlejohns P, Mbeeli T, Rao KD, Edejer TT & Wikler D (2014). Guidance on priority setting in health care (GPS-Health): the inclusion of equity criteria not captured by cost-effectiveness analysis. *Cost Effectiveness and Resource Allocation* 12:1-8, doi: 10.1186/1478-7547-12-18.

93 Hofmann B, Droste S, Oortwijn W, Cleemput I & Sacchini D (2014). Harmonization of ethics in health technology assessment: a revision of the Socratic approach. *International Journal of Technology Assessment in Health Care* 30(1):3-9, doi: 10.1017/S0266462313000688.

94 Staniszewska S & Werkö S (2017). Patient-based evidence in HTA. In: *Patient involvement in health technology assessment*, Facey K, Ploug Hansen H & Single A (eds), Adis, Singapore, 43-50.

95 De Groot S, Rijnsburger AJ, Versteegh MM, Heymans JM, Kleijnen S, Redekop WK & Verstijnen IM (2015). Which factors may determine the necessary and feasible type of effectiveness evidence? A mixed methods approach to develop an instrument to help coverage decision-makers. *BMJ Open* 5(7):1-10, doi: 10.1136/bmjopen-2014-007241.

96 Hinde S & Spackman E (2015). Bidirectional citation searching to completion: an exploration of literature searching methods. *Pharmacoeconomics* 33(1):5-11, doi: 10.1007/s40273-014-0205-3.

97 Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Rader T, Shokraneh F, Thomas J & Wieland L (2021). Chapter 4: Searching for and selecting studies. In: *Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021)*, Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M & Welch V (eds), The Cochrane Collaboration.

98 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J & Moher D (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 62(10):e1-e34, doi: 10.1016/j.jclinepi.2009.06.006.

99 Moher D, Liberati A, Tetzlaff J, Altman DG & PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 62(10):1006-1012, doi: 10.1016/j.jclinepi.2009.06.005.

100 Tacconelli E (2010). Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases* 10(4):226, doi: 10.1016/S1473-3099(10)70065-7.

101 Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E & Henry DA (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358:1-9, doi: 10.1136/bmj.j4008.

102 Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hrobjartsson A, Kirkham J, Juni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schunemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF & Higgins JP (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355:1-7, doi: 10.1136/bmj.i4919.

103 Hayden JA, van der Windt DA, Cartwright JL, Cote P & Bombardier C (2013). Assessing bias in studies of prognostic factors. *Annals of Internal Medicine* 158(4):280-286, doi: 10.7326/0003-4819-158-4-201302190-00009.

104 Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S & Guyatt GH (2011). GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 64(4):401-406, doi: 10.1016/j.jclinepi.2010.07.015.

105 Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J & Schunemann HJ (2011). GRADE guidelines: 4. Rating the quality of evidence – study limitations (risk of bias). *J Clin Epidemiol* 64(4):407-415, doi: 10.1016/j.jclinepi.2010.07.017.

106 Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J & Schunemann HJ (2011). GRADE guidelines: 6. Rating the quality of evidence – imprecision. *J Clin Epidemiol* 64(12):1283-1293, doi: 10.1016/j.jclinepi.2011.01.012.

107 Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schunemann HJ & GRADE Working Group (2011). GRADE guidelines: 7. Rating the quality of

evidence – inconsistency. *J Clin Epidemiol* 64(12):1294-1302, doi: 10.1016/j.jclinepi.2011.03.017.

108 Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schunemann HJ & GRADE Working Group (2011). GRADE guidelines: 8. Rating the quality of evidence – indirectness. *J Clin Epidemiol* 64(12):1303-1310, doi: 10.1016/j.jclinepi.2011.04.014.

109 Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA & Schunemann HJ (2011). GRADE guidelines: 5. Rating the quality of evidence – publication bias. *J Clin Epidemiol* 64(12):1277-1282, doi: 10.1016/j.jclinepi.2011.01.011.

110 Singh S, Chang SM, Matchar DB & Bass EB (2012). Chapter 7: grading a body of evidence on diagnostic tests. *Journal of General Internal Medicine* 27(Suppl 1):S47-S55, doi: 10.1007/s11606-012-2021-9.

111 Lewin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, Garside R, Noyes J, Booth A, Tuncalp O, Wainwright M, Flottorp S, Tucker JD & Carlsen B (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings – paper 2: how to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. *Implementation Science* 13(Suppl 1):11-23, doi: 10.1186/s13012-017-0689-2.

112 Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gulmezoglu M, Noyes J, Booth A, Garside R & Rashidian A (2015). Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Medicine* 12(10):1-18, doi: 10.1371/journal.pmed.1001895.

113 Booth A, Lewin S, Glenton C, Munthe-Kaas H, Toews I, Noyes J, Rashidian A, Berg RC, Nyakang'o B, Meerpohl JJ & GRADE-CERQual Coordinating Team (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings – paper 7: understanding the potential impacts of dissemination bias. *Implementation Science* 13(Suppl 1):63-70, doi: 10.1186/s13012-017-0694-5.

114 Cleophas TJ & Zwinderman AH (2009). Meta-analyses of diagnostic studies. *Clinical Chemistry and Laboratory Medicine* 47(11):1351-1354, doi: 10.1515/cclm.2009.317.

115 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM & Zwinderman AH (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58(10):982-990, doi: 10.1016/j.jclinepi.2005.02.022.

116 Hoyer A & Kuss O (2015). Meta-analysis of diagnostic tests accounting for disease prevalence: a new model using trivariate copulas. *Statistics in Medicine* 34(11):1912-1924, doi: 10.1002/sim.6463.

117 Hoyer A & Kuss O (2018). Meta-analysis for the comparison of two diagnostic tests to a common gold standard: a generalized linear mixed model approach. *Statistical Methods in Medical Research* 27(5):1410-1421, doi: 10.1177/0962280216661587.

118 Takwoingi Y, Guo B, Riley RD & Deeks JJ (2017). Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Statistical Methods in Medical Research* 26(4):1896-1911, doi: 10.1177/0962280215592269.

119 Trikalinos TA, Balion CM, Coleman CI, Griffith L, Santaguida PL, Vandermeer B & Fu R (2012). Chapter 8: meta-analysis of test performance when there is a "gold standard". *Journal of General Internal Medicine* 27(Suppl 1):S56-S66, doi: 10.1007/s11606-012-2029-1.

120 Nyaga VN, Arbyn M & Aerts M (2014). Metaprop: a Stata command to perform meta-analysis of binomial data. *Archives of Public Health* 72(1):1-10, doi: 10.1186/2049-3258-72-39.

121 Glas AS, Lijmer JG, Prins MH, Bonsel GJ & Bossuyt PM (2003). The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 56(11):1129-1135, doi: 10.1016/s0895-4356(03)00177-x.

122 Lee J, Kim KW, Choi SH, Huh J & Park SH (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers – part II. Statistical methods of meta-analysis. *Korean Journal of Radiology* 16(6):1188-1196, doi: 10.3348/kjr.2015.16.6.1188.

123 Bossuyt PM & Leeflang MM (2008). Developing criteria for including studies. In: *Cochrane handbook for systematic reviews of diagnostic test accuracy*, The Cochrane Collaboration.

124 Millard LA, Flach PA & Higgins JP (2016). Machine learning to assist risk-of-bias assessments in systematic reviews. *International Journal of Epidemiology* 45(1):266-277, doi: 10.1093/ije/dyv306.

125 Swets JA (1988). Measuring the accuracy of diagnostic systems. *Science* 240(4857):1285-1293, doi: 10.1126/science.3287615.

126 Cochran WG (1954). The combination of estimates from different experiments. *Biometrics* 10(1):101-129, doi: 10.2307/3001666.

127 Higgins JPT, Thompson SG, Deeks JJ & Altman DG (2003). Measuring inconsistency in meta-analyses. *BMJ* 327(7414):557-560, doi: 10.1136/bmj.327.7414.557.

128 Egger M, Smith GD, Schneider M & Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315(7109):629-634, doi: 10.1136/bmj.315.7109.629.

129 Begg CB & Mazumdar M (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50(4):1088-1101, doi: 10.2307/2533446.

130 van Enst W, Ochodo E, Scholten R, Hooft L & Leeflang M (2014). Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Medical Research Methodology* 14:70.

131 Deeks JJ, Macaskill P & Irwig L (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 58(9):882-893, doi: 10.1016/j.jclinepi.2005.01.016.

132 Macaskill P, Gatsonis C, Deeks JJ, Harbord RM & Takwoingi Y (2010). Analysing and presenting results. In: *Cochrane handbook for systematic reviews of diagnostic test accuracy*, Deeks JJ, Bossuyt PM & Gatsonis C (eds), The Cochrane Collaboration.