



Australian Government

Department of Health

**Technical Guidelines for preparing
assessment reports for the Medical
Services Advisory Committee –
Service Type: Investigative
(Version 3.0)**

Final

July 2017

Medical Services Advisory Committee

Foreword

The Medical Services Advisory Committee (MSAC) is an independent committee that provides advice to the Minister for Health on the strength of the evidence relating to the comparative safety, clinical effectiveness and cost-effectiveness of any new or existing medical services or technology, and the circumstances under which public funding should be supported through listing on the Medical Benefits Schedule (MBS).

To achieve this, MSAC undertakes health technology assessments (HTA) using the best available evidence to assess proposals for their comparative safety, clinical effectiveness, and cost effectiveness.

Applications for investigative services provide applicants with a number of challenges, requiring them to prove that both the proposed service provides accurate, meaningful information and also that the information improves the subsequent treatment (and health outcomes) of patients.

This document provides detailed advice to assist applicants with determining content and presentation of submissions of evidence for consideration by MSAC and the Evaluation Sub-committee (ESC).

Chair
Medical Services Advisory Committee

Contents

Foreword	i
Contacts	x
Record of updates	xi
Abbreviations	xii
PART I General information	1
1 Medical Services Advisory Committee	2
1.1 Purpose and roles of MSAC.....	2
1.2 Membership of MSAC.....	2
1.3 MSAC sub-committees	2
1.4 Overview of MSAC processes.....	3
1.4.1 Regulatory framework	3
1.4.2 The application and assessment process	3
1.4.3 Sources of advice	3
1.4.4 Publication of assessment report.....	3
1.4.5 Timing of MSAC processes.....	4
1.4.6 MSAC appraisal.....	4
2 Introduction to the Guidelines	5
2.1 Structure of these Guidelines	5
2.2 Associated documents.....	5
2.3 What is an investigative medical service?	6
2.4 Writing and style conventions used in these Guidelines.....	6
2.5 The future.....	7
3 Rationale and basis for the economic evaluation in the Australian context	8
3.1 Analysis of cost-effectiveness.....	8
3.2 Australian context	8
3.3 Relevant factors influencing MSAC decision making.....	8
3.4 Flexibility in interpretation of these Guidelines.....	9

4	Organisation of a standard assessment report	10
4.1	Choice of information	10
4.2	Overview of a standard assessment report	11
4.2.1	Sections of a standard assessment report	11
4.3	Presentation of the assessment report	11
4.3.1	Standard assessment report	11
4.3.2	Two stage approach to an assessment report	13
4.3.3	Co-dependent/integrated assessment reports	13
5	Lodging an assessment report	14
5.1	Assessment report checklist	14
5.1.1	Information requirements	14
5.1.2	Information requests	14
5.2	Provision of information to allow independent verification of computer analyses	16
5.2.1	Information requirements	16
5.3	Provision of information after lodgement of the assessment report	16
5.3.1	Post-lodgement communication with MSAC	16
5.3.2	Provision of information sourced from the TGA after lodgement of the assessment report	16
	Part II Preferred clinical and economic evidence for proposed medical services to be considered by MSAC	17
Section A	Details of the proposed investigative medical service and its intended use on the Medical Benefits Schedule (or for other public funding)	18
A1	Address all items in the agreed PICO Confirmation.....	18
A2	Proposed medical service	18
A3	Proposed MBS listing or other public funding sought.....	19
A4	Proposed population.....	19
A5	Comparator details	19
A6	Clinical management algorithm(s).....	19
A6.1	Algorithms for intended and current contexts.....	20
A7	Differences between the proposed medical service and the main comparator	21
A8	Clinical claim	21
A9	Summarise the primary elements of the decision analysis (PICO).....	21

Section B Clinical evaluation for the proposed investigative medical service	22
B1 Direct Evidence.....	23
B1.1 Description of search strategies (direct evidence)	23
B1.2 Presentation of direct evidence	24
B2 Overview of linked evidence approach.....	25
B2.1 Basis for linked evidence	25
B2.2 Steps for linked analysis	26
B3 Diagnostic performance	27
B3.1 Identification of a reference standard.....	28
B3.2 Description of search strategies (accuracy studies)	29
B3.3 Listing of all accuracy studies.....	31
B3.3a Listing accuracy studies of direct comparison (where relevant).....	31
B3.3b Listing accuracy studies for inclusion in an indirect comparison (where relevant)	35
B3.4 Assessment of the measures taken by investigators to minimise bias in accuracy studies.....	38
B3.5 Characteristics of accuracy studies	42
B3.6 Systematic overview of the results of the accuracy studies	44
B3.7 Extended assessment of reliability evidence.....	49
B3.8 Presenting a concordance analysis in the absence of a reference standard	51
B3.9 Interpretation of evidence on diagnostic performance.....	52
B4 Clinical validity.....	53
B4.1 Measures of clinical validity	53
B4.2 Supplementary analysis if the service is to be used to inform prognosis or predisposition	54
B5 Clinical utility	55
B5.1 Impact on clinical management (therapeutic efficacy).....	55
B5.2 Therapeutic effectiveness (including impact of effect modification).....	57
B6 Impact of repeat testing/use of service as part of a monitoring strategy.....	58
B7 Extended assessment of comparative harms	60
B8 Overall interpretation of all the clinical evidence presented.....	61
Section C Translation Issues.....	63
C1 Identification of issues to be addressed.....	64
C2 Focused analytical plan	69
C3 Results of pre-modelling studies.....	73

C4	Relationship of each pre-modelling study to the economic evaluation	75
Section D	Economic evaluation for the main indication.....	76
	Introduction.....	76
D1	Overview of the economic evaluation.....	78
D2	Population and circumstances of use reflected in the economic evaluation	85
D3	Structure and rationale of the economic evaluation	90
D4	Variables in the economic evaluation	99
D5	Results of the economic evaluation.....	105
D6	Sensitivity analyses	113
Section E	Estimated extent of use and financial implications.....	115
	Introduction.....	115
E1	Justification of the selection of sources of data	118
E2	Estimation of use and costs of the proposed medical service	120
E3	Estimation of changes in use and cost of other medical services.....	123
E4	Estimated financial implications for the MBS	125
E5	Identification, estimation and reduction of uncertainty	125
Section F	Options to present additional relevant information	129
	Introduction.....	129
F1	Other relevant factors.....	129
Part III	Alternative clinical evidence for proposed medical services to be considered by MSAC	132
Section D(i)	Economic evaluation for the main indication: presenting a cost-minimisation approach.....	133
	Introduction.....	133
D(i)1	Presentation of a cost-minimisation analysis or a cost analysis.....	134

Appendices	136
Appendix 1	Relevant factors influencing provision of advice by MSAC 137
Appendix 2	Expert opinion..... 139
A2.1	Uses of expert opinion 139
A2.2	Presenting expert opinion..... 140
A2.3	Describing the collection and collation of expert opinion 140
Appendix 3	Assessment of non-inferiority 144
A3.1	Introduction..... 144
A3.2	Service delivery information..... 144
A3.3	Non-inferiority threshold 144
A3.4	Method of analysis 144
A3.5	Presenting an assessment of non-inferiority 145
A3.6	Assessing comparative harms in the context of non-inferiority..... 146
A3.7	Interpretation of the clinical evidence..... 146
Appendix 4	Utility valuation of health outcomes..... 147
A4.1	Use of health-related QALYs gained and cost-utility analysis 147
A4.2	Obtaining utility weights..... 148
A4.3	Trial-based utility valuation of health outcomes..... 149
A4.4	Scenario-based utility valuation of health outcomes 151
A4.5	Other methods for obtaining utilities 158
Appendix 5	Monetary valuation of health outcomes..... 161
A5.1	Preference for cost-utility analyses over cost-benefit analyses..... 161
A5.2	Scenario-based monetary valuation of health outcomes..... 162
Appendix 6	Including non-health care resources and non-health outcomes in a supplementary analysis 167
A6.1	Identifying, measuring and valuing non-health care resources..... 167
A6.2	Identifying, measuring and valuing non-health outcomes 167
A6.3	Resources and outcomes to be excluded..... 170

Appendix 7 – Codependent applications across MSAC and PBAC	171
Introduction.....	171
What are codependent technologies?.....	171
When is a codependent submission required?	172
A7.1 Overview of information requested in codependent submissions.....	175
Types of evidence	175
Integrated codependent submissions.....	175
Streamlined codependent submissions.....	180
A7.2 Specific codependent technology information requests.....	181
Section 1 – Context.....	181
Details about the biomarker, the test and the medicine	181
Rationale for the codependency	185
Proposed impact of codependent technologies on current clinical practice	188
Section 2 – Clinical evaluation	190
Direct evidence approach.....	191
Section 2a Evidence of prognostic effect of the biomarker.....	191
Section 2d Clinical evaluation of the codependent technologies (combined)	191
Linked-evidence approach	194
Section 2b Test performance and accuracy.....	194
Section 2c Change in clinical management	198
Section 2d Clinical evaluation of the codependent technologies (separate).....	198
Applicability of the effectiveness of the codependent technology	201
Section 3 – Economic evaluation.....	202
Structure of the model.....	202
Transition probabilities relating to test outcomes	203
Resource items and costs included in the model.....	206
Uncertainties in the model	208
Section 4 – Use of the medicine in practice.....	209
59 (O) Budget impact analysis incorporating both MBS and PBS components	209
60 (O) Epidemiology estimate for disease burden	209
61 (T) Likely use and overall financial cost of the test.....	210
62 (O) Other MBS costs	210
 Appendix 8 Screening investigative medical services.....	 211

Tables and Figures

Figure 4.1 Sections of a standard assessment report.....	12
--	----

Table 5.1	Checklist of information to be included in an investigative assessment report	15
Table B3.3.1	Summary of identification of accuracy studies of direct comparison from the search of the published literature	32
Table B3.3.2	Accuracy studies of direct comparison (and associated reports) presented in the assessment report	33
Table B3.3.3	Reasons to exclude each accuracy study of direct comparison from further detailed assessment.....	34
Table B3.3.4	Comparative summary of characteristics of accuracy studies of direct comparison.....	34
Table B3.3.5	Comparative summary of results of accuracy studies of direct comparison.....	34
Table B3.3.6	Accuracy studies (and associated reports) presented in the assessment report	36
Table B3.3.7	Reasons to exclude each accuracy study from the indirect comparison.....	37
Box B3.4.1	Indicative checklist for assessing the quality of specific accuracy studies	39
Table B3.4.1	Summary of the measures undertaken to minimise bias.....	41
Table B3.4.2	Flow of participants in each accuracy study of direct comparison under consideration	42
Table B3.5.1	Eligibility criteria in the accuracy studies under consideration	43
Table B3.5.2	Characteristics of participants in the accuracy studies.....	43
Table B3.6.1	Results of accuracy studies directly comparing proposed and comparator investigative medical services	48
Table B3.6.2	Summary of results of the indirect comparison	49
Table B3.7.1	Characteristics and summary results of accuracy studies reporting reliability / reproducibility.....	51
Table B8.1	Classification of the therapeutic relativity of the proposed investigative medical service over its main comparator	62
Figure D1	Key information requests for assessment report Section D of a standard assessment for MSAC	77
Table D1.1	Classification of a service under MSAC consideration/Classification of the effectiveness of the proposed medical service over its main comparator and guide to the suitable type of economic evaluation.....	80
Table D1.2	Type of economic evaluation that should be presented for various classifications of a service under MSAC consideration.....	82
Table D2.1	Definitions for populations and circumstance of use that should be taken into account in the evaluation.....	85
Table D2.2	Comparison of characteristics of trial and requested populations and circumstances of use	87

Table D2.3 Example of a comparison of the characteristics of target, study and wider populations and settings	89
Figure D3.1 Example of the structure of a basic economic evaluation.....	92
Table D3.1 Key elements of the base-case economic evaluation.....	94
Table D4.1 List of health care resource items and unit costs included in the economic evaluation	102
Table D5.1 List of health care resource items and summary of cost impacts in the economic evaluation	106
Table D5.2 List of health care resource items and summary of cost impacts for each health state in a state transition model.....	107
Table D5.3 List of health states and summary of cost impacts included in the economic evaluation	107
Table D5.4 List of health states and summary of health outcomes included in the economic evaluation	108
Table D5.5 Assessment of the implications for the economic evaluation of applying the clinical evaluation (Step 1 then Step 2).....	111
Table D5.6 Assessment of the implications for the economic evaluation of extrapolating and transforming the clinical evaluation (Step 3)	112
Table D6.1 Analyses of the implications for the economic evaluation of usage beyond the requested populations and circumstances of use	114
Figure E1 Key information requests for assessment report Section E of a standard assessment for MSAC	116
Table E.1.1 Categories of data sources	119
Table A1.1 Examples of factors that are more readily quantified	137
Table A1.2 Examples of factors that are less readily quantified	138
Table A2.1 Methods to collect and collate expert opinion.....	142
Flowchart A7.1 Classification of integrated and streamlined codependent submissions	174
Figure A7.2 Preferred structure of an integrated codependent submission for Section 1.....	177
Figure A7.3 Preferred structure of an integrated codependent submission for Section 2.....	178
Figure A8.1 The two types of randomised trials used to evaluate screening investigative medical services.....	213

Contacts

The Medical Services Advisory Committee (MSAC) and its two sub-committees have secretariats within the Australian Government Department of Health.

Departmental Staff are available through the Health Technology Assessment (HTA) Team on the contact numbers and email below to discuss proposals for MSAC consideration or related matters. Any correspondence or assessment reports should also be lodged at via the address below. Staff within the HTA Team are also the first point of contact concerning the relevant committee or sub-committee's discussions and decisions.

HTA Team
Australian Government Department of Health
MDP 851
GPO Box 9848
CANBERRA ACT 2601

Or courier

Sirius Building
23 Furzer Street
PHILLIP ACT 2602

Phone: +61 2 6289 7550
Fax: +61 2 6289 5540
Website: <http://www.msac.gov.au>
Email: hta@heath.gov.au

Record of updates

Date	Version	Summary of changes
March 2012	1.0	Draft Guidelines prepared by the MSAC Guidelines Working Group with technical and editing assistance from Biotext
January 2013	1.1	Draft Guidelines – amendments
February 2014	1.2	Draft Guidelines – amendments
March 2015	1.3	Draft Guidelines – amendments
August 2015	1.4	Draft Guidelines – amendments from public consultation
February 2016	1.5	Final Guidelines – amendments from internal departmental consultation as well as following consultation with members of MSAC and ESC
February 2016	1.6	Accessibility Standards applied
March 2016	2.0	Document renamed and released for publishing Online
December 2016	3.0	Appendix 7 revised
28 July 2017	3.0	Heading Section B – B1 Direct Evidence formatted correctly for accessibility

Abbreviations

ACTR	Australian Clinical Trials Registry
AHMAC	Australian Health Ministers' Advisory Council
AR-DRG	Australian Refined Diagnostic Related Group
ARTG	Australian Register of Therapeutic Goods
CEA	Cost-effectiveness analysis
CI	Confidence intervals
CMA	Cost-minimisation analysis
CUA	Cost-utility analysis
CV	Contingent valuation
DOR	diagnostic odds ratio
DRG	diagnostic-related group
EAV	Effective analytical validity
ESC	Evaluation Sub-committee
Guidelines	<i>Guidelines for Preparing Investigative Assessment Reports for the Medical Services Advisory Committee</i>
Health	Australian Government Department of Health
HTA	health technology assessment
ID	Identification
ITT	Intention-to-treat
MAUI	Multi- attribute utility instrument
MBS	Medicare Benefits Schedule
MCID	Minimal clinical important difference
MSAC	Medical Services Advisory Committee
NHMRC	National Health and Medical Research Council
PASC	PICO Advisory Sub-committee
PBAC	Pharmaceutical Benefits Advisory Committee
PBAC Guidelines	<i>Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee</i>
PBS	Pharmaceutical Benefits Scheme
PLAC	Prostheses List Advisory Committee
PSD	public summary document
ROC	receiver operator characteristic
SG	Standard gamble
SROC	Standard receiver operator characteristic
STARD	Standards for Reporting of Diagnostic Accuracy
TGA	Therapeutic Goods Administration
Therapeutic Guidelines	<i>Guidelines for Preparing Therapeutic Assessment Reports for the Medical Services Advisory Committee</i>
TTO	Time trade-off
QALYs	Quality adjusted life years
Q-Twist	Quality- adjusted time without symptoms of the disease or toxicity

QUADAS
WTP

Quality Assessment of Diagnostic Accuracy Studies
Willingness-to-pay

PART I

General information

1 Medical Services Advisory Committee

1.1 Purpose and roles of MSAC

The Medical Services Advisory Committee (MSAC) is a non-statutory committee established by the Australian Government Minister for Health in 1998. MSAC appraises new medical services proposed for public funding, and provides advice to Government about the level and quality of evidence relating to the comparative safety, clinical effectiveness, and cost-effectiveness of such services. Amendments and reviews of existing services funded by the Medicare Benefit Schedule (MBS) or other programs (for example, blood products or screening programs) are also considered by MSAC.

The MSAC advises the Minister for Health on medical services in relation to:

- the strength of evidence about the comparative safety, effectiveness, cost-effectiveness and total cost of the medical service;
- whether public funding should be supported for the medical service 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; and
- other matters related to the public funding of health services referred by the Minister for Health.

MSAC also advises the Australian Health Ministers' Advisory Council (AHMAC) on health technology assessments referred under AHMAC arrangements.

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

1.2 Membership of MSAC

MSAC is an independent expert committee comprising professionals 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 <http://www.msac.gov.au>.

1.3 MSAC sub-committees

MSAC currently has two sub-committees: 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 also the Deputy Chair of MSAC) to manage MSAC activities between formal committee meetings.

1.4 Overview of MSAC processes

1.4.1 Regulatory framework

All therapeutic goods used in the provision of medical services must be assessed by the Therapeutic Goods Administration (TGA) and included on the Australian Register of Therapeutic Goods (ARTG) before they can be marketed in Australia.

As a general rule, MSAC does not support public funding for a service that uses a therapeutic good for indications beyond those for which it was included on the ARTG.

An application to MSAC can be lodged before relevant therapeutic goods are included on the ARTG provided that the applicant has evidence that the relevant sponsor has commenced the TGA process. Confirmation of inclusion on the ARTG is required before MSAC can finalise its own appraisal of the corresponding medical service.

In considering whether to advise listing a service on the MBS, MSAC considers whether the service meets the criteria laid down in the *Health Insurance Act 1973*, and takes advice from the Department of Health on legal and policy matters as required.

1.4.2 The application and assessment process

The approach to seeking MSAC advice to government for public funding is broken up into stages that provide stakeholders and the general public with opportunities to be actively engaged in the consultation phases, as well as opportunities for further applicant engagement throughout the process.

Unlike applications to the Pharmaceutical Benefits Advisory Committee (PBAC) and the Prostheses List Advisory Committee (PLAC), the costs incurred in submitting an application to MSAC is not recovered from the applicant. To ensure that only relevant information is collected, the scope of every application is determined before evidence is compiled.

MSAC may seek co-applicants or co-sponsors to broaden the scope of an application. In some instances, a professional body and more than one commercial company might be co-applicants in a combined application.

1.4.3 Sources of advice

In formulating its advice, MSAC and its sub-committees may seek expert opinion from relevant professional bodies or appropriate specialists, and input from appropriate consumer bodies. Where external advice is obtained, the applicant is informed of the advice and given an opportunity to reply.

1.4.4 Publication of assessment report

Reports will be published as submitted to ESC. Any agreed errors of fact will be provided separately as an errata. If commercial-in-confidence information has been identified by the Applicant in the Assessment Report and where agreed confidentiality has been reached, applicants will have access to and will be required to confirm that the modified version of the assessment report (submitted or contracted), and the modified MSAC Public Summary Document (PSD) can be uploaded on the MSAC website.

1.4.5 Timing of MSAC processes

MSAC advises all interested stakeholders of the meeting dates for the following year, as well as the associated cut-off dates via the MSAC website.

Assessment reports should be presented on time, complete, in the format requested in the associated template and with the correct number of copies. No guarantee can be given that material supplied late will be incorporated into the Assessment Report or included in the agenda papers to MSAC and its sub-committees.

For PBAC co-dependent integrated reports (material being presented to both PBAC and MSAC), the PBAC requirement for report formatting and publication will prevail in acknowledgement of the different government public funding arrangements, Pharmaceutical Benefit Scheme (PBS) listing and established memorandum of understanding arrangements.

Initial advice of committee decisions for co-dependent applications to MSAC and PBAC will, where possible, follow the PBAC approach to provision of advice to applicants.

1.4.6 MSAC appraisal

MSAC will appraise the evidence presented in the Assessment Report and ESC reports to inform its advice to Government. MSAC prepares a detailed rationale for its conclusions in the form of a PSD.

Where specific material is agreed to be confidential, the PSD will be published with the confidential material redacted. The Department offers debrief meetings to applicants following the public release of MSAC's advice, if requested.

Following MSAC's consideration, the Department of Health is required to consider the financial impact to Government, consult with relevant stakeholders, seek Cabinet agreement and draft and implement legislative change to amend or add an item to the MBS. As previously advised in Section 1.1, there is no obligation on Government to accept or implement the advice MSAC provides.

Please note that ESC and MSAC does not meet with or accept face-to-face presentations of evidence from applicants.

2 Introduction to the Guidelines

These *Guidelines for Preparing Investigative Assessment Reports for the Medical Services Advisory Committee* (referred to in this document as the Guidelines), provide practical information on how to present evidence to MSAC when seeking Australian Government funding of a medical service.

Although these Guidelines have been written for applicants from the medical profession and industry, they are also intended to provide information to other interested stakeholders, including clinical and patient groups, and the general community.

2.1 Structure of these Guidelines

These Guidelines are organised into four parts, as follows:

- **Part I General information**
This part covers information on the preferred layout and style conventions, different types of applications and a checklist with a navigation aid of the information that is to be contained in reports for particular types of assessment reports.
- **Part II Clinical and economic evidence provided in the most preferred format**
This part covers the evidence for public funding for the proposed medical service when it is available in the most preferred format. The Sections in Part II follow the order in which information should be presented in the assessment report:
 - A Context (details of the proposed medical service and its intended use on the MBS)
 - B Clinical evaluation
 - C Translation issues
 - D Economic evaluation
 - E Financial implications
 - F Other
- **Part III Clinical and economic evidence provided in alternative formats**
This part covers situations where the evidence is not available in the most preferred format.
- **Appendices** include additional information on various aspects of the assessment report.

Further information is available in the associated template for the investigative assessment report.

2.2 Associated documents

A template for the investigative assessment report is available on the MSAC website and should be used when developing an investigative assessment report in line with these Guidelines.

Applicants may also need to refer to the *Guidelines for Preparing Therapeutic Assessment Reports for the Medical Services Advisory Committee* (referred to in this document as the *Therapeutic Guidelines*).

2.3 What is an investigative medical service?

An investigative medical service is one that generates clinically relevant information about the individual to whom the service is rendered. There are a number of sub-categories of such services depending on the purpose of performing the investigative medical service. The purpose of performing an investigative medical service may be to:

- establish (or add to an existing) diagnosis in a patient presenting with a particular set of clinical symptoms (and/or set of results from prior tests);
- establish a predisposition or estimate a prognosis;
- identify a patient as suitable for a therapeutic medical service by predicting a variation in the effect of the therapeutic medical service;
- measure an early treatment effect on a surrogate outcome as the basis for predicting the extent of a later treatment effect on more patient relevant outcomes;
- monitor a patient over time after an initial investigation to guide subsequent treatment decisions if the service requires to be repeated; or
- form the basis of a screening program in asymptomatic populations (see Appendix 8).

To achieve an improvement in health outcomes, the investigative information must result in a change in the management of an intermediate therapeutic service. In this sense, it can only **indirectly** improve health outcomes and any improvement also needs to be balanced against any harm that the service might cause. This purpose defines the need for a co-dependent technology of some sort to be (or to have been) assessed.

A co-dependency occurs where the use of one health technology to directly improve health outcomes (e.g. a medicine, or medical device or procedure) is improved by the use of another health technology (e.g. a pathology or an imaging technology) and where both technologies require consideration for public funding. Possible co-dependencies involving investigative medical services include:

- investigative medical service + therapeutic medical service (both requiring funding approval through MSAC);
- investigative medical service (funding approval through MSAC) + therapeutic; or
- medical service (requiring funding approval through another committee e.g. co-dependent pharmaceutical that requires coordinated consideration for PBS funding by PBAC).

Further information on co-dependencies is available in Sub-section 4.3.3 of these Guidelines.

2.4 Writing and style conventions used in these Guidelines

Several conventions have informed the revision of these Guidelines to assist users of the document to navigate their way to the information required when preparing their assessment reports.

These Guidelines include a series of requests for specific types of information. The aim is to provide an ordered series of reference points (requests for information) against which the specific information presented in an assessment report can be evaluated to ensure that the assessment report is complete.

The ‘default’ writing style for requests for information uses the imperative voice, as follows:

‘Describe the proposed course of treatment’ or ‘Justify the exclusion of the study’.

Readers should interpret these imperative statements as indicating what **should** be done. This allows requests for information that is known to be more persuasive or influential to be communicated as simply as possible in these Guidelines. Following these requests helps to improve the comparability of assessment reports considered by MSAC, and hence the consistency of decision making.

Within each Section, the main requests for information expected to be addressed by each standard assessment report are highlighted as ‘**Information requests**’. Other subsidiary requests and background information are provided in normal text.

In some instances, the request includes the word ‘**must**’. In each case, the requirement is included in the information request under the separate heading of ‘**Information requirements**’. Failure to comply with these requirements is sufficient to render the assessment report unacceptable, and for the assessment report to be returned to the applicant.

In other instances, there is no basis to indicate a preference for one type of information over another. In these instances, options about what **could** be presented are usually given. MSAC is generally indifferent about which option is presented, although the context of a particular assessment report might suggest the basis for expressing a preference. The assessment report should therefore explain the basis for selecting the information presented.

2.5 The future

Future revisions of these Guidelines will be disseminated via the MSAC website. A summary of each change will be recorded at the front of the electronic version published on the website, and those involved in preparing assessment reports will be notified.

Further feedback on these Guidelines is welcome and should be forwarded to The HTA Team at hta@health.gov.au.

3 Rationale and basis for the economic evaluation in the Australian context

KEY POINTS — APPROACH TO ECONOMIC EVALUATION

- MSAC is required to assess the degree to which new, amended or revised medical services represent ‘value for money’ for the Australian community.
- The economic evaluation should focus on the effectiveness of the proposed medical service compared with existing medical services, its cost and the likely changes in the provision of health care resources after its introduction (including changes in the provision of other health care resources not funded through the MBS).
- Economic evaluations should be relevant to the Australian context.
- The practical aspects of the economic evaluation of the performance of medical services are challenging; therefore, there will be continued flexibility in the interpretation of these Guidelines.

Australia, like other countries, is faced with a steady increase in the total cost of medical services. Although the medical service budget is not ‘capped’ in Australia, choices must be made as to which medical service will be subsidised by the Australian Government. Economic evaluation is one factor to be considered when making choices among competing medical services. Other important factors that are considered include uncertainty, equity, extent of use and total costs.

3.1 Analysis of cost-effectiveness

MSAC considers the results of economic analyses in its decision making to assess the degree to which new or revised medical services represent ‘value for money’ for the Australian community.

3.2 Australian context

Although the results of clinical trials or studies of sufficient scientific rigour done overseas are a reasonable basis for economic evaluations relevant to the Australian health care system, an economic evaluation performed overseas will often not be relevant in Australia. This is because of standard differences in unit costs, the patterns of resource provision and the way in which health care is funded in other countries. Applicants are therefore encouraged to submit an economic evaluation that is relevant to the Australian context in Australian dollars.

3.3 Relevant factors influencing MSAC decision making

MSAC considers many factors when proposing that a medical service be publicly funded. Each of these factors might have a separate influence on the decision to list the proposed medical service and, depending on the circumstances of each consideration, might influence MSAC in favour of, or against, listing. More than one factor might be relevant to each consideration.

Tables A1.1 and A1.2 in Appendix 1 list relevant factors, which are divided into two groups: quantitative and qualitative. The qualitative factors (Table A1.2) include some of the underlying assumptions implicit in such concepts as quality-adjusted life-years (QALYs) and discounting. To enable consistency within an assessment report regarding these factors, MSAC has adopted a particular position (which is specified in these Guidelines in the Sections indicated by the cross-references in the tables). However, in certain circumstances, it might be reasonable to argue that a different position should be considered.

Individual factors are not weighted equally by MSAC in its decision-making process, and different factors might be more or less important in different situations. In other words, the importance of any particular factor cannot be quantified. The descriptions provided in Appendix 1 represent MSAC's understanding at the present time. MSAC continues to reflect on its processes and further develop its understanding of these matters.

3.4 Flexibility in interpretation of these Guidelines

Despite the differences in data available and uncertainties that might exist in the base case, it is in the interests of the community, industry and MSAC that uniformity be maintained in the way that economic analyses are conducted and evaluated. However, the practical aspects of the economic evaluation of the performance of medical services are challenging for applicants, MSAC and the administrative arm of government. For this reason, although applicants should present the economic analysis as outlined in these Guidelines, there will continue to be the need to be flexible in the interpretation of these Guidelines.

4 Organisation of a standard assessment report

KEY POINTS — ORGANISATION OF A STANDARD ASSESSMENT REPORT

- Assessment reports must consist of an executive summary, the main text of the assessment report and additional information (attachments and technical documents).
- Part II (for the majority of assessment reports) and Part III (for supplementary and alternative information in some assessment reports) of these Guidelines provide the preferred order for presenting information in the main text of standard assessment reports.
- The preferred order for presenting information consists of six Sections (A–F). If possible, do not present information in any other order, because this will reduce MSAC’s ability to effectively evaluate the assessment reports.
- Use frequent, accurate cross-referencing between the executive summary, main text and other technical documents.
- Use succinct, plain English wherever possible (while maintaining scientific rigour).
- Provide justification for any variations to the requested information.
- If using a new analytical technique, present the base case using both the requested methods and the new technique for comparison.

These Guidelines are designed to assist applicants to identify and present the basic information required by MSAC to determine its advice and to provide guidance to applicants on the most appropriate form of economic evaluation for the specific assessment reports.

This Section outlines the information that should be presented in a standard assessment report. A flowchart showing MSAC’s key decisions in evaluating standard assessment reports is also included, along with advice on presenting alternative information in particular circumstances.

4.1 Choice of information

Information should address the PICO Confirmation agreed by the PASC or the MSAC Executive, however, an applicant, if they choose to take the risk of not to adhere to the PICO Confirmation should note that this may impact the final decision made by MSAC. These Guidelines set out the information requested, and while additional information might be included, it must be clear that this additional information addresses matters that are outside the Final PICO Confirmation.

A wide array of information should be presented in a standard assessment report to MSAC. Some information is requested for all assessment reports, whereas some additional information requests only apply according to the type of service for which funding is being sought. In addition, a large number of information requests provide guidance on presenting the ‘next best’ option when it is not possible to provide the preferred information.

Each assessment report should be as succinct and informative as possible. MSAC and ESC are most likely to be influenced by arguments based on scientifically rigorous data rather than opinions. Assessment reports should use suitable scientific language, but avoid jargon.

4.2 Overview of a standard assessment report

4.2.1 Sections of a standard assessment report

To achieve a ‘base case’ estimate and decision analysis with uncertainty identified, a standard assessment report needs to include an executive summary and Sections A–F as shown in Figure 4.1.

These Guidelines are set out to provide a stepped approach for the presentation of the best and most persuasive evidence. The most preferred option is described in Part II and alternative, less preferred options are provided in Part III.

The order of the information requests indicates the preferred flow of information. The requests in Parts II and III refer to all medical services and products.

4.3 Presentation of the assessment report

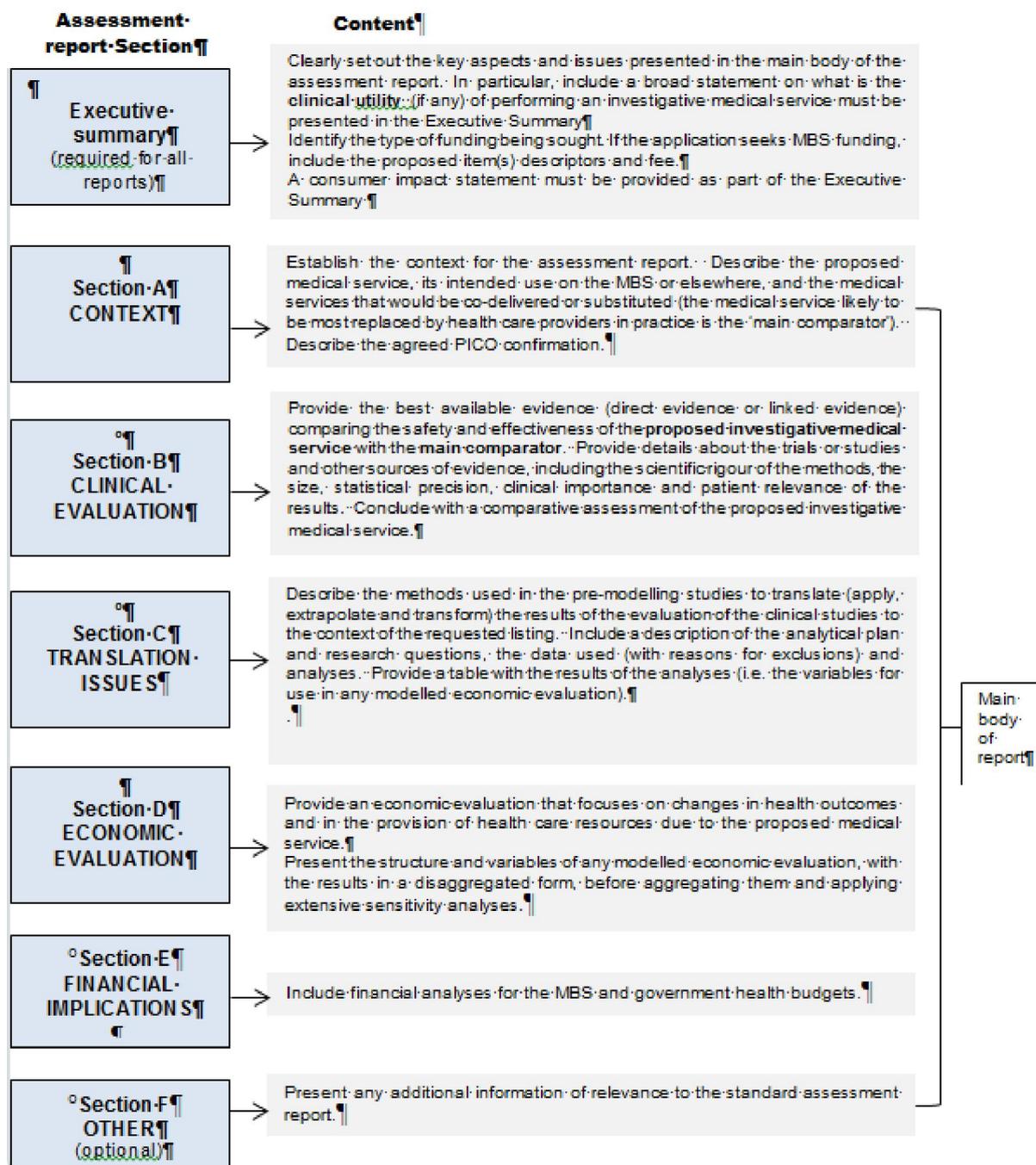
4.3.1 Standard assessment report

The main body of the assessment reports **must** be presented according to the MSAC *Investigative Assessment Report Template* available on the MSAC website. Key reports of the relevant trials on which the report is based must be provided separately. Other information might be provided as attachments or technical documents. This other supplementary material is made available to committee members on request. Where the report relies on specific information, it should be referenced (if publicly available) or included within the report and available for publication. Where the conclusions in a report rely on agreed commercial-in-confidence material it might, by agreement, be provided as a not-for-public-release attachment.

It is vital that the assessment report includes frequent and accurate cross-references between the executive summary and the main body of the assessment report, and between the main body of the assessment report and reports of the key trials, attachments, technical documents and material in electronic formats. This will assist those who have to evaluate and consider the assessment report.

The key steps for preparation of a standard assessment report and how these decisions relate to the Sections of the assessment report, are set out in Parts II and III of these Guidelines. The order of the information requests in Part II and/or Part III indicates the preferred order for the information that should be presented to optimise its evaluation by MSAC. Arranging the same information in another order has generally been found to be unhelpful.

Figure 4.1 Sections of a standard assessment report



Note: PICO = population/problem; intervention (medical service), investigation (diagnosis); comparator; outcomes

4.3.2 Two stage approach to an assessment report

Some applicants may advise their preference for submitting the clinical component (Sections A and B) of their submission based assessment before commencing their economic component (Sections C to F). This would provide the benefit of the ESC and MSAC's feedback on both the clinical evidence and the proposed structure of the economic model.

In commencing this pathway, applicants should be aware that it will require a two stage approach to the submission, ie. it will be considered by ESC and MSAC twice. The first stage would be presenting Sections A and B, and then the second stage, presenting Sections C to F.

If an applicant chooses this pathway, consideration should be given to submitting the draft financial component (Section E), alongside that of Sections A and B.

4.3.3 Co-dependent/integrated assessment reports

For co-dependent applications, applicants are not required to submit all of the therapeutic evidence of health outcomes in the investigative (MSAC) application because this evidence is presented in full in the therapeutic (either PBAC or MSAC) application.

In this case, the investigative application is only required to include answers to the information requests listed in Appendix 7.

5 Lodging an assessment report

5.1 Assessment report checklist

5.1.1 Information requirements

As indicated in the template, the investigative assessment report must consist of the following components:

- contents;
- executive summary (including a consumer impact statement);
- main body of report;
- attachments; and
- appendices.

Each hard copy of the main body of a standard assessment report must be suitably bound, identified, indexed, pages numbered and divided with labelled tabs.

All economic calculations must be provided in Australian dollars.

The investigator's summary of each trial report, the main published paper, and an adequate account of the methods and results for each trial or study must be included as attachments within the main body of the assessment report.

All submitted information must be legible and in English.

All assessment reports, unless otherwise specified, will be made public in the format in which they were lodged.

5.1.2 Information requests

Before lodging an assessment report, the applicant must notify the Health Technology Assessment Team (hta@health.gov.au or (02) 6289 7550) of their intention to lodge an assessment report using the timeframe published on the MSAC website and to also receive up-to-date details on the information requirements (i.e. the number of hard copies required etc).

A checklist is provided at Table 5.1 as an initial guide to assist applicants in this process.

Table 5.1 Checklist of information to be included in an investigative assessment report

Component	Included?
<ul style="list-style-type: none"> The original, signed covering letter for the assessment report (with an attachment containing the complete index to the assessment report). 	Yes/No
<ul style="list-style-type: none"> A comprehensive index attached to the covering letter, which serves as a checklist for all documentation and other materials comprising the assessment report and confirming: <ul style="list-style-type: none"> – the numbers of copies of the main body of the assessment report and details of its contents – the numbers of copies of other parts of the assessment reports and details of their contents 	Yes/No
<ul style="list-style-type: none"> The current TGA-approved product information with approval date (if and when available, with the latest draft product information in the meantime; each copied single-sided and stapled) (where relevant). 	Yes/No
<ul style="list-style-type: none"> The letter of registration with details of marketing approval and registration (if and when available; each copied single-sided and stapled) (where relevant). 	Yes/No
<ul style="list-style-type: none"> Any additional technical documents, attachments and references provided separately to the main body of the assessment report (where relevant), which should: <ul style="list-style-type: none"> – be suitably bound (i.e. each folder is robust enough to withstand regular use, with the width of its spine matching the number of pages it contains) – have the contents identified on the cover – be legible and in English (or accompanied by a reputable translation) 	Yes/No
<ul style="list-style-type: none"> Bound copies of the main body of the assessment report (using the agreed template), which must: <ul style="list-style-type: none"> – be suitably bound (i.e. each folder is robust enough to withstand regular use, with the width of its spine matching the number of pages it contains) – have the contents identified on the cover – have a clear and adequate index (which encompasses both the main body of the assessment report and the contents of all other documentation contained in separate volumes, and also identifies all other materials supplied as part of the assessment report, which is also attached to the covering letter of the assessment report) – have consistent pagination throughout – include dividers between each Section, attachments and references, with an appropriately labelled tab extending beyond the page width – have all cost calculations in Australian dollars (A\$) – incorporate attachments containing reports of each of the relevant randomised trials (or each of the relevant non-randomised studies, if necessary), which must be: <ul style="list-style-type: none"> (i) the investigator’s summary of each applicant’s trial report and the main published paper (where available), together with adequate details of the trial methods, analysis and all trial results presented in the assessment report for use in the economic evaluation; OR the main published paper alone if the applicant has no access to a more detailed report (ii) legible and in English (or be accompanied by a reputable translation) 	Yes/No
Electronic versions of the assessment report on a USB	
<ul style="list-style-type: none"> Supply the whole assessment report and any accompanying calculations and models in electronic format (with any spreadsheet compatible with Microsoft Excel 2010-2013, RevMan, any word-processing document compatible with Word 2010-2013, and any other software package consistent with Sub-section 5.2.1). Ensure that all components of these electronic documents, spreadsheets and analyses are fully accessible (e.g. do not have password protection); fully enabled to allow all document text, tables and figures to be accessed for copying; and fully executable to allow all spreadsheet cells and all statistical or decision analysis input variables to be changed. 	Yes/No
<ul style="list-style-type: none"> Supply electronic copy of key articles that the conclusions in the report are based on. 	Yes/No

5.2 Provision of information to allow independent verification of computer analyses

5.2.1 Information requirements

Provide sufficient information to permit independent verification of computer-based analyses to generate information for the assessment report Sections C or D (e.g. input data, methods of analysis, outputs).

Provide an electronic copy of all computer-based analyses (including the economic evaluation) in the form in which it was conducted, together with any associated data files, and a technical document or an attachment with clear cross-references to the assessment report.

Use a software package that can be readily evaluated by MSAC or, before lodging the assessment report, discuss the arrangements with the HTA Team to ensure the acceptability for the evaluation of any software that is not on the maintained list of software packages. Examples of software include Word (2010 - 2013 compatible), Excel (2010 – 2013 compatible), STATA, Triage, RevMan, Endnote, etc.

5.3 Provision of information after lodgement of the assessment report

5.3.1 Post-lodgement communication with MSAC

MSAC procedures provide post-lodgement opportunities for applicants to communicate with MSAC.

It is expected that applicant responses will address issues raised in the relevant papers rather than introduce substantive changes, such as a different population identified by a modification to the requested restriction, a different nomination for the main comparator, new data or new analyses. Such changes might result in an MSAC request for a standard reassessment to examine the implications of the substantive change.

Before the departmental papers are finalised, applicants might be approached by either the Department or an assessment group for further information or clarification of aspects of their assessment report. Applicants are expected to deal with these requests expeditiously.

5.3.2 Provision of information sourced from the TGA after lodgement of the assessment report

Upon receipt of notification of TGA registration approval, applicants are requested to advise the HTA team (via hta@health.gov.au) immediately, in writing, of any aspect of an assessment report that is not consistent with the final TGA registration. At this time, also provide a copy of the TGA-approved product information, accompanied by a document highlighting any variation between the most recent draft provided with the assessment report and the subsequent TGA-approved product information that would have any bearing on the consideration of the assessment report or on the consideration of any subsequent MSAC recommendation to list.

Part II

Preferred clinical and economic evidence for proposed medical services to be considered by MSAC

Section A

Details of the proposed investigative medical service and its intended use on the Medical Benefits Schedule (or for other public funding)

Introduction

Section A of the assessment report establishes the context for the report. It provides the information outlined in the PICO Confirmation that has been agreed to by PASC or the MSAC Executive in the pre-assessment phase of the application to MSAC.

A1 Address all items in the agreed PICO Confirmation

- All items in the agreed PICO Confirmation should be addressed in the assessment report.
- If any items are not addressed this presents a risk to the applicant; these items should be identified and reasons provided for not addressing them.
- Confirm that the assessment report has fully addressed the questions defined in the agreed PICO Confirmation.
- Indicate if any additional information provided in the assessment report has been compared to the agreed PICO Confirmation.

A2 Proposed medical service

Describe the proposed medical service as set out in the agreed PICO Confirmation, including, the purpose of the investigative medical service, methods used (eg point of care vs laboratory), mode of delivery and other details. If the proposed medical service is a genetic test, broadly outlined the sampling procedures, the range of testing which needs to be done (i.e. is the test limited to monogenic testing including single mutation testing or is it a panel test) and laboratory methods used in both processing sample and validating the result.

It should be clearly articulated whether the purpose or intent of performing the investigative medical service will:

- establish (or add to an existing) diagnosis in a patient presenting with a particular set of clinical symptoms (and/or set of results from prior tests);
- establish a predisposition or estimate a prognosis;
- identify a patient as suitable for a therapeutic medical service by predicting a variation in the effect of the therapeutic medical service;
- measure an early treatment effect on a surrogate outcome as the basis for predicting the extent of a later treatment effect on more patient relevant outcome;

- monitor a patient over time after an initial investigation to guide subsequent treatment decisions if service requires to be repeated; or
- form the basis of a screening program in asymptomatic populations (see Appendix 8).

A3 Proposed MBS listing or other public funding sought

Provide MBS or other public funding descriptors, as set out in the agreed PICO Confirmation. Differences between the proposed descriptor and the descriptor provided in the agreed PICO Confirmation should be highlighted and a justification provided in the assessment report.

A4 Proposed population

- Identify the main population(s) described in the agreed PICO Confirmation including key inclusion and exclusion criteria. This may involve the results of prior tests to exclude or include patients in the proposed population.
- If more than one population is being proposed, for example, in genetic testing more than one population may or may not be considered (the proband population and/or first degree relatives of the proband).
- Include a high level summary of the frequency (prevalence and/or incidence) of the population or disease in question and where relevant the natural history/pathophysiology of the condition of interest

A5 Comparator details

- Identify the main comparator(s) described in the agreed PICO Confirmation.
- Depending upon the population being studied, the primary comparison is likely to be either another investigative medical service in terms of alternate diagnostic method or modality or in some instances ‘no testing’/‘usual care’, for example, a clinical diagnosis in the proposed population based on the symptoms and signs of presenting complaint, previous medical or family history and the outcome of other tests (as appropriate).
- If there are any additional comparator(s) justify their selection.
- Identify any other factors that might affect the identification of the main comparator in the future.

A6 Clinical management algorithm(s)

- Present the clinical management algorithm(s) described in the agreed PICO Confirmation.
- Present the clinical management algorithm that depicts the context of the intended use of the proposed medical service following a listing on the MBS or other public funding, as listed in the agreed PICO Confirmation.
- Present the corresponding algorithm depicting the current context as listed in the agreed PICO Confirmation.

- Highlight the differences between the two algorithms to summarise the changes in the patterns of resource provision, both those required by any requested indication and those that would be expected to follow as consequences of the requested listing.
- Indicate whether multiple-listing scenarios are presented.

A6.1 Algorithms for intended and current contexts

Clinical management algorithms are most relevant to an assessment report presenting a modelled economic evaluation (see Sub-section D1). They are also helpful for estimating changes in use and cost of other medical services (see Sub-section E3). An assessment report not presenting a modelled economic evaluation might only need to present straightforward algorithms.

The objective of these clinical management algorithms is to help clarify the comparison addressed in the assessment report through the following three steps:

- Define the eligible patients and the circumstances of use if the listing or public funding were implemented as requested (algorithm 1).
- Identify the current situation in terms of the expected substitution of service options for these patients and their circumstances of use, both at the time of substitution and subsequently (algorithm 2).
- Identify the full nature of the comparison(s) being made in the assessment report and limit the comparison to these contexts (highlight the differences between algorithms 1 and 2).

The algorithms are expected to be of varying complexity, depending on the particular contexts to be described in each assessment report. Overall, ensure that the algorithms identify the nature of any and all differences across the full streams of resource provision consequences, both before and after the point(s) in the algorithm at which the proposed medical service is introduced. This ensures greater clarity about the context of the intended use of the proposed medical service in terms of patients and circumstances, from which the comparative health outcomes, comparative costs, comparative cost-effectiveness and financial implications can all be estimated.

In each algorithm, summarise all:

- relevant diagnostic and treatment steps, including all:
 - required previous medical services; and
 - diagnostic criteria and/or tests (including those demonstrating that one or more previous medical services cannot be used to manage the indication, and including those required to support any continuation criteria in the requested restriction);
- required co-delivered services; and
- consequences for subsequent service options.

Specify any other important characteristics of patients and types of circumstances of use. Examples include specifying the characteristics of the medical condition in the eligible patients (e.g. in terms of risk factors) and the aspects of the spectrum of the medical condition (e.g. in terms of severity of disease or remaining treatment options). Sub-section D2 provides further examples.

Justify the basis for the selection of the algorithm with reference to a literature review of relevant published clinical management guidelines. Provide a copy of those clinical management guidelines in an attachment or technical document. If expert opinion or survey has been used to help specify the clinical management algorithms, please advise.

A7 Differences between the proposed medical service and the main comparator

Describe the main differences in the indications, contraindications, likelihood and severity of adverse events between the proposed medical service and the main comparator(s).

A8 Clinical claim

Provide information about the clinical claim with respect to the proposed investigative medical service, as set out in the PICO confirmation, against the main comparator. The clinical claim is to be tested in Section B.

A9 Summarise the primary elements of the decision analysis (PICO)

Provide the PPICO (population/problem, prior tests, investigation/index test, comparator, and outcome) criteria and decision option(s) for the proposed investigative medical service, as set out in the PICO Confirmation.

Primary elements for an investigative medical service include:

- population and medical condition (e.g. patients with non-small cell lung cancer) including results of prior tests (i.e. any tests that would be done before the proposed investigative medical service is used) that would inform which patients are included in or excluded from the proposed population;
- proposed investigative service
- comparator service; and
- outcome claims.

Section B

Clinical evaluation for the proposed investigative medical service

Introduction

The purpose of Section B is to identify and present the best available clinical evidence for the main indication. As alluded to in the introduction of this document, the information generated by the investigative medical service under consideration has the potential to change clinical management and thereby improve health outcomes indirectly through the use of a therapeutic medical service. It is this characteristic that generates the essential difference for the presentation of evidence compared to the presentation of evidence for therapeutic medical services. That said MSAC still has a preference for clinical and economic evaluations of investigative medical services that are based on direct evidence on health outcomes if this has been conducted. Direct evidence is considered to be a more convincing source of evidence. If direct evidence is available, it is proposed that it be presented as a priority, but given the scarcity of such evidence, a linked evidence approach is required to evaluate the majority of investigative medical services presenting for consideration by MSAC.

Overall, the assessment of an investigative medical service has two stages:

- search for and assessment of (if it exists) direct clinical trial evidence on patient management and health outcomes. If there is evidence from clinical trials or studies that have been specifically designed to prove a linkage between the investigative medical service and a therapeutic outcome, applicants are redirected to the *Therapeutic Guidelines* (Section B), because the principles of presenting such evidence are the same as for therapeutic medical services — although the terminology relating to presenting the evidence might occasionally differ (e.g. ‘intention to test or screen’ versus ‘intention to treat’); and
- in the absence of direct clinical trial evidence, a linked evidence approach is recommended. Given the scarcity of direct clinical trial evidence or studies specifically designed to test the linkage between an investigative service and a therapeutic health outcome, applicants need to consider a linked evidence approach to evaluate potential health outcomes from use of the investigative medical service. This involves the narrative linking of evidence assessing components of the test-treatment pathway in order to come to a conclusion as to the impact on patient health outcomes as a result of performing the test. Along this test-treatment pathway the components that are linked up include diagnostic performance, clinical validity (where relevant), therapeutic efficacy (change in management) and therapeutic effectiveness (informed by direct evidence supporting the effectiveness of treatments that are selected as a consequences of performing the investigative medical service).

B1 Direct Evidence

B1.1 Description of search strategies (direct evidence)

INFORMATION REQUEST

- Describe the search strategies and characteristics used to locate reports of potentially relevant trials from the published literature and registers of clinical trials in relation to any direct clinical trial evidence of health outcomes.

Search strategies

The primary objective of the search strategies is to locate *all* direct clinical trial evidence of health outcomes that compare clinical management with the proposed investigative medical service, with the main comparator for participants with characteristics that overlap with patients who would be eligible for use of the proposed investigative medical service. The search should involve at least three approaches:

- (a) a search of the published literature;
- (b) a search of registers of clinical trials; and
- (c) manual checking of reference lists of all relevant articles that are obtained by other means.

When describing the search strategies and characteristics, sufficient detail should be provided so that an independent replication of the search would yield the same results.

The methods used to search the published literature are pivotal to assessing the completeness of the overall search. Therefore, specify the following characteristics of the search strategy:

- the specific databases and registers of clinical trials searched, including at least MEDLINE, EMBASE, the Cochrane Library (including the Cochrane Database of Systematic Reviews and the Cochrane Central Register of Controlled Trials), the ACTR; the search should also include databases internal to the company and any other known registers of randomised trials relevant to the therapeutic area;
- the date the search was conducted;
- the date span of the search (including the most recent update of each database searched);
- the complete search strategies used, including the search terms (key or MeSH words) and the relationship (sets and Boolean) between the search terms; and
- any supplementary searches, especially manual checking of references in the retrieved papers from the database searches.

If the literature review identifies clinical trials/ that constitutes direct evidence of the impact of the proposed investigative medical service on clinical management and health outcomes with characteristics that overlap with patients who would be eligible for use of the investigative medical service, go to Sub-section B1.2.

If there is no direct evidence, Section B2 provides the information request for presentation of linked evidence.

B1.2 Presentation of direct evidence

INFORMATION REQUESTS

- If there is direct clinical trial evidence of the effect of the proposed investigative medical service on clinical management and/or patient health outcomes, go to the *Therapeutic Guidelines* (Part II, Section B) and follow the information requests as for a therapeutic medical service.

Direct clinical trial evidence of therapeutic health outcomes refers to evidence from trials or studies specifically designed to prove the linkage between an investigative service and a therapeutic outcome (both safety and clinical effectiveness). This includes trials and studies that allocate (preferably randomise) patients to groups:

- with or without the investigative service and subsequent clinical management; and/or
- who have been identified for treatment by the results of the investigative service to groups to receive a particular treatment or alternate treatment strategy.

The principles for presenting direct clinical trial evidence of the effect of an investigative medical service on patient health outcomes are the same as for presenting clinical trial evidence relating to a therapeutic medical service. The information requested in the *Therapeutic Guidelines* (Part II, Section B and Part III, Section B) therefore also apply in this case.

For investigative medical services that have direct evidence of their effect on patient health outcomes, the following additional pieces of information needs to be considered in conjunction with the presentation of direct evidence to inform an overall conclusion as to whether the proposed service is superior or non-inferior to the main comparator (see Section B8). The following pieces of information may or may not have been reported as part of any direct clinical trials:

- The diagnostic performance and clinical validity of the investigative medical service where relevant (see Section B3 and B4).
- The relative clinical impact of false negatives and false positives arising from the test if this information cannot be extracted from direct evidence presented (see Section B5).
- Impact of repeat testing (if relevant) (see Section B6).
- The relative safety of performing the test (see Section B7).

B2 Overview of linked evidence approach

B2.1 Basis for linked evidence

INFORMATION REQUEST

- Provide information as to whether there is a basis to present a linked analysis.
- Provide a summary of the subsequent decision options in terms of therapeutic medical services that would arise from performing the investigative medical service.
- Determine whether there is evidence on health outcomes for each therapeutic medical service option and whether the evidence supporting these subsequent therapeutic options has been generated in similar populations.

For those investigative medical services for which no direct evidence of the effect of the service on clinical management and/or patient health outcomes exists, consider whether a linked analysis is feasible to determine the investigative medical service's impact on clinical management and health outcomes. In other words, can different types of evidence from different sources be linked in a chain of argument to estimate this impact? A full, linked evidence approach is only meaningful when the evidence for accuracy for an investigative medical service under consideration and the evidence supporting subsequent options in terms of therapeutic medical services have been generated in similar patient populations. Therefore, it is clinically sensible to link the two datasets.

When determining whether a linked analysis is justified, the first question to consider is whether separate evidence has been generated on health outcomes for each therapeutic medical service option arising from performing the investigative medical service. When evidence from accuracy studies is linked to evidence of treatment effectiveness, the second question to ask is whether the spectrum of disease in patients defined as having the target information or clinical information of interest using the investigative medical service is representative of those treated in studies evaluating the effectiveness of the therapeutic medical service. Comparing the characteristics of patients selected to accuracy studies with those selected to studies of standard treatment might identify factors that preclude the transferability of results between the two populations.

The treated population and treatment comparison (treatment and control arms) must be applicable to the intended use of the investigative medical service. The results of the investigative medical service can be used to start, stop or modify treatment. The link between the results of an investigative medical service and treatment decisions can be reasonably assumed when the proposed investigative medical service will be used to replace an existing investigative medical service and standard treatment for the target condition is well established. In other cases, studies evaluating the impact of the results of the investigative medical service on patient management might be required to demonstrate that the test results are sufficient to alter the clinician's threshold for changing clinical management.

The treated population and treatment comparison (treatment and control arms) is less likely to be applicable to the intended use of the proposed investigative medical service if a positive result using the proposed investigative medical service leads to earlier, new or alternative therapeutic medical services that have not been evaluated in clinical trials. In other words, if the investigative medical service identifies patients earlier or with a

different spectrum of disease from the patients in whom the therapeutic medical service has been trialled, then it is not clinically sensible to link this evidence. In such circumstances, direct evidence is needed.

If the proposed investigative medical service is intended to be a new screening or early detection tool, direct evidence of health outcomes comparing the effectiveness of earlier diagnosis and treatment versus later (standard) diagnosis and treatment is preferred. In the absence of direct evidence a linked analysis may be feasible if evidence exists demonstrating that the proposed test results in earlier diagnosis and separate evidence exists that earlier treatment results in better health outcomes. However it is generally regarded that direct evidence of the merits of a screening test is more persuasive evidence (refer to Appendix 8 for further details)

B2.2 Steps for linked analysis

The following pieces of information are generally required to construct a linked evidence analysis with variations of evidence requirements depending upon the purpose of performing the investigative medical service. The differing evidence requirements are explained in the subsequent sections:

- consideration of the diagnostic performance and clinical validity (where relevant) of the investigative medical service (Section B3 and B4);
- consideration of the clinical utility of the investigative medical service (Section B5) in terms of the impact of positive versus negative test results on patient management, the contribution and clinical importance of false negatives versus false positives and direct impact of each therapeutic medical service option on health outcomes;
- consideration of the impact of repeat testing (if relevant) (Section B6); and
- consideration of the relative safety of performing the investigative service both immediate safety issues of directly performing the test and ‘flow on’ safety issues that arise as a result of conducting the investigative service (Section B7).

Once a narrative linking of the above Sections has taken place, an overall conclusion is made in Section B8 as to whether the proposed investigative medical service is superior or non-inferior to the main comparator.

B3 Diagnostic performance

Introduction

It is important to ascertain how well an investigative medical service performs in terms of its ability to distinguish the presence or absence of the target disease or clinical information of interest. Assessing diagnostic performance depends upon both whether there is a reference standard and the nature of the clinical comparison.

A reference standard enables the determination of the presence or absence of the target condition or clinical information of interest. That is, it ‘truly’ distinguishes patients who have the target condition or clinical information of interest from those who do not. Therefore, the availability of a reference standard creates more certainty around the evidence presented. This allows quantitative assessment of *analytical validity*¹ (*accuracy* and *reliability*). This assessment informs a diagnostic conclusion of whether the proposed service is superior or non-inferior to the main comparator in terms of accuracy and reliability. The key measures of accuracy of interest to MSAC are **sensitivity** and **specificity**. It should be noted that if a genetic test is being presented to MSAC for consideration, a distinction is made between **analytical sensitivity and specificity** (which should be presented in Section B3) and **clinical sensitivity and specificity** (which should be presented in Section B4).

For investigative medical services for which there is no reference standard (e.g. studies simply looking at diagnostic yield of either the proposed service or its comparator) evidence of concordance needs to be presented alongside evidence of reproducibility. The absence of a reference standard creates more uncertainty around any claim of diagnostic performance, but concordance analyses remain useful in this circumstance. The clear preference for a reference standard does not imply that a minimum standard must be met. MSAC has considered and will continue to consider all types of evidence, whether or not there is a reference standard. However, MSAC will be most influenced by the results of studies for which there is the most rigorous source of data (refer to Sub-section B3.8 on how to conduct a concordance analysis).

The nature of the main comparison outlined in the PICO also informs whether it is feasible or not to generate a summary estimate of the comparative accuracy of investigative medical service versus the main comparator. If the main comparator is not a specific investigative method or modality per se (for example the comparator is ‘no testing’ or ‘usual care’ without the proposed service), ascertaining comparative/incremental accuracy may be difficult if the accuracy of the main comparator cannot be pinned down or is simply not known. This may be an issue in the example of genetic testing especially if a definitive diagnosis cannot be made other than through genetic testing.

¹ *Analytical validity* relates to whether an investigative medical service measures what it claims to measure and combines the concepts of accuracy and reliability. Accuracy (which is technically analogous to the concept of validity) refers to the amount of agreement between the results of the investigative medical service under consideration and those from the reference standard (i.e. the proportion of participants whom the investigative medical service correctly identifies as positive or negative. Reliability (which is analogous to the concept of precision) refers to the rate of agreement among different operators or instruments applying the same investigative medical service. A reliable investigative medical service consistently gives the same result. However, an investigative medical service might reliably provide an inaccurate result. In other words, reliability is necessary but not sufficient for analytical validity, and vice versa.

If assumptions are made about accuracy of the main comparator, then the rationale for these assumptions must be presented. If an ‘all or nothing’ assumption is made about the accuracy of the main comparator, for example, patients in the comparator arm of a ‘no testing’ scenario will be clinically managed as if they all (100%) test negative (if not offered the proposed investigative medical service) where in fact a numerical proportion of these patients (depending on how frequent the medical condition under consideration is) may in fact actually have the medical condition under consideration and thus will experience the clinical consequences of a false negative result (see Section B5 for further details). This will inform what numerical probabilities will feed into the main comparator arm the decision analytic model of the economic evaluation (see Section D for further details).

B3.1 Identification of a reference standard

INFORMATION REQUESTS

- Identify a reference standard for the proposed investigative medical service (or state if there is not a reference standard).

The reference standard² is an investigative medical service or series of investigative medical services that is 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. With dichotomous investigative medical services and a single-target condition, accuracy is often expressed as the proportion of people with the target condition who indeed have a positive result (sensitivity, or true positive fraction) and the proportion of people without the target condition who have a negative result (the specificity, or true negative fraction). Estimates of accuracy are based on the assumption that the investigative medical service under consideration is being compared to a reference standard that is theoretically 100% sensitive and specific. If there are any disagreements between the reference standard and the investigative medical service, then it is assumed that the investigative medical service is incorrect. Thus, the choice of an appropriate reference standard is a very important determinant in establishing the accuracy of an investigative medical service.

Briefly outline in this section of the assessment report the widely accepted reference standard in relation to the proposed population being assessed including a brief discussion as to the fallibility of the proposed reference standard.

If a reference standard does not exist, Section B3.8 explains that evidence of concordance needs to be presented. If a reference standard does not exist, usual measures of accuracy cannot be generated.

² If the reference standard, for the purpose of an assessment, is the same reference standard as reported in the primary accuracy studies under consideration, then this is sometimes referred to as the *evidentiary standard*.

B3.2 Description of search strategies (accuracy studies)

INFORMATION REQUESTS

- Describe the search strategies and characteristics used to locate reports of potentially relevant accuracy studies from the published literature and registers of clinical trials that (a) report accuracy studies, with preference for studies directly comparing the proposed (index) investigative medical service and main comparator service, followed by (b) indirect comparisons where there is a common reference standard, etc.
- If the main comparator is not a specific investigative method or modality per se (for example 'no testing' or usual care without the proposed service), ascertaining comparative accuracy may or may not be feasible if the accuracy of the comparator is not known. If assumptions are made about accuracy of the main comparator, then rationale for these assumptions must be presented and a detailed assessment of evidence of the incremental accuracy of the proposed service over the comparator as outlined in the remainder of Section B3 is not necessary. In this circumstance simply present 'single arm' evidence of accuracy of the proposed investigative service only against the reference standard. Provide justification why an estimate of comparative accuracy with the main comparator cannot be generated from the published literature.

In the presence of a reference standard, the accuracy of the proposed investigative medical service and the main comparator (provided a detailed assessment of comparative accuracy is feasible) can be compared in three different ways:

- **Fully paired direct comparison (the strongest design):** the proposed investigative medical service and the comparator medical service are evaluated in the same patient population and all study participants receive the proposed investigative medical service and the comparator medical service, as well as the reference standard. Fully paired comparisons are efficient, in terms of the resulting precision relative to the number of study participants.
- **Direct comparison without full pairing:** participants receive only a subset of the investigative medical service(s) under consideration. In this instance, study participants ideally should be randomly allocated to receive either the proposed investigative medical service or the comparator, and the results subsequently verified by the reference standard.
- **Indirect comparisons:** estimates of the accuracy of the respective investigative medical services are obtained in different patient groups. The accuracy of the proposed investigative medical service is estimated in one set of studies, while the accuracy of the comparator test is estimated in a different set of not or only partially overlapping studies. Such indirect comparisons can be prone to selection bias.

If a detailed assessment of comparative accuracy between the proposed investigative service and the main comparator is not feasible and/or assumptions around accuracy, for the purpose of the economic model (Section D), will be made in relation to the main comparator, then simply search for accuracy studies looking at the accuracy of the proposed service only against the reference standard. The principles outlined in the remainder of Section B3 still apply in relation to the assessment of the individual accuracy studies that generate information about the proposed investigative service only against the nominated reference standard.

Search strategies to identify accuracy studies

The primary objective of the search strategies is to locate all accuracy studies for the main indication, and compare the proposed investigative medical service directly with the main comparator. The search should involve at least three approaches:

- 1) a search of the published literature;
- 2) a search of registers of clinical trials (as many contain accuracy studies), as well as a search of registers of accuracy studies; and
- 3) manual checking of reference lists of all relevant articles that are obtained by other means.

When describing the search strategies and characteristics for (a), (b) and (c), sufficient detail should be provided so that an independent replication of the search would yield the same results.

The methods used to search the published literature are pivotal to assessing the completeness of the overall search. Therefore, specify the following characteristics of the search strategy:

- the specific databases and registers of clinical trials searched, including at least MEDLINE, EMBASE, the Cochrane Library (including the Cochrane Database of Systematic Reviews, and the Cochrane Central Register of Controlled Trials), the National Institutes of Health, the Australian Clinical Trials Registry (ACTR) and the Cochrane Register of Diagnostic Test Accuracy Studies. The search should also include databases internal to the company or professional group and any other known registers of randomised trials relevant to the therapeutic area;
- the date the search was conducted;
- the date span of the search (which should include the most recent update of each database searched);
- the complete search strategies used, including the search terms (key or MeSH words) and the relationship (sets and Boolean) between the search terms; and
- any supplementary searches, especially manual checking of references in the retrieved papers from the database searches.

If no relevant studies directly comparing the proposed investigative medical service and its main comparator have been retrieved in response to the systematic searches, the search criteria should be broadened to identify all accuracy studies. This involves relaxing the inclusion criteria to identify all accuracy studies involving either the proposed investigative medical service or the main comparator, preferably with similar study populations and a common reference standard.

B3.3 Listing of all accuracy studies

INFORMATION REQUIREMENTS

- The assessment report must identify and list all relevant literature relating to accuracy.
- If no relevant studies on accuracy are found in the searches, a 'nil return' must be included in the assessment report.

INFORMATION REQUESTS

- Present tables listing all citations of accuracy studies identified from the search of the published literature and other sources. Studies directly comparing the proposed investigative medical service and the main comparator service should be cited first, followed by indirect comparisons where there is a common reference standard.
- Alternatively if accuracy studies assessing the proposed service against the nominated reference standard only apply then list all relevant citations
- Show the inclusion and exclusion criteria for identifying relevant accuracy studies and state which accuracy studies have been published.
- On the hard copy of each of the search printouts supplied as technical documents with the assessment report, annotate each citation to indicate excluded citations with the reason for the exclusion.
- Collate all reports of each accuracy study to create a master list and indicate the preferred identification (ID) for each study to be used throughout the assessment report for consistency.
- Justify the exclusion of any relevant accuracy study. Tabulate a summary highlighting key aspects of the identified study, presenting included and then excluded studies.
- Separately identify any meta-analysis of accuracy studies and assess their exclusion or inclusion using the same criteria as above. Include any relevant systematic reviews (e.g. from the Cochrane Database of Systematic Reviews of Diagnostic Test Accuracy).
- Include copies (or sufficient details) of the included accuracy studies as attachments in the main body of the assessment report and ensure that the location of each item is clearly shown in the assessment report index.

If no relevant accuracy studies are found in the searches, the assessment report **must** include a statement to this effect with the results of the searches.

B3.3a Listing accuracy studies of direct comparison (where relevant)

Search results

Assess all citations retrieved by the searches to extract all accuracy studies that meet each of the following inclusion criteria:

- (a) the study directly compared the proposed (or index) investigative medical service and the main comparator;
- (b) the study contained the nominated reference standard; and
- (c) the accuracy study recruited participants with characteristics that overlap with those of patients with characteristics of the main indication.

Of these criteria, (b) and (c) require an element of judgement. If there is any uncertainty about whether to include or exclude an accuracy study, it is usually wiser to include it.

Tables B3.3.1 provides a suggested format for presenting the search results to summarise the inclusion and exclusion of citations from the results of searches reported in this Section. Other suggested formats are the PRISMA flowchart (<http://www.prisma-statement.org/statement.htm>).

Table B3.3.1 Summary of identification of accuracy studies of direct comparison from the search of the published literature

	MEDLINE	EMBASE	Trial or accuracy registries	Other databases
Number of citations retrieved by search				
Number of citations excluded after title/abstract review: <ul style="list-style-type: none"> • study does not include a direct comparison of the proposed investigative medical service and the main comparator • study did not include the nominated reference standard • characteristics of the recruited participants do not overlap with the main indication TOTAL				
Number of citations excluded after full text review: <ul style="list-style-type: none"> • study does not include a direct comparison of the proposed investigative medical service and the main comparator • study did not include the nominated reference standard • characteristics of the recruited participants do not overlap with the main indication TOTAL				
Number of citations of accuracy studies included from each database				
Consolidated number of citations of accuracy studies (removing exact duplicates across different databases)				
Number of multiple (additional) citations of accuracy studies identified				
Number of published accuracy studies included				

Note: Present columns that correspond with submitted printouts (e.g. if the printouts combine MEDLINE and EMBASE, these results can be combined in the table).

Annotated search printouts

On the hard copy of each of the search printouts supplied as technical documents with the assessment report, annotate each citation as appropriate with the letter (a), (b) or (c) to indicate which of the above criteria was invoked to exclude that citation. Provide only key citations that were excluded that may be of interest to the Committee. Each citation without an annotation should thus be a report of an accuracy study included in the assessment report.

Master list of trials

Table B3.3.2 provides a suggested format for presentation of a master list of all the direct accuracy studies identified in the search.

Table B3.3.2 Accuracy studies of direct comparison (and associated reports) presented in the assessment report

Accuracy study	Reports
Unique identification (ID) of accuracy study used in remainder of assessment report	Internal study report title. Date. Author(s). Title. <i>Journal</i> Year; Vol(No);pages Author(s). Title. <i>Journal</i> Year; Vol(No);pages

Meta-analyses

Separately identify any meta-analysis of accuracy studies that directly compare the proposed investigative medical service and the main comparator from the suite of searches above, and assess their exclusion or inclusion using the criteria above. This should include any relevant systematic reviews from the Cochrane Database of Systematic Reviews.

If a published meta-analysis of accuracy studies that directly compares the proposed investigative medical service and the main comparator is the principal source of the presented clinical evaluation, provide a copy of the publication as an attachment in the main body of the assessment report. Assess whether the published meta-analysis has a well-defined clinical question relevant to the intended listing of the proposed investigative medical service, a reproducible literature search strategy and appropriate criteria for any exclusions of identified accuracy studies. Assess the meta-analysis using the framework provided in this Sub-section alongside the presentation of the individual accuracy studies. Where there is more than one such meta-analysis, tabulate these assessments.

Exclusion of studies

Justify the exclusion of any accuracy study included in the master list in Table B3.3.3 from further detailed assessment in the assessment report. The grounds for exclusion might include the quality of the study, the patient characteristics or the absence of the nominated reference standard. This might minimise observable differences across the accuracy studies, or examine and explain, where possible, their contribution to heterogeneity across all the trials.

It is not possible to give unequivocal guidance on the exclusion of accuracy studies at this stage. If a decision to exclude or include one or more accuracy studies is likely to be controversial, it is usually wiser to also present a sensitivity analysis examining whether that decision makes a difference to the conclusions from the overall clinical evaluation.

If one or more accuracy studies that directly compare the proposed investigative medical service and the main comparator are to be excluded, identify those aspects of each study that cause the exclusion (see Table B3.3.3). If there is more than one type of reason for exclusion, arrange the excluded trials in Table B3.3.3 by the reason for exclusion.

Table B3.3.3 Reasons to exclude each accuracy study of direct comparison from further detailed assessment

Study ID	Ground(s) for seeking exclusion	Details ^a
Quality of studies		
Study A		
Etc.		
Patient characteristics and circumstances of use in the studies		
Study B		
Etc.		
Measure(s) of accuracy reported in the studies		
Etc.		

ID = identification

^a Cross-reference each set of details to the source of information (specifying the trial report with page, table, and figure number).

Present tables summarising key aspects and the results of all the identified accuracy studies (included and the excluded; see Tables B3.3.4 and B3.3.5).

Table B3.3.4 Comparative summary of characteristics of accuracy studies of direct comparison

Study ID	Design characteristics ^a	Compared investigative medical services	Summary of main population characteristics
Included studies			
Study 1			
Etc.			
Excluded studies			
Study A			
Etc.			

ID = identification

^a Cross-reference each set of details to the source of information (specifying the trial report with page, table, and figure number).

Table B3.3.5 Comparative summary of results of accuracy studies of direct comparison

Study ID	Measure(s) of accuracy (95% CI)	Major adverse events
Included studies		
Study1		
Etc.		
Excluded studies		
Study A		
Etc.		

CI = confidence interval; ID = identification

Presentation of non-inferiority (equivalence) accuracy studies

Most accuracy studies are designed to show a difference between the investigative medical services being compared.

If any accuracy study was designed as a non-inferiority study, and/or the investigative conclusion presented in Sub-section 3.9 is non-inferiority or equivalence, refer to the additional guidance on presenting the information in Appendix 3.

Non-inferiority means that, in terms of accuracy, the proposed investigative medical service is no worse than its main comparator. Non-inferiority is used to *support* a claim of equivalence; it is not adequate to demonstrate the *absence* of a statistically significant difference between the investigative medical services to claim equivalence. This lack of a significant difference might occur when the studies are too small to demonstrate a real difference in the accuracy of the investigative medical services being compared. The appropriate comparison to present is the point estimate of the difference with its 95% confidence interval. This allows MSAC to assess whether the confidence interval contains the minimal clinically important difference.

Study details

Include sufficient details of the relevant accuracy studies as attachments in the main body of the assessment report. Where there is more than one report of accuracy study (e.g. one or more published papers), provide both the published paper(s) and key extracts. The results might vary between reports of the same accuracy study. If so, justify and cross-reference the selection of the source of results extracted for the assessment report. Provide a copy of each publication that reports data from a listed accuracy study. Ensure that the assessment report index shows the location of all submitted papers, both in the main body of the assessment report and in the attachments.

For any relevant accuracy study identified from a meta-analysis, include the individual study report or publication(s), as above. If no separate report is available, indicate the efforts made to retrieve them and to obtain any missing information from the authors of the published meta-analysis.

B3.3b Listing accuracy studies for inclusion in an indirect comparison (where relevant)

Search results

Assess all citations retrieved by the expanded searches to extract all accuracy studies that meet the following inclusion criteria for accuracy study to support one or more indirect comparisons involving a common reference standard that is also the nominated reference standard from Sub-section 3.1:

- accuracy studies simply comparing the proposed (index) investigative medical service against the nominated reference standard;
- accuracy studies simply comparing the main comparator against the nominated reference standard; and
- accuracy studies that recruited participants with characteristics that overlap those of patients who would be eligible for the main indication.

Adapt the guidance given earlier in this Sub-section to present the results of the searches, and to list and provide details of all the accuracy studies that meet the inclusion criteria separately for the proposed investigative medical service and the main comparator. In addition to the two tables presented to list any accuracy studies of direct comparison, replicate the format of those tables to present the expanded searches for all accuracy studies of the investigative medical service and the main comparator.

Annotated search printouts

Present annotated search printouts as described above in Section B3.3a.

Master list of trials

From the two tables reporting the results of the expanded searches for the proposed investigative medical service, list the most relevant and persuasive citations for the proposed investigative medical service. Similarly, list all identified relevant citations of accuracy studies for the main comparator. Table B3.3.6 provides a suggested format for presenting a master list of all the relevant accuracy studies identified in the search for the indirect comparison.

Table B3.3.6 Accuracy studies (and associated reports) presented in the assessment report

Study	Description	Reports	Comparable?
Common reference (should also be the nominated reference standard)			
Proposed investigative medical service			
Unique identification (ID) of study used in remainder of the assessment report	Brief description of study	Internal study report title. Date. Author(s). Title. <i>Journal</i> Year; Vol(No):pages Author(s). Title. <i>Journal</i> Year; Vol(No):pages	Yes/No
Main comparator			
ID of study used in remainder of the assessment report	Brief description of trial	Internal study report title. Date. Author(s). Title. <i>Journal</i> Year; Vol(No):pages Author(s). Title. <i>Journal</i> Year; Vol(No):pages	Yes/No

Presentation of non-inferiority (equivalence) trials

If an indirect comparison is provided to support a diagnostic conclusion of non-inferiority or equivalence in Sub-section 3.9, see Appendix 3 for additional guidance on the presentation of the information.

Assess comparability of identified accuracy studies to justify any exclusions

Observable differences across the accuracy studies should be minimised, or their contribution to heterogeneity across the studies examined and adjusted where possible. By definition, non-observable differences cannot be minimised or adjusted, and this contributes to the residual uncertainty inevitably associated with indirect comparisons.

Aspects that might justify the exclusion of accuracy studies from an indirect comparison include:

- important differences in the quality of the studies being compared;

- important differences in baseline patient characteristics; for example, measure(s) of accuracy generated in one study involving patients with severe disease might not be comparable with measure(s) of accuracy generated in another study involving patients with mild disease;
- differences in measure(s) of accuracy reported; and
- differences in the reference standard.

It is not possible to give unequivocal guidance on the exclusion of accuracy studies from an indirect comparison. The justification to exclude an accuracy study should anticipate whether this would raise issues of selection bias, while the justification to include an accuracy study should anticipate whether this would raise issues of comparability. If a decision to exclude or include one or more accuracy studies is likely to be controversial, it is usually wiser to also present a sensitivity analysis examining whether the decision makes a difference to the conclusions from the overall clinical evaluation.

If one or more accuracy studies are to be excluded from an indirect comparison, identify the aspect(s) of each study that form the reasons for the proposed exclusion (as per Table B3.3.7). Indicate whether each reason relates to the quality of the trials, the patient characteristics and circumstances of use, and/or indices of accuracy reported in the studies.

Table B3.3.7 Reasons to exclude each accuracy study from the indirect comparison

Study ID	Ground(s) for seeking exclusion	Details ^a
Quality of the study		
Study 1		
Patient characteristics and circumstances of use in the trial		
Study 2		
Measure(s) of accuracy reported in the trial		
Etc.		

ID = identification

^a Cross-reference each set of details to the source of information (specifying the trial report with page, table, and figure number).

Study details

Present the included comparable accuracy studies in the main body of the assessment report and attach a report of each study to the main body of the assessment report. Provide a report of each included (single-arm) accuracy study in a separate volume of the assessment report. Provide clear cross-references between the presentation of the trials and the reports.

B3.4 Assessment of the measures taken by investigators to minimise bias in accuracy studies

INFORMATION REQUESTS

- For each accuracy study listed, provide information on the measures taken to minimise bias, using the checklist provided.
- For each checklist response, specify the source document in the reports or papers accompanying the main body of the assessment report, together with the page or table from which the information was extracted.

Assessment of measures to minimise bias

The purpose of assessments of measures to minimise bias is to provide MSAC with a clear idea of which accuracy studies are of greater scientific rigour. There is no minimum standard but MSAC is most likely to be persuaded by the data of the highest scientific rigour.

Several quality assessment instruments are available for accuracy studies, including the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS 2) tool (refer to <http://www.ncbi.nlm.nih.gov/pubmed/22007046>), the Standards for Reporting of Diagnostic Accuracy (STARD) initiative and the ACCE³ Model Project (for genetic tests).

For simplification, the remainder of this Section is based on the QUADAS 2 quality appraisal tool. The indicative checklist in Box B3.4.1 adapted from the QUADAS tool includes the topics that help to assess the methodological quality of each study. This is a useful guide to help MSAC review the scientific rigour of the evidence by assessing the measures taken by the investigators to minimise bias. It is not intended to discourage the presentation of data.

Several factors threaten the internal and external validity of an accuracy study. Some of these factors relate to the design of such studies; others relate to the selection of participants, the execution of the tests or the analysis of the data.

³ ACCE gets its name from [analytic validity](#); [clinical validity](#); [clinical utility](#); and associated [ethical, legal and social implications](#).

Box B3.4.1 Indicative checklist for assessing the quality of specific accuracy studies

1. Was the spectrum of patients representative of the patients who will receive the investigative medical service under consideration?
2. Were selection criteria clearly described?
3. Is the reference standard likely to correctly classify the target condition?
4. Is the time period between reference standard and investigative medical service short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample, or a random selection of the sample, receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the result arising from the investigative medical service?
7. Was the reference standard independent of the investigative medical service (i.e. the investigative medical service under consideration did not form part of the reference standard)?
8. Was the execution of the investigative medical service described in sufficient detail to permit replication of the service?
9. Was the execution of the reference standard described in sufficient detail to permit its replication?
10. Were the results arising from the investigative medical service interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results arising from the investigative medical service?
12. Were the same clinical data available when test results were interpreted as would be available when the investigative medical service is used in practice?
13. Were uninterpretable/intermediate results arising from the investigative medical service reported?
14. Were withdrawals from the study explained?

Notes for quality checklist***(a) Participant characteristics and recruitment***

An important element of any accuracy study is how eligible participants were identified and recruited, because demographic and clinical features of the study population can affect measures of accuracy. There are two aspects to this item: first, whether the right participant group was recruited to the study to address the review question; and second, whether the method of sampling participants for inclusion from this group was likely to yield a representative sample. Ideally, the study should enrol consecutive participants clinically suspected of the target condition because of presenting symptoms or referral by another health care professional. The participants then undergo the investigative medical service under consideration, as well as the reference standard at the same time. However, when a delay occurs between performing the investigative medical service and the reference standard, the condition of the participant might change, leading to worsening or improvement of the target condition or alternative conditions. Similar concerns apply if a therapeutic medical service is started after performing the investigative medical service, but before performing the reference standard.

Other studies, such as diagnostic case-control studies, select a sample of participants already known to have the target condition and are then compared with a separate group

of normal/healthy people known to be free of the target condition. In this situation, participants with borderline or mild expressions of the target condition are potentially excluded, which can lead to exaggeration of both sensitivity and specificity. This is called *spectrum bias*, because the spectrum of study participants will not be representative of patients seen in practice. It is therefore important that accuracy studies include an appropriate spectrum of participants and also a clear description of the population included in the study. An appropriate participant spectrum should be defined in light of the research question, stating key factors that could affect accuracy, such as setting, disease severity and prevalence, and prior testing. Where it is possible that a small proportion of inappropriate participants would be tolerated, this proportion should be stated. In some reviews, exclusion of inappropriate sampling methods might be part of the eligibility criteria (e.g. exclusion of studies that have enrolled a group of healthy controls). Reported estimates of accuracy might have limited clinical applicability (generalisability) if the spectrum of tested participants is not similar to the patients in whom the test will be used in practice.

(b) Reference standard

The validity of the reference standard used in each accuracy study should be determined in the context of the condition under review. Criteria for determining the validity of the reference standard should be pre-specified. This can include the choice of the reference standard(s) and its timing in relation to the investigative medical service under consideration. When the accuracy of an investigative medical service is determined by comparison with an imperfect reference standard, some target condition misclassification will be introduced. Misclassification of the target condition by the reference standard will tend to result in underestimation of the accuracy of the investigative medical service under consideration. Underestimation of the sensitivity is most likely when the prevalence of the target condition is low, and the estimated sensitivity will be closer to the true sensitivity with increasing prevalence. Underestimation of the specificity will occur most when the prevalence of the target condition is high, and the estimated specificity will be closer to the true specificity when the prevalence of the target condition is low. Good reference standards are independent of the investigative medical service under consideration, and are applied blindly or objectively applied to all patients. Poor reference standards are haphazardly applied, but still independent of the investigative medical service under consideration. Use of a non-independent reference standard implies a lower level study. Other lower level accuracy studies include studies of diagnostic yield, which generate a yield of diagnosed patients, as determined by an investigative medical service, without confirmation of the accuracy of this diagnosis by a reference standard. These might be the only alternative when there is no reliable reference standard.

(c) Blinding

Blinding of participants, investigators or those responsible for assessing the outcomes helps prevent several important biases in accuracy studies. Blinding of participants and investigators might influence several aspects of the study. Knowledge of the results of the reference standard can influence the reading of the investigative medical service under consideration, and vice versa. Such knowledge is likely to increase the agreement between results of the investigative medical service and those of the reference standard, leading to inflated measures of accuracy. The distortion of measures of accuracy caused by knowledge of the result of the reference standard while interpreting the investigative medical service is known as *test review bias*. Conversely, knowing the result of the investigative medical service while interpreting the reference standard is known as

diagnostic review bias. The observation that interpretations become more accurate by providing additional clinical information to interpreters is known as *clinical review bias*. If blinding was used in an accuracy study under consideration, describe the methods used. If blinding was not used, list the reasons why the study did not blind the participants, investigator(s) or outcome assessors. Discuss the effect, if any, that the absence of blinding might have had on the measure(s) of accuracy generated.

(d) Withdrawals

Withdrawals occur when participants withdraw from the study before the results of either or both the investigative medical service and reference standard are known. If participants lost to follow-up differ systematically from those who remain, for whatever reason, then estimates of accuracy might be biased.

(e) Indirect comparisons

It is not possible to minimise bias across the indirect comparison beyond the assessment of comparability and selection bias. For studies deemed comparable for the assessment report, it is important to identify any differences that might exist in the quality of the trials across the indirect comparison.

Tabulate responses

If there are multiple accuracy studies, tabulate the responses in the main body of the assessment report, with the detailed responses to the above questions in an accompanying attachment or technical document. In this detailed presentation, also provide adequate cross-references to the trial report (including page, table or figure numbers of the source document) from which each aspect of the information was extracted.

Tables B3.4.1 and B3.4.2 provide a suggested format for presenting the summary in the main body of the assessment report.

Table B3.4.1 Summary of the measures undertaken to minimise bias

Study	RISK OF BIAS			APPLICABILITY CONCERNS			
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
-	-	-	-	-	-	-	-

☺ Low Risk ☹ High Risk ? Unclear Risk

Table B3.4.2 Flow of participants in each accuracy study of direct comparison under consideration

Study ID	Did not receive	Withdrawals	Analysed	Source of information
Study 1				(Add this column to tables and submit in a separate technical attachment)
<ul style="list-style-type: none"> • Proposed investigative medical service 	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	
<ul style="list-style-type: none"> • Main comparator 	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	

ID = identification; *n* = number of participants with event

For indirect comparisons, compare and assess the minimisation of bias in the studies across each set of studies forming the indirect comparison. Present additional tables for indirect comparisons similar to the above table.

Source documents

For each of the responses provided in Tables B3.4.1 and B3.4.2, specify the source document in the reports or papers accompanying the main body of the assessment report. Provide adequate detail of cross-referencing to page, table or figure number of the relevant study report(s) in a way that does not detract from the presentation of the requested results.

For the presentation of a complex systematic overview, consider re-presenting the tables from the main body of the assessment report in a technical document or attachment, and add an additional column to each table to provide adequate detail of cross-referencing (as illustrated by the shaded column in Table B3.4.2). Alternatively, if it is clearer for some tables, identify the source of information cell by cell, using footnotes.

B3.5 Characteristics of accuracy studies

INFORMATION REQUESTS

- For each accuracy study, provide the following details of the study protocols and participants:
 - the eligibility criteria for participants considered for recruitment into the study;
 - the baseline demographic and clinical characteristics of each group; and
 - the nature of follow-up (median and range) and whether the study has been completed or is ongoing.
- For each response, specify the source document in the reports or papers accompanying the main body of the assessment report, together with the page or table from which the information was extracted.

Details of accuracy studies

If there are multiple accuracy studies, tabulate the responses in the main body of the assessment report. Tables B3.5.1 and B3.5.2 provide a suggested format.

Table B3.5.1 Eligibility criteria in the accuracy studies under consideration

Study ID	Inclusion criteria	Exclusion criteria
Study 1		
Study 2		
Etc.		

ID = identification

Indicate any significant differences in the baseline characteristics of patients across the accuracy study and discuss any impact this might have on the interpretation of accuracy results. Table B3.5.2 provides a suggested format for this information.

Table B3.5.2 Characteristics of participants in the accuracy studies

Study ID (baseline characteristics)	Spectrum of participants receiving proposed investigative medical service	Spectrum of participants receiving proposed investigative medical service
Study 1		
Age		
Sex		
Prior testing		
Disease severity		
Study setting		
Study 2		
Age		
Sex		
Prior testing		
Disease severity		
Study setting		
Etc.		

ID = identification

Provide any additional information about the trial or participant characteristics that is not requested elsewhere in Section B but is relied on in assessing the applicability of the accuracy evidence to the listing requested.

Characteristics of the accuracy studies included in an indirect comparison (where relevant)

For accuracy studies deemed indirectly comparable for the assessment report, it is particularly important to assess the baseline characteristics of the participants recruited into the studies and the nature of common reference standard.

Similarly, assess how far apart in time and place the studies were conducted. This is necessary because changes in medical practice and participant characteristics might mean that, nominally, the studies might not be comparable, especially when they have been conducted at different times or in different geographical regions. Such changes might confound the indirect comparison.

Source documents

For each of the responses provided above, provide adequate cross-referencing to page, table or figure numbers of the relevant trial report(s), if necessary in a separate technical document or attachment.

B3.6 Systematic overview of the results of the accuracy studies

INFORMATION REQUESTS

- For each accuracy study, present the results of the primary analysis for that trial.
- Present an analysis of the results for each type of primary measure of accuracy reported in tables with graphed forest plots.
- Where there are multiple accuracy studies reporting the same measures of accuracy, statistically combine (meta-analyse) the results, where possible.
- Assess the potential for outcomes reporting bias by reporting, in a footnote, the presentation of the forest plot for each outcome:
 - the number of accuracy studies contributing to the forest plot; and
 - the proportion of these accuracy studies over the total number of accuracy studies included in Table B3.3.1.
- For each response, specify the source document in the reports or papers accompanying the main body of the assessment report, together with the page or table from which the information was extracted.

Once the methodological quality of the studies included in the review has been assessed and a narrative summary of quality presented, applicants should consider how the information collated should be considered in the analysis and when informing an investigative conclusion. Applicants need to distinguish between the inclusion of studies for the application and the inclusion of studies into specific meta-analyses. The inclusion criteria for the application can be broad, while specific meta-analyses might be focused on a subgroup of studies that can be reasonably combined.

The presentation of the results of accuracy studies serves two purposes:

- first, the presentation of the results of the primary analyses, as established for each accuracy study, is part of the assessment of the scientific rigour of the study dataset and becomes a reference point for interpreting other measures of accuracy generated for that study; and
- second, the presentation of the results of common measure(s) of accuracy across more than one study enables an assessment to be made of the comparative accuracy of the proposed investigative medical service and the main comparator under the circumstances of the studies as designed and conducted.

Primary analysis

For each study listed in Sub-section B3.3, present the results of that study according to the design of the pre-specified primary analysis for that study. Primary measures of accuracy likely to be reported in such studies include sensitivity, specificity, likelihood ratios, receiver operator characteristics (ROC) curves and the diagnostic odds ratio (DOR).

Estimates of accuracy are subject to chance variation, with larger studies usually resulting in more precise estimates. Authors should therefore quantify the amount of statistical uncertainty around the observations. Two different plots, the paired forest plot and the summary ROC plot, can be used to report the results of the individual studies.

If an accuracy study stipulates issues around the impracticality of the reference standard available, and only partially used it to the extent possible, comment on how the authors calculated estimates of sensitivity and specificity adjusted to correct for any (verification) bias that might have been introduced by not using the reference standard to its fullest extent.

The usual formulas for calculating sensitivity and specificity will give biased estimates of sensitivity and specificity (i.e. *verification or workup bias*) in the following situations:

- if the study authors determined that using a reference standard on all participants was impractical or not feasible;
- if study authors estimated sensitivity and specificity using the proposed investigative medical service and a comparative method (other than a reference standard) on all participants; and
- if the study authors used the reference standard on just a subset of participants (sometimes called partial verification studies or two-stage studies).

If the designated reference standard was applied to a random subset of all participants, or to all participants where the investigative medical service and the main comparator disagree, and to a random sample of participants where they agree, then it is possible to compute adjusted estimates (and variances) of sensitivity and specificity. In this case, comment on whether the authors retested a sufficient number of participants to estimate sensitivity and specificity with reasonable precision.

Analysis (including meta-analysis)

Because evaluating accuracy requires knowledge of two quantities — sensitivity and specificity — meta-analysis methods for accuracy have to deal with two summary statistics simultaneously rather than one (as is the case for reviews of therapeutic medical services). A meta-analysis of accuracy also has to allow for the trade-off between sensitivity and specificity that occurs between studies that vary in the threshold value used to define test positives and test negatives. Methods for undertaking analyses that account for both sensitivity and specificity, the relationship between them, and the heterogeneity in accuracy, require fitting hierarchical random effects models. Therefore, collaboration with a statistical expert is highly recommended.

Meta-analyses of accuracy are useful because they might increase the precision of the estimates of differences between the proposed investigative medical service and the main comparator. It is also useful when there are conflicting results from studies of similar

scientific rigour. Meta-analysis can also highlight advantages of a proposed investigative medical service that are too small to be detected reliably in individual accuracy studies, but might be clinically important. Justify any decision *not* to present a meta-analysis.

Where there is more than one study reporting a particular primary measure of accuracy, the presentation of a meta-analysis, which statistically pools results across trials, is generally preferred where appropriate. Collate the results of each study reporting into a meta-analysis and present the results of each meta-analysis in a table and as a graphed forest plot, including the pooled results from across the studies.

Where a meta-analysis is based on a subset of all available accuracy studies (e.g. only accuracy studies where there has been a direct comparison between the proposed investigative medical service and its main comparator), identify the studies in the subset. Report the number of studies in the subset and the proportion that this number represents of the total number of studies listed in Sub-section B3.3. Examine whether there are any differences between the results of the subset and the total set of studies using group-level data, and assess the impact of any bias across any differences detected.

Explain and justify any other method used for statistically combining the results of the accuracy studies and any additional statistical tests used. Clearly document and reference the methods used to make them independently reproducible and verifiable. Provide adequate detail of all sources of information relied on for these other analyses and then present their results.

Identifying and interpreting heterogeneity between accuracy studies

Heterogeneity is to be expected in meta-analyses of accuracy. Variation in accuracy is common and can be due to random error or true differences (heterogeneity) between the studies. Any differences in the results of studies that address the same research question should be clearly identified and interpreted in the assessment report. Potential sources of heterogeneity include:

- different definitions of target condition or reference standard;
- different test or test procedures;
- different test thresholds;
- different spectrum of disease in the tested population due to different criteria for subject selection, including prior tests or referral setting; and
- different spectrum of non-disease in the tested population due to different criteria for subject selection and different prevalence of differential diagnoses in diverse referral settings; for example, comorbidities that result in a higher rate of false positives.

If differences in the results cannot be attributed to these known sources of variation, then pooling the results should not be attempted because it will not be possible to interpret the summary estimate. However, reporting the range of sensitivities and specificities provides a useful summary of the variation observed.

Three simple statistical methods to test for heterogeneity of sensitivity and specificity are:

- plot the sensitivity and specificity from each study with their 95% confidence intervals in a table and/or forest plot to illustrate the range of estimates and identify outliers;
- if sufficient data are available, plot the paired sensitivity and one minus the specificity results for each study on the ROC plane to detect heterogeneity and identify outliers; study variation due to differences in the threshold used to define a positive test will produce a symmetrical curve that resembles the underlying ROC curve for the test; and
- use a relevant test statistic to test the hypothesis that there is no statistically significant difference in the sensitivity and specificity reported. This statistical test for differences in proportions provides a conservative test of the null hypothesis that the study results are homogeneous.

Report results for statistical heterogeneity alongside relevant test statistic with its 95% uncertainty interval.

Pooling data

Data pooling should only be considered for studies that address the same clinical question, meet the pre-specified quality criteria, have patients recruited from clinically similar populations, and use comparable investigative medical services and reference standards. The steps required for pooling data are:

- (a) define the criteria used to select studies for the meta-analysis;
- (b) assemble the dataset from eligible studies;
- (c) justify assumptions made about included studies; for example, random/non-random sample of population and test procedures;
- (d) test for the presence/absence of heterogeneity (see above);
- (e) test for the presence/absence of threshold effect (see below); and
- (f) report on the methods and results for pooling.

Pooling sensitivity and specificity

If there is no heterogeneity between studies and no threshold effect is observed, then sensitivities and specificities for each investigative medical service might be pooled and compared. When pooling results, the statistical model used depends on what assumptions can be made about the group of studies selected. The two models are:

- **Fixed effects model:** all the studies are assumed to represent a random sample of one large common study. Under this assumption, the differences between study results are considered to be the result of random error. The data can be pooled with a weighting for individual studies based on the inverse of the variance of the parameter of test accuracy (i.e. precision) or the number of participants.
- **Random effects model:** differences between the studies are assumed to be due to real differences between the study populations and procedures, not just random differences. Under this assumption, a more complex mathematical model is used to weight studies, taking within-study and between-study variation into account.

Pooling likelihood ratios

Summary likelihood ratios can be estimated from the pooled estimates of sensitivity and specificity, or, more preferably, by using standard methods of meta-analysis of risk ratios. The latter method allows testing for heterogeneity between studies. Pooling likelihood ratios is useful to transfer the results of the meta-analysis to a clinical context; for example, by providing a summary estimate of what proportion of participants with a negative result will have the target condition (on the basis of the average prevalence of the condition in the studies available).

Producing a summary DOR and ROC curve

If the DOR is constant regardless of the diagnostic threshold, then the summary DOR for the proposed test and comparator can be presented with a 95% confidence interval to compare differences in diagnostic performance.

Identifying a test threshold effect

Variation in accuracy across studies included in the review might be due to differences in the explicit (pre-specified) or implicit (observer-related) threshold used to define a positive result generated by an investigative medical service. This effect can be observed when the paired sensitivity and one minus the specificity results for each study are plotted in the ROC plane and display the trade-off between sensitivity and specificity. The regression model used to fit the summary ROC (SROC) curve can be used to test for this effect. In general terms, if the logarithmic transformed DORs (ln DOR) are homogenous across different studies, then the SROC curve will be symmetrical around the line sensitivity (which is equivalent to the specificity) and variation in these end points between studies can be attributed to a threshold effect. When this is the case, the SROC represents the most informative synthesis of evidence about test accuracy and the pooled DOR is a useful single summary measure. If a threshold effect exists, the DOR changes with the test threshold and the SROC will be asymmetric. Similar patterns of variation can also be observed across studies due to differences in the spectrum of disease, and so some caution is required when interpreting variability as a threshold effect.

Table B3.6.1 provides a suggested format for presenting and comparing accuracy data generated from several studies.

Table B3.6.1 Results of accuracy studies directly comparing proposed and comparator investigative medical services

Study ID	Proposed investigative medical service Measures of accuracy 95% CI	Main comparator Measures of accuracy 95% CI	Forest plot here
Study 1			
Study 2			
Etc.			
Pooled result from relevant statistical model			
Chi-square (Q) for heterogeneity: $P =$ I^2 statistic with 95% uncertainty interval =			

CI = confidence interval; ID = identification

Note: Provide number and percentage of the identified relevant accuracy studies that contributed data to this meta-analysis.

Presenting the results of an indirect comparison

Table B3.6.2 suggests how to present results from accuracy studies where a primary analysis has been generated from an indirect comparison of the proposed investigative medical service and its main comparator with a common reference standard.

Table B3.6.2 Summary of results of the indirect comparison

Study ID	Studies of proposed investigative medical service	Common reference standard	Studies of main comparator
	Relevant measures of accuracy (95% CI)		Relevant measures of accuracy (95% CI)
Study 1		-	
Study 2			
Etc.			
Pooled			
			-

CI = confidence interval; ID = identification

When documenting and referencing any other methods used to quantify the results of the indirect comparison, ensure that the methods are reproducible and able to be independently verified. Where appropriate, assess the implications for the conclusions of the indirect comparison of excluding studies considered to be less comparable (e.g. in terms of patient populations). Alternatively, justify, describe and present any other adjustment of the indirect comparison.

Adverse event data

As a minimum, report important adverse events as the number of participants reporting:

- any adverse event;
- any adverse event resulting in discontinuation of an investigative medical service before it can be completed;
- any adverse event resulting in hospitalisation; and
- any adverse event resulting in death.

Sub-section B7 outlines how to present an extended assessment around comparative harms between the proposed and comparator investigative medical service.

B3.7 Extended assessment of reliability evidence

INFORMATION REQUESTS

- For each accuracy study listed, provide information on whether reliability of the investigative medical service was also measured or whether information was provided on how each study ensured reliability, using the checklist provided.
- For each checklist response, specify the source document in the reports or papers accompanying the main body of the assessment report, together with the page or table from which the information was extracted.

Analytical validity relates to whether an investigative medical service measures what it claims to measure, and combines the concepts of *accuracy* and *reliability*. The term *accuracy* refers to the amount of agreement between the investigative medical service under consideration and the reference standard; that is, the proportion of participants whom the investigative medical service correctly identifies as positive or negative. Section B has focused on accuracy, up to this point.

The term *reliability* (which is analogous to the concept of *precision*) refers to the amount of agreement of different operators or instruments applying the same investigative medical service. That is, a reliable investigative medical service is measuring something consistently. Reliability is sometimes referred to as *reproducibility* or *repeatability*.

The reproducibility of an observation depends on the variability of the same person or instrument making the observation 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*). Other terms for this form of variation include imprecision, analytic methodological variation or analytical noise (error). Reproducibility might be further affected by factors such as number of observers, tissue storage and processing, and so on. An investigative medical service that has poor reliability cannot have high validity. On the other hand, good reliability does not assure high validity.

Kappa statistics are the method of choice in an extended assessment of reliability. The kappa value is a statistical measurement for the intra-observer and inter-observer agreement corrected by chance. The kappa value is intended to give the reader a quantitative measure of the magnitude of agreement, standardised to lie on a –1 to 1 scale, where 1 is perfect agreement, 0 is exactly what would be expected by chance, and negative values indicate agreement less than chance (i.e. potential systematic disagreement among the observers).

When interpreting the kappa value, it is also important to keep in mind that the estimated kappa itself could be due to chance. To report a *P* value of a kappa requires calculation of the variance of kappa and deriving a *z* statistic, which are beyond the scope of these Guidelines. A confidence interval for kappa, which might be even more informative, can also be calculated. *P* values and confidence intervals are sensitive to sample size, and — with a large enough sample size — any kappa above 0 will become statistically significant.

Primary analysis of studies reporting reliability/reproducibility

Identify studies that clearly included reproducibility analysis of either the proposed investigative medical service or its main comparator; for example, if they reported assessing the same investigative medical service on the same specimens but under different conditions (such as different time intervals, operators or laboratories). If a specimen from the patient is required to perform the investigative medical service under consideration, identify whether the specimen required has been clearly identified, and whether this specimen needs to be collected specifically for the purposes of performing the test or has already been collected for another purpose.

Present any differences across laboratories in how they characterise results, such as the kappa or other relevant statistic. Identify whether there is an external quality assurance program by which the studies have specified how laboratories have benchmarked their assays. Evaluate the reproducibility of the methods used in a study. The characteristics and summary results of the included reproducibility studies should be presented as shown in Table B3.7.1.

Table B3.7.1 Characteristics and summary results of accuracy studies reporting reliability / reproducibility

Study ID	Study characteristics	Summary of reproducibility results
Study 1		
Study 2		
Etc.		

ID = identification

B3.8 Presenting a concordance analysis in the absence of a reference standard

Other indicators of validity (construct validity) may need to be used in the absence of a reference standard when assessing investigative medical services.

If a reference standard is not available or unacceptable for the requested use and/or requested population, the step to consider is whether one can be potentially constructed. If so, calculate estimated sensitivity and specificity under the constructed standard. In this situation:

- specify the designated reference standard that was constructed;
- create the new reference standard independently from the analysis of results of the proposed test (ideally, in advance of collecting any specimens); and
- consult with statisticians and health professionals prior to constructing the reference standard.

If a reference standard is not available and cannot be constructed, calculate and report measures of agreement (the terms sensitivity and specificity are not appropriate to describe these comparative results). Instead, the same numerical calculations are made, but the estimates are called positive percent agreement and negative percent agreement, rather than sensitivity and specificity. This reflects that the estimates are not of accuracy but of agreement of the investigative medical service with the non-reference standard. In addition, quantities such as positive predictive value, negative predictive value, and positive and negative likelihood ratios cannot be computed since the subjects' condition status (as determined by a reference standard) is unknown. In this situation:

- report the 2 x 2 table of results comparing the candidate test with the comparative method;
- describe the comparative method and how it was performed; and
- report the agreement measures along with their confidence intervals or kappa statistics.

Alternatively odds ratios could be reported indicating the likelihood of an outcome, given that particular test result.

The statistical methods underpinning a concordance analysis are similar to those described in Sub-section B3.7 of these Guidelines when considering the reliability of an investigative medical service. Engagement of a biostatistician is recommended if an applicant decides to undertake a concordance analysis.

B3.9 Interpretation of evidence on diagnostic performance

INFORMATION REQUESTS

- Provide a summary assessment of the overall evidence presented for diagnostic performance.

It is proposed that a conclusion be drawn at the end of Section B3 about whether the proposed investigative medical service is non-inferior (no worse than) or superior compared to its alternatives in terms of diagnostic performance.

The interpretation of the clinical data presented in Section B of the assessment report is crucial in determining the success of the assessment report. It is important to classify the diagnostic performance of the proposed investigative medical service in relation to its main comparator (i.e. whether it is superior, inferior or equivalent to the comparator in terms of diagnostic performance).

The essential difference between assessing whether the proposed investigative medical service is superior diagnostically or non-inferior to the main comparator is that the 95% confidence interval for superiority excludes the possibility that there is no difference between the medical services. The 95% confidence interval for non-inferiority, however, excludes the possibility that the investigative medical service is inferior to a clinically important extent. In each case, the interpretation of the point estimate and its 95% confidence interval is compared to the null hypothesis of the assessment. In the case of a superiority assessment, the null hypothesis is that there is no difference between the compared alternatives. In the case of a non-inferiority assessment, the null hypothesis is that the difference between the compared alternatives is no worse than the minimal clinically important difference.

Distinguishing between non-inferiority and superiority in relation to diagnostic performance at this stage of the assessment report is important as this distinction may or may not translate into being the same overall clinical conclusion made in Section B8 and thus provide opportunities for an abridged linked evidence approach for some applications, with evidentiary requirements satisfied by accuracy studies meeting the transferability assumption. In other words, evidence of investigative service accuracy may be a suitable proxy in some circumstances for clinical effectiveness of an investigative service. For example, if an investigative medical service is non-inferior in terms of diagnostic performance it is usually expected that there would be no change in clinical management or change in health outcomes (Section B5) provided that there is indeed (a) no studies demonstrating change in management (i.e. do physicians trust the test results); (b) there is truly no differences between the proposed service and main comparator in terms of safety (Section B7). In this instance a cost-minimisation analysis is sufficient (see to Part III, Section D(i)).

A superior claim of diagnostic performance is necessary to support a subsequent claim of a change in clinical management, translating to a change in health outcomes for at least proportion of tested patients and thus an incremental cost-effectiveness analysis. Superiority of accuracy can only be translated into a change in health outcomes through a linked analysis or through direct evidence. However, it must not always be assumed that a superior investigative service in terms of diagnostic performance will translate into superior health outcomes. A scenario may arise where an investigative service may be more accurate than the main comparator but it (a) may be more invasive than the main comparator to perform and thus be associated with higher rate of complications, (b) either identify inconsequential disease or identify disease where benefit of treatment for patients true positives is small or non-existent and is outweighed by the harms of treatment. Hence the importance to complete the remaining components of a linked analysis to see whether or not superior diagnostic performance does indeed translate (or not) into superior health outcomes.

B4 Clinical validity

INFORMATION REQUESTS

- For applications to MSAC where this section is relevant, provide information on whether clinical validity was measured in the literature.
- Specify the source document in the reports or papers accompanying the main body of the assessment report, together with the page or table from which the information on clinical validity was extracted.

B4.1 Measures of clinical validity

Clinical validity relates to whether an investigative medical service answers the clinical question being asked, and refers to how an investigative medical service detects or predicts the target condition under consideration. Prevalence (or pre-test probability) of the target condition or clinical information of interest comes into consideration when determining clinical validity. **Positive and negative predictive values** are indices of clinical validity, and are the probabilities of disease or absence of disease in a tested individual. If an individual is symptomatic and the test is being used for diagnostic purposes, the probability of actually having underlying disease (if the test is positive) or not disease (if the disease is negative) is heavily dependent upon prevalence. Although this is the most clinically informative measure of an investigative medical service, these measures are not generally useful for systematic reviews of investigative medical services, because they are dependent on the prevalence of the target condition in the study population, and thus cannot readily be transferred to different populations or pooled to produce a summary estimate. An estimate of the prevalence of the target condition or clinical information of interest is based on data available for the target population or a systematic review of prevalence studies.

In the field of genetics, clinical validity refers to a test's ability to detect or predict the clinical disorder or phenotype associated with the genotype, the four most relevant measures being **clinical sensitivity/clinical specificity** and **clinical positive/negative predictive values**. The approach to the literature search outlined in Section B3 will also apply for clinical sensitivity and clinical specificity. It should be pointed out that the distinction between analytical sensitivity/specificity vs clinical sensitivity/specificity is unique to genetic testing. For other investigative medical services that are considered by MSAC this two layered distinction is not made in the assessment of a sensitivity and specificity.

In relation to clinical predictive values, this measure has slightly different meaning depending on which population is being considered be it the symptomatic individuals versus the asymptomatic first degree biological relative of an individual with a known genetic mutation. In this later population the positive predictive value is the risk of developing disease over a specified time period if the test result is positive and the negative predictive value is the probability of not developing disease over a specified time period if the test result is negative. This application of the concept of predictive values is distinctly different to the ‘symptomatic/diagnostic scenario and for the asymptomatic individual who is currently disease free, the calculated predictive value is essentially a reiteration of the concept of absolute risk and provides prognostic information in terms of both informing a baseline (pre-intervention) risk of developing the condition or outcome of interest over a nominated time period which would then feed into a discussion in clinical utility (Section B5) of the likely absolute benefit of receiving subsequent therapeutic interventions (which is determined by a combination of baseline risk and the proportional benefit conferred by receiving subsequent therapeutic interventions).

B4.2 Supplementary analysis if the service is to be used to inform prognosis or predisposition

INFORMATION REQUESTS

- State whether the information generated as a result of providing the investigative medical service under consideration is of prognostic value or generates information about predisposition. Present summary of key evidence supporting this.

An investigative medical service might, in addition to being used for diagnostic purposes may additionally generate prognostic information about overall health outcomes irrespective of whether therapeutic medical services are offered. However, the prognostic impact of the information generated by the investigative medical service can potentially allow better targeting of a broader range of existing treatment options, or to provide other clinical information of interest. The question of prognosis (and also predisposition) is one of ‘causality’. Therefore in terms of a hierarchy of evidence, a systematic review of prospective cohort studies is considered more persuasive evidence than, for example, case-control studies or cross-sectional studies (although in some instances, case-control studies are the only available study type, especially if the health outcome or target condition of interest is rare). The National Health and Medical Research Council (NHMRC) has developed a hierarchy of evidence for prognostic studies which is available on the NHMRC website.

In this instance, provide a summary of the key literature supporting the prognostic value of the information generated by the proposed investigative medical service or literature supporting its use as a predisposition test. To some extent, the structure around presenting such evidence will be left to the discretion of the applicant. However, MSAC will be most influenced by the results of more rigorous prognostic data over less persuasive evidence.

Summarise the key measures of effect generated out of the cited literature (relative risk, etiologic fraction, odds ratio, hazard ratios etc). These findings will provide a baseline for the analysis conducted in Section B5 on clinical utility for those investigative medical services for which these measures would be referenced against as a ‘baseline’ if subsequent treatment were to be offered.

B5 Clinical utility

INFORMATION REQUEST

- If a patient was identified as having the target condition or clinical information of interest (regardless of whether they were correctly identified), determine whether this translates to a net change in clinical management and present key evidence supporting this.
- Summarise the relative impact and clinical importance of false negatives and false positives
- If evidence on health outcomes for each therapeutic medical service option has been identified and that the evidence supporting these subsequent therapeutic options can be linked, then present a summary of key findings here.
- If a subsequent therapeutic option is linked to an investigative medical service through co-dependence because the investigative medical service predicts treatment effect modification, go to Appendix 7.

Clinical utility refers to how likely the test is to significantly impact on patient management and health outcomes. This section primarily focuses on demonstrating the clinical utility of the proposed investigative medical service relative to the main comparator in terms of the net change (if any) in clinical management and health outcomes. However a scenario may arise where regardless of how an investigative service performs diagnostically (be it the proposed service or main comparator), there may be no clinical utility, in an absolute sense, and thus no clinical justification in performing either service. Thus a brief statement about the clinical utility more broadly needs to be made upfront in the Executive Summary in the assessment report to provide a baseline for an analysis of clinical utility to be presented in this section of the assessment report.

B5.1 Impact on clinical management (therapeutic efficacy)

Presenting evidence on the impact on clinical management of investigative medical services involves the following:

- Separately outline the broad impact on clinical management if a **positive** result is generated by the investigative medical service versus the impact on clinical management. If a **negative** result is generated (regardless of whether the result is truly correct or not) – link this in with Sub-section B5.2.
- Will knowledge of the result (be it a positive result or negative result) generated by the investigative medical service (regardless of whether the result is truly correct or not) cause a **change in the management** of the patient by the treating clinician? Are there instances where clinical management would not change?
- The relative consequences of incorrect test results (**false negatives** versus **false positives**) across the ‘proposed arm’ and ‘comparator arm’ (depending upon the point estimates for comparative accuracy outlined in Section B3) including providing rationale why the assessment should provide more weight on the clinical consequences arising from a either false negative or false positive result.

As an investigative medical service will often be used to guide decisions around therapeutic medical services, this connection would need to be explicitly addressed. Once listed, these issues could be informed by data that compare the numbers of ‘positive’ results generated by the investigative medical service and the number of therapeutic medical services actually provided. Additional positive results detected by the proposed investigative medical service can lead to changes in management by identifying patients with more advanced disease than suspected (through the main comparator) in whom the instigation of more aggressive treatment might be appropriate. Conversely, where the proposed investigative medical service indicates more limited disease than suspected (through the main comparator), the addition of the proposed investigative medical service can potentially lead to less aggressive treatment appropriate to patients with more limited disease, thereby potentially improving patient outcomes by avoiding morbidity associated with more aggressive treatment.

It should be pointed out that there might be ‘leakage’ issues identified through an evaluation of the ‘change in management’ part of a linked evidence analysis. Often, an investigative medical service is done to rule out the use of a therapeutic medical service, but the therapeutic medical service is provided to the patient anyway. Or, alternatively, the investigative medical service is used to select a specific therapeutic medical service, but the therapeutic medical service is not provided to the patient. Also discuss situations in which clinical management would not change, despite the investigative medical service indicating the presence or absence of the target condition under consideration. That is, where there is no effective alternative therapeutic medical service (i.e. not last-line versus last-line) and thus no loss of alternative source of potential effectiveness.

Assigning the correct test result is essential, if the knowledge of the result generated by the investigative medical service can potentially change clinical management. That said it must not always be assumed that correct test assignment (true positives and true negatives) is without any downside especially in a clinical scenario in which the subsequent treatment offered as a result of correct test assignment has little to no clinical benefit but carries with it potential for harm (see Subsection B5.2 and Sections B.7) Also provide a summary of the relative consequences of false negative and false positive results across the ‘proposed arm’ and ‘comparator arm’ including providing rationale why the assessment should have more weight on the clinical consequences arising from either false negatives or false positives. Incorrect test assignment can compound clinical problems created by the complexity of disease and potentially exposes the patient to an inferior health outcome or denies consideration of an alternative therapeutic medical service. In some cases incorrect test assignment of treatment based on a wrong test result can potentially impact health outcomes (shorten progression-free survival). For example, a false negative result might result in a patient being denied a superior therapeutic medical service. The consequences of incorrect test assignment vary according to target condition or clinical information of interest, and stage or severity of disease. This is where consideration of trade-offs between sensitivity and specificity becomes important (Section B3). Where there is no effective alternative therapeutic medical service, a false negative may likely outweigh a false positive in terms of clinical importance. Where there are effective alternative therapeutic medical services, a false positive may likely outweigh a false negative in terms of clinical importance in informing an overall conclusion in Section B8.

In addition, assess whether there is a **threshold** below which either the proposed service or main comparator should not be used (eg the false positives are too great or the false negatives are too great). The greatest value of an investigative medical service lies at the threshold between normal and abnormal, or between disease A or B (eg an investigative medical service that detects florid disease might not be much use, while an investigative medical service that correctly identifies subtle disease might be potentially more useful).

If overall evidence of a net change in clinical management does exist, then present key findings in this Sub-section.

B5.2 Therapeutic effectiveness (including impact of effect modification)

For each therapeutic medical service option for which there is evidence of health outcomes, it is important to present the key findings of this evidence (provided it is clinically meaningful to link this evidence). Rather than go down an exhaustive approach as described in the *Therapeutic Guidelines* (Part II, Section B) for each therapeutic medical service option, it is recommended that applicants present a summary of the body of evidence supporting each option (including point estimates and accompanying confidence intervals for the magnitude of benefit of each treatment option). To some extent, this will be left to the discretion of the applicant. However, it is recommended that the *Therapeutic Guidelines* (Part II, Section B) is used as a guide to identifying the most rigorous source of data because MSAC will be most influenced by the results of such data as opposed to less persuasive evidence. Identify whether the health outcomes measured were surrogate outcomes or definitive health outcomes. Should evidence exist that either the timing of performing the test and thus timing of receiving treatment (for example earlier versus late treatment) is one of the impacts on health outcomes of performing the investigative medical service, then present a summary of the key evidence supporting this in this section.

If an investigative medical service generates information in terms of identifying and predicting which patients will respond to treatment differently (better or worse) compared to other patients, summarise the evidence supporting this. This is otherwise known as treatment effect modification. In studies where effect modification has been demonstrated, in addition to a 2 X 2 table outlining the summary measure of association between the relevant 'study factor' and 'outcome factor' that is being assessed, an additional stratified analysis is conducted analysing separately in strata of those with the 'other variable' of interest (for example a particular biomarker) that is a potential effect modifier and those who do not which generates two new 2 X 2 tables, one for each stratified group. When the measure of association is calculated for each stratified group, if a major difference between the two measure of associations is calculated for the two stratified groups the variable of interest is regarded as an effect modifier. Effect modification essentially occurs when the magnitude of the chosen measure of association between the study factor and the "outcome factor" differs according to the level of the "other variable" (called *statistical interaction*). This could be something inherent or intrinsic within individuals or the environment that allows this to occur. The most common effect modification scenario presented to MSAC is the biomarker-drug scenario i.e. how well a patient responds to a particular drug may vary depending upon the presence or absence of the biomarker. In this scenario of effect modification the investigative medical service is paired with a **co-dependent** therapeutic service. In this instance applicants do not need to submit a full clinical analysis of the therapeutic evidence of health outcomes if this to be submitted in full in a paired application across MSAC and PBAC. However, to link the clinical conclusion for the investigative medical service, to an economic analysis, some core information about the co-dependent therapy is requested. This is outlined in Appendix 7.

B6 Impact of repeat testing/use of service as part of a monitoring strategy

INFORMATION REQUESTS

- For applications for which this section is relevant, provide evidence supporting the repeat use of the proposed investigative medical service as part of a monitoring strategy against four criteria for assessing the merits of using a service for monitoring (a) reiteration of clinical validity, (b) responsiveness, (c) detectability of long term change and (d) practicality.

Monitoring and repeat testing is an important element of a patient's longer term care. Monitoring in clinical practice occurs in three main phases depending upon the theoretical trajectory overtime: before treatment, response to treatment, and long-term monitoring⁴. Four important criteria may be used to choose the best investigative medical service for monitoring a patient in each of these phases (a) clinical validity, (b) responsiveness, (c) detectability of long term change, and (d) practicality. These criteria are to be considered within the context of the risk factor-disease-treatment pathway that leads to the initiation of treatment to the outcome that treatment aims to produce and also informs the appropriate time interval between conducting the service and whether this time period is constant or varies (for example, a short cluster of repeat testing followed by longer period of no testing).

Clinical validity describes the ability of the investigative service to predict the clinically relevant outcome of interest (see Section B4). It must be on the risk factor / outcome pathway, either directly or as a proxy for another marker on the pathway that is less easily measured. The later the service is on the pathway (closer to outcome) the more predictive it is likely to be, with the most valid investigative service being a measure of the outcome itself.

Responsiveness describes how much the investigative service changes in response to a therapeutic intervention relative to background random variation. 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 investigative service 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 an investigative service should show 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 drugs (monitor glucose) or bleeding risk for patients on warfarin (monitor international normalized ratio). In other situations in which the change in outcome is much slower, it is still preferred that the investigative service response can be quickly judged whether treatment is working as expected, for example, risk of a cardiovascular event (monitor cholesterol and blood pressure). Not all responsive investigative services show rapid changes in response to treatment; in fact, some take months/years to change, for example, HbA1c. Because changes in the results of the investigative services reflect average treatment effects over a longer period of time, these services may be preferred for judging effects over the medium to long term.

⁴ Bell KJL, Glasziou PP, Hayen A, Irwig L. Criteria for monitoring tests were described: validity, responsiveness, detectability of long term change and practicality. *Journal of Clinical Epidemiology*; 67: 152-159. 2014

The concepts of “**signal**” and “**noise**” are relevant to both response monitoring and long-term monitoring. For response monitoring, signal includes both mean change and between-person variation in response. If the between-person variation component of the signal is small, then 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 services used for monitoring may not be appreciated by clinicians, and variations due to noise may be wrongly attributed to real change.

Detectability of long-term change describes the size of changes in the results of an investigative medical service over the long term relative to background random variation. Long-term change describes the ability of the service 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. Noise is the same as for response monitoring: short-term random variation in level within an individual. Unlike response monitoring in which the between-person variation component of the signal is often small, rendering 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 the need to also estimate the individual’s true deviation from the mean change under conditions of a favourable signal-to-noise ratio.

Finally the **practicality** of the investigative medical service as a monitoring tool describes its ease of use, level of invasiveness and cost (relevant later in this document).

Assessing the performance of an investigative medical service as part of a monitoring strategy generally requires longitudinal data prospectively generated from trial and/or cohort studies. Present in Section B6 of the assessment report, the relevant data that addresses the above principles to inform whether the proposed investigative medical service is justified to be used as part of a monitoring strategy.

If an investigative medical service that is being considered by MSAC is being proposed to be used as part of a monitoring strategy that informs the use of a pharmaceutical concurrently seeking PBS listing through PBAC (and thus the pharmaceutical is co-dependent on the investigative medical service) then a paired application across both committees applies (see Appendix 7).

B7 Extended assessment of comparative harms

INFORMATION REQUESTS

- State whether there is any evidence of direct harm as a result of the proposed investigative medical service (immediate or delayed) as well outline flow on safety consequences.
- Specify and justify the search strategy used to identify suitable sources of evidence.
- Succinctly present any such evidence identified, with appropriate cross-referencing to any source documents provided in a technical document or attachment to the assessment report.
- Provide appropriate cross-referencing to any source documents provided in a technical document or attachment to the assessment report.
- Indicate how the harm profile compares with that of the main comparator.

Consideration of the relative safety of performing the investigative service versus the main comparator requires the:

- presentation of immediate or delayed safety consequences of physically performing the investigative medical service; and
- presentation of ‘flow on’ safety consequences that arise as a result of conducting the investigative medical service.

Accuracy studies are often not a complete source of data on comparative harms that are a direct consequence of physically conducting the investigative medical service. Thus, a wider basis of assessment of comparative harms from other sources (i.e. beyond the results of accuracy studies) is required. This wide assessment is especially important for serious adverse reactions that might be delayed (e.g. delayed bleeding following an investigative procedure such as a liver biopsy). Specify and justify the search strategy used to identify suitable sources of information about any such reactions. Harms directly resulting from an investigative medical service of more catastrophic nature (i.e. significant life-threatening complications) are likely to have more weighting in informing comparative harms as a consequence of directly performing the investigative service.

In regards to ‘flow on’ safety issues, separately summarise the safety considerations as a consequence of each of the four outputs of test result that is true positives, false positives, true negatives and false negatives (refer Section B5.1).

Come to an overall conclusion on comparative safety of the proposed investigative medical service versus the main comparator.

B8 Overall interpretation of all the clinical evidence presented

INFORMATION REQUESTS

- Provide a summary assessment of the evidence presented in Section B for the overall clinical claim.
- Use this assessment to state the category from Table B8.1 that best reflects the therapeutic conclusion of the proposed (index) investigative medical service over its main comparator, supported by the evidence presented.

It is proposed that an overall conclusion be drawn about whether the proposed investigative medical service is non-inferior (no worse than) or superior compared to its main comparator in terms of therapeutic performance (both safety and clinical effectiveness). Include in this assessment of the evidence the main findings of:

- Comparative direct evidence of clinical impact on health outcomes;
- In the absence of direct evidence, the main findings of the linked evidence approach including:
 - relative diagnostic performance (and clinical validity where relevant) of the proposed investigative service versus the main comparator;
 - proportion of patients in both the proposed arm' and 'comparator arm' respectively that are assigned as true positives, false positives, true negatives and false negatives based on the point estimates of comparative accuracy outlined in Section B.3
 - impact on the clinical management and health outcomes (clinical effectiveness) across the 'proposed arm' and comparator arm'; and
 - impact on safety across the 'proposed arm' and the 'comparator arm' including the relative weight given to the clinical importance of safety (depending on the severity and nature of harms) versus clinical effectiveness
 - combining both evidence on clinical effectiveness and safety to come to an overall conclusion in terms of the relative benefit/harm ratio across the 'proposed arm' and 'comparator arm'. Include how repeat testing impacts on this benefit/harm ratio.

The interpretation of the clinical data presented in assessment report Section B is crucial in determining the success of the assessment report. Table B8.1 sets out a framework for this classification.

Table B8.1 Classification of the therapeutic relativity of the proposed investigative medical service over its main comparator

Comparative safety	Comparative clinical effectiveness (generated by either direct evidence or linked evidence approach)			
	Inferior	Uncertain	Non-inferior	Superior
Inferior	Health forgone: need other supportive factors	Health forgone possible: need other supportive factors	Health foregone: need other supportive factors	? Likely Linked evidence approach ^c to inform subsequent incremental cost-effectiveness analysis
Uncertain	Health forgone possible: need other supportive factors	?	?	? Likely Linked evidence approach to inform subsequent incremental cost-effectiveness analysis
Non-inferior	Health forgone: need other supportive factors	?	Cost-minimisation analysis	Linked evidence approach to inform subsequent incremental cost-effectiveness analysis
Superior	Linked evidence approach to inform subsequent incremental cost-effectiveness analysis	? Likely Linked evidence approach to inform subsequent incremental cost-effectiveness analysis	Linked evidence approach to inform subsequent incremental cost-effectiveness analysis	Linked evidence approach to inform subsequent incremental cost-effectiveness analysis

? = reflect uncertainties and any identified health trade-offs in the economic evaluation, as a minimum in a cost-consequences analysis

The advantage of distinguishing between non-inferiority and superiority is that non-inferiority can result in a simpler economic evaluation of cost-minimisation (see Part III, Section D(i)). On the other hand a conclusion of superiority requires a more complex economic evaluation to be conducted (see Part II Section D)

At the conclusion of Section B proceed to Section C which outlines how to translate the outcome of the clinical evaluation in Section B into an economic evaluation in Section D.

Section C

Translation Issues

Introduction

The primary purpose of Section C in the assessment report is to guide the presentation of analyses conducted to translate the systematic overview of the results of clinical evidence to the listing requested, and thus to the framework of the economic evaluation (Section D) of the assessment report. This is particularly important when one or more variables incorporated into the economic evaluation are derived from, but not directly based on, the clinical evaluation presented in assessment report Section B. These variables might be derived using a number of analyses that modify the results of the clinical evaluation to help construct a modelled economic evaluation. Such analyses are referred to in these Guidelines as ‘pre-modelling studies’.

The need for pre-modelling studies arises because the study protocols for the studies used for the clinical evaluation might differ from the proposed clinical practice setting for the main indication in one of the following ways:

- The participants and circumstances of use in the studies presented in the evidence base in Section B might not be the same as the intended population for treatment in Australia (and might therefore have a different profile of risks of future events and circumstances of use). In this case, the clinical evaluation would need to be applied from the baseline risk of the sample of study participants and their circumstances of use to the expected absolute risks of future events of the intended Australian population and their circumstances of use. Examples of pre-modelling studies of **applicability** include subgroup analyses and surveys of the patterns of health care resource provision in Australia corresponding to one or more health states included in a modelled economic evaluation.
- The length of follow-up (time horizon) of participants in the studies presented in the evidence base in Section B might be less than the expected duration of therapy or expected duration of overall health and health care resource impacts. In this case, the clinical evaluation would need to be extrapolated to the intended duration of therapy or expected health and resource impacts. Examples of pre-modelling studies of **extrapolation** include extrapolating integrals of time-to-event analyses and a review of the literature for single-arm follow-up studies of the natural history of the condition to estimate rates of disease progression.
- The outcomes measured in the studies presented in the evidence base in Section B might not be the patient-relevant final outcomes of treatment. In this case, the clinical evaluation would need to be transformed to take account of the patient-relevant final outcomes (in terms of QALYs gained). Examples of pre-modelling studies of **transformation** include transforming comparative treatment effects measured on surrogate outcomes to final outcomes and scenario-based studies to value health outcomes using utilities.

Thus, the results of the studies presented in Section B might need to be applied, extrapolated and transformed (collectively referred to in these Guidelines as ‘translated’) into a decision analysis appropriate for the intended clinical use of the proposed investigative medical service if publicly funded in Australia, taking into account the

above issues. These pre-modelling studies provide a clearer and more systematic basis to support the necessary variables for inclusion in the economic evaluation (see Section D). As indicated by the examples above, the types of pre-modelling studies relevant to this process of translation can vary widely.

The methods of translation are described in Sub-section C2. The methods also help examine any impact of reintroducing sources of random error (the play of chance) and systematic error (bias), which were minimised in the systematic overview of the studies presented in Section B and to be presented in Section B of the assessment report.

The results of pre-modelling studies are intended to inform:

- the underlying structure of the model and the selection of options for examination in an analysis of the structure of the model, and the scenarios it is examining; and
- the selection of values for variables in the economic evaluation, and ranges of plausible extremes to include in the associated sensitivity analyses.

Importantly, Section C requests a consistent format for the presentation of all pre-modelling studies. Each presentation has the following components:

- a succinct question to address a particular issue (Sub-section C1);
- a focused analytical plan that is presented and justified (Sub-section C2);
- a set of results (Sub-section C3); and
- an explanation of how these results contribute to the economic evaluation presented in Section D of the assessment report (Sub-section C4).

Presentation of Section C in the assessment report would be helped by listing the issues to be addressed in pre-modelling studies in a single response to Sub-section C1, preferably with a concluding tabulated summary. Then present the pre-modelling studies sequentially in a series of Sub-section C2 and C3 pairs (i.e. the focused analytical plan in response to Sub-section C2 requests and the results in response to Sub-section C3 requests). A single response to Sub-section C4 should then summarise the main results of the pre-modelling studies together and indicate how their results are to be used in the economic evaluation presented in Section D of the assessment report.

C1 Identification of issues to be addressed

INFORMATION REQUEST

- Define application issues: Describe any ways in which the participants and circumstances of use in the studies presented in Section B differ from the proposed population for treatment (including the baseline risk of participants and circumstances of use).
- Define extrapolation issues: State whether there is a need to extrapolate the outcomes reported in the clinical evaluation beyond the study horizon.
- Define transformation issues: State whether there is a need to transform the nature of the outcomes measured in the clinical evaluation (i.e. taking a surrogate or intermediate endpoint, and transforming it to a QALY or equivalent).

- Define any other translation issues: State whether there is any other need to translate from the clinical evaluation.
- Convert each defined translation issue into a succinct question that can be addressed in a pre-modelling study.

The following guidance is intended to help an applicant decide whether additional analyses are needed and to identify methodological options that might be considered. It is recognised that not all the necessary information will be available to inform every aspect of each circumstance and the resulting analyses. Methodological experts might also disagree about the most appropriate methodological option to pursue in particular circumstances. However, this detailed guidance is warranted because many assessment reports have had difficulties in this area.

The issues identified in response to Sub-section C1 should focus on those for which pre-modelling studies are presented in Section C of the assessment report. At the end of the response to this Sub-section C1, tabulate a summary list of these material translation issues in the order identified. Separately tabulate a summary list of any other translation issues identified, but for which pre-modelling studies are not presented. In each case, summarise in the table why a pre-modelling study is not presented (e.g. not expected to make a material difference).

Applicability issues

Define any issues that indicate a need to apply the study data to the intended population and circumstances of use. Applicability issues might arise due to differences between participants enrolled in the trials and patients who would be likely to obtain the proposed investigative medical service if publicly funded in Australia, and between the circumstances of use in the studies presented in Section B and those that would occur if publicly funded in Australia.

There are a number of important patient factors that might affect outcomes. There might also be important differences in the mix of patients who would receive the proposed investigative medical service if publicly funded in Australia. For example, it is a concern of MSAC that there might be patients in the community who have a disease that is less severe than that of participants in the studies presented in Section B. There might also be patients in the community for whom the main comparator can be expected to perform better than in the studies presented. Both could diminish the difference in effectiveness between the proposed investigative medical service and the main comparator, and therefore make the incremental cost-effectiveness ratio less favourable for the proposed investigative medical service.

There are also important factors relating to the circumstances of use. These factors might also include extrapolating results of clinical studies conducted in hospitals to use outside the hospital and the effect of more rigorous follow-up, which might swamp important differences in the convenience and acceptability of the proposed investigative medical service compared with alternate services. This might have resulting effects on patient compliance and subsequent response to flow on treatments.

The fact that one or more differences might be demonstrated does not necessarily raise an applicability issue, because the differences might not help to predict any variation in treatment effect. However, the demonstration of such differences does identify areas that could be examined, such as in the examples given in the following Sub-sections.

Population characteristics

There might be evidence within the studies presented in Section B and/or other sources to indicate that patients vary in their expected risk of adverse major clinical outcomes. In such cases, which are common for many medical conditions, additional analysis of the comparative treatment effect detected in studies in Section B, presented as a pre-modelling study, might indicate that this effect is best summarised as a constant relative reduction in the risk of these outcomes across the study population of varying baseline (expected) risks.

If this is the case, such an analysis forms an acceptable basis to apply the study data to specific subgroups. For example, this evidence would be sufficient to justify targeting a requested restriction to those patients with a greater expected absolute risk of future events at the point of deciding whether to start a therapeutic medical service (i.e. a poorer prognosis) as being the patients likely to benefit most from a therapeutic medical service flowing on from the conduction of the investigative medical service. Any thresholds of greater expected absolute risk to identify the population that would be eligible to start a therapeutic medical service would need to be justified and supplemented by sensitivity analyses on different thresholds. The absolute or incremental treatment effect would then be calculated by multiplying the expected absolute risks across the eligible population by the estimated overall relative treatment effect. As a check, present the results of the targeted subgroup that might be recruited in the therapeutic studies presented in Section B5 as the absolute risk difference, or explain why this is not possible.

The comparative treatment effect detected in the studies presented in Section B5 might indicate that this effect is best summarised as a varying relative reduction in the risk of these outcomes across the study population of varying baseline risks. In this case, which is less common than the previous example, the pre-modelling analysis would need to identify treatment effect variation when measured in relative terms (e.g. relative risk, hazard ratio, odds ratio). This analysis of the relative treatment effect would need to show sufficient heterogeneity within the clinical evidence available to support statistically a claim regarding the nature (qualitative or quantitative) and extent of each treatment effect variation, and thus any resulting subgroup analysis.

Variations in the relative treatment effect might arise with varying characteristics of the patient, the relevant flow on therapeutic medical service(s) or the medical condition. Together with a justification of any thresholds as necessary (supplemented by sensitivity analysis on different thresholds), this evidence contributes to an argument to target a requested restriction to these patients (see Section A) and to calculate the absolute treatment effect by applying the estimated relative treatment effect for the subgroup to the expected risk for the subgroup.

Circumstances of use

- One or more of the studies presented in Section B might include flow on therapeutic medical services that are not recommended or approved by the TGA, or that might otherwise have an impact on the direction and/or magnitude of the treatment effect.

- One or more of the studies in relation to the proposed investigative medical service and one or studies in relation to flow on therapeutic medical services (as presented in Section B) might have been conducted in settings that are not applicable to the requested listing on the MBS or with some study participants who would not be eligible for the proposed investigative medical service according to the requested restriction (or who would not be eligible for the flow on therapeutic medical service according to a requested or existing restriction)
- One or more of the studies presented in Section B in relation to the proposed investigative medical service and studies in relation to flow on therapeutic medical service(s) might have delivered the proposed investigative medical service and flow on therapeutic medical service(s) in a way that differs from how it would be delivered if publicly funded in Australia.

There is no limit to the types of difference in populations and circumstances of use, but only a small number of these might modify the extent of treatment effect detected by the overall results of the studies presented in Section B. Thus, the general rule is to apply the overall treatment effect from the intention-to-treat (ITT) population, rather than to explore for possible variations in treatment effects in subgroups.

As discussed in Sub-section C2, an analysis to support a claim of treatment effect variation according to a particular patient characteristic or circumstance of use is more convincing if it was pre-specified with a biologically plausible rationale before the collection of any data in the trial(s) providing the source data for the analytical plan. Thus, for each analytical plan relying on and examining an applicability issue, state whether the data were collected before or after finalisation of the analytical plan (see below).

If an applicability issue involves introducing one or more diagnostic criteria or tests specifically to identify patients who are eligible according to the requested restriction that was not relied on in the studies presented in Section B, then separately present additional information on the accuracy (specificity, sensitivity), reliability and comparability of these criteria and tests, both across all studies presented and in regular Australian practice. This is necessary to examine the impact of false positive and false negative identification of eligible patients, as well as the impact of false positive and false negative identification of treatment response, on the application of the trial results. Section D4 provide further advice on specifying and costing these diagnostic criteria and tests in the diagnostic and treatment algorithm, and on the implications of misclassification for estimating incremental effectiveness and incremental cost-effectiveness.

If there is no applicability issue, state this.

Extrapolation issues

Define any issues that indicate a need to extrapolate the within-study patterns of resource provision (cost) and within-study health outcome results, including time-to-event data, beyond the time horizon of the studies presented in Section B. Such extrapolation might be considered necessary in the context of a modelled economic evaluation, to determine comparative effectiveness and cost-effectiveness beyond the median duration of therapy and/or follow-up in the presented evidence in Section B.

If there is no need to extrapolate the evidence from the clinical evaluation, state this.

Transformation issues

Define any issues with outcomes that indicate a need to transform the nature of the outcome(s) measured in the studies presented in Section B to those relied on in the economic evaluation. For example, some studies in Section B might only report outcomes that are of less patient relevance than intended final outcomes of treatment. These less relevant outcomes are known as *surrogate outcomes*. Arguably, the closer a surrogate outcome is to the final outcome, the more useful it is, but generally the more difficult it is to measure accurately.

To transform the surrogate outcomes measured in the evidence presented in Section B to final outcomes and to extend the range of outcomes (e.g. the number of patients with unhealed peptic ulcers who eventually need surgery), the trial results might need to be supplemented by estimates obtained from other sources.

For most proposed investigative medical services, the ultimate outcome is to improve quality of life and/or survival (be it indirectly through a therapeutic medical service. In theory, all outcomes could be expressed as QALYs gained (see Appendix 4). In practice, few clinical studies have directly measured the impact of proposed investigative medical service on QALYs, because few are large enough or long enough to measure changes in final outcomes directly.

Another common need is to transform the outcome(s) measured in the clinical evaluation to value them in utility terms for the economic evaluation (see Appendix 4 for more information on utility terms). If this transformation supplements any other transformation, for example, from surrogate outcomes measured in the clinical evidence to patient-relevant outcomes, present the links between these two transformations and any assumptions involved in combining them.

Other transformations that have been considered include:

- converting outcomes reported as continuous data to dichotomous data; and
- converting outcomes reported as dichotomous data to time-to-event data to estimate periods of time in one or more health states, or periods of time free from being in one or more health states.

Although these transformations increase uncertainty, they can allow for a more readily interpretable health outcome (see Sub-section C2).

If there is no need to transform the outcomes measured in the clinical evidence in Section B, state this.

Other translation issues

Define any other issues that required pre-modelling studies to justify an aspect of the economic evaluation (see Section D). Examples of other issues that might be included here, are as follows:

- one or more of the studies presented in Section B was less successful in minimising bias (e.g. inadequate concealment of randomisation, inadequate blinding of subjective outcomes, unable to reconstruct full ITT analysis);
- one or more of the studies presented in Section B reported fewer patient-relevant outcomes or no patient-relevant outcomes;

- one or more of the studies presented in Section B was of insufficient duration to detect the most patient-relevant outcomes; and / or
- the patterns of resource provision measured in the studies presented in Section B did not closely reflect those in Australia (and/or the likely changes in patterns of resource provision were not measured in the trials).

Clinical evidence generated overseas are an acceptable basis for an economic evaluation relevant to Australian practice. However, although the overall estimate of the change in a final or surrogate outcome might be transferable to Australia, estimates of the costs of resources provided (such as further investigations, procedures or operations) are often not readily transferable, as:

- it is usually apparent that the unit costs are quite different;
- less apparent, but also important, is the fact that the frequency or patterns of resource provision might not be relevant to Australia because of major differences in medical practice or different incentives in different economies and health care systems;
- sometimes assumptions need to be made during the translation of overseas clinical evidence to create a modelled economic evaluation that is relevant to the Australian context;
- the clinical evidence did not measure provision of all types of health care relevant resources (which might change and therefore would need to be added in a model); and / or
- the protocols of the primary clinical studies outlined in Section B required more resources to be provided than would be typical in normal management of the medical condition (such as extra monitoring to demonstrate safety or effectiveness). In this case, only resources provided or avoided in regular clinical practice need to be included in a model.

If there are no other issues that require pre-modelling, state this.

C2 Focused analytical plan

INFORMATION REQUEST

- Present and justify a focused analytical plan (which specifies details of data, sources, methods and analyses) to address each translation issue identified.

For each translation issue identified in Sub-section C1, provide a focused analytical plan that clearly describes the:

- issue;
- specific question to be addressed by a pre-modelling study;
- data to be used and their sources; and
- methods of the pre-modelling study (with sufficient details to enable independent verification of the analysis).

A range of methods that might inform the development of an analytical plan are shown below. Justify the choice of method where more than one option exists. Comment on any implications of this choice for the results of the pre-modelling study, including how the choice of the method will be assessed; for example, in the sensitivity analyses of the economic evaluation.

Methods to address applicability issues

Addressing applicability issues might involve investigations of heterogeneity, variation in treatment effect, subgroup analysis and/or meta-regression.

Heterogeneity analysis

Discuss and explain any suggested heterogeneity of study results. This is particularly relevant if a meta analysis is presented in Section B. Reasons for heterogeneity might include differences in study population or design. If there are strong biological or methodological grounds for heterogeneity, consider presenting a pre-modelling study to examine the impact of these grounds for heterogeneity by comparing relevant pooled analyses with the overall estimate. Unexplained heterogeneity, depending on its direction and magnitude, generally makes the summary estimator less meaningful.

Assessment of heterogeneity is an important aspect of interpreting meta-analyses where there are a large number of studies. Refer to biological or clinical reasoning as appropriate when justifying the inclusion of further analyses in pre-modelling studies to take into account heterogeneity when considering the application of the results of the trials.

Explain and justify the presentation of any additional meta-analyses in which studies are excluded (e.g. on the grounds of inadequately minimising bias or of reporting fewer patient-relevant outcomes) and examine the impact each exclusion has on the overall meta-analysis. Similarly, explain and justify the presentation of any additional meta-analyses in which study groups are excluded and examine the impact each exclusion has on the overall meta-analysis.

Support any claimed treatment effect variation on the basis of observed heterogeneity with reference to the excluded studies and/or study groups, and the covariate that predicts the treatment effect variation, such as:

- varying duration of use;
- settings of use; and
- patient baseline characteristics, including risk factors and disease severity.

If any heterogeneity is thought to be due to the studies presented in Section B having different periods of follow-up, presenting the pooled incidence rate differences might be useful.

Assessment of possible publication bias, where there are sufficient trials, might be assisted by presentation of a funnel plot.

Methods to address extrapolation issues

Extrapolating time-to-event data

Several different methods might be used to extrapolate time-to-event data, and a range of assumptions need to be tested in an extrapolation of survival or time-to-event data beyond the horizon of the studies presented in Section B (be it direct evidence or therapeutic evidence that has incorporated in a linked analysis). Justify the assumption (whether made directly or indirectly) in relation to the measure of effect (eg hazard ratio) reflecting the comparative treatment effect beyond the time horizon of the studies presented in Section B.

If the economic evaluation is based on an extrapolation of time-to-event data, also present the within study case (i.e. within the time horizon of the studies presented in Section B) alongside the extrapolation, because this allows an at-a-glance assessment of the extent to which the incremental gains arise within the time horizon of the trial compared with the extrapolated time horizon. Similarly, if the proposed approach to extrapolating the time-to-event results does not result in a convergence of the two extrapolated curves, present an analysis that incorporates a linear triangulation from each of the observed curves at the point of median duration of follow-up to a single common maximum end point justified as being clinically plausible. Another method to converge these curves would be to project the curve representing the outcome with the main comparator beyond the median duration of the follow-up within studies contained in either Section B, and apply a hazard ratio of one to estimate the projection of the curve representing the outcome with the proposed therapeutic medical service from this time point. Particular justification would be needed to apply a hazard ratio representing a continued differential treatment effect beyond the median duration of the key studies presented in Section B.

Methods to address transformation issues

Use of surrogate outcomes to estimate final outcomes

The claim that an incremental treatment effect on a surrogate outcome measured in respect to the therapeutic medical services (that flows on from the proposed investigative medical service) quantitatively predicts a subsequent incremental treatment effect on a final outcome is more persuasively shown if attention is given to the following issues:

- **Step 1** - Present a systematic review of the literature to examine whether epidemiological evidence and biological reasoning has established that there is a relationship between the surrogate outcome and the final outcome independent of any intervention. In a few instances, relationships have been established, or have been proposed, between surrogate outcomes and final outcomes. Examples include blood left ventricular ejection fraction and survival after myocardial infarction, or viral load and cure of viral hepatitis.
- **Step 2** - Present a systematic review of key literature to examine whether direct evidence using other active medical services has shown that there is a basis to conclude that a treatment effect on the surrogate outcome has satisfactorily predicted a treatment effect on the final outcome. Based on this evidence, quantify the relationship between these treatment effects with an assessment of the uncertainty of the relationship. Discuss the reproducibility of these.

- **Step 3** - Explain why this relationship between the treatment effects on these outcomes with these other active medical services is likely to apply to the therapeutic medical services that flow on from the proposed investigative medical service. At present, it is difficult to give categorical advice. Consider which outcomes are most appropriate and most feasible, given the data available. The clinical importance and patient relevance of the outcomes should be established and, where possible, supported with data.

Having addressed the three steps above in transforming a treatment effect on a surrogate outcome to a treatment effect on a final outcome, explain in response to Sub-section D4 how this is included in the economic evaluation, including by specifying and referencing the sources of the longer term natural history (e.g. longitudinal population studies) as well as the transformed treatment effects.

Valuing health outcomes

Where the final outcome flowing from the proposed investigative medical service is a change in quality of life (with or without a change in the number of projected life-years gained), a separate utility analysis is appropriate to transform this change into a preference-based measure. Appendix 4 provides further guidance on the presentation of a pre-modelling study to elicit the utility valuations.

Other useful transformations of health outcomes

Outcomes that are expressed as dichotomous outcomes measured on a per patient basis (are easier to interpret and to incorporate into an economic evaluation than a difference in means for a quality-of-life scale or a physiological variable. Further, converting these proportions, as appropriate, to estimate periods of time free of an event, time with an event or time in a health state allows for a more interpretable incremental cost-effectiveness ratio if there is no limit to the duration of therapeutic medical service that flows on from the proposed investigative medical service. Consider providing a technical document or an attachment to the assessment report to give the details of the methods of these transformations.

Methods to address other translation issues

Examination of exclusion of studies from Section B

Examination of the impact of removing studies from clinical evaluation contained in Section B can sometimes suggest explanations for translating the clinical evaluation. If one or more studies are to be excluded from the clinical evaluation contained in Section B, identify the aspect(s) of each study that justify the exclusion (see Table C2.1). Indicate whether each reason relates to an applicability, extrapolation or transformation issue (see above), or whether a translation issue arises because one or more of the studies was less successful in minimising bias, or reported fewer or no patient-relevant outcomes. If there is more than one type of reason for exclusion, arrange the studies for exclusion in Table C2.1 by reason for exclusion. Present each relevant clinical evaluation both with and without the studies excluded. Discuss any implications of the exclusions for the interpretation of the results of the clinical evaluation contained in Section B.

Table C2.1 Reasons to exclude studies

Trial ID	Ground(s) for seeking exclusion	Details ^a
Trial 1		
Etc.		

ID = identification

^a Cross-reference each set of details to the source of information (specifying the trial report with page, table, figure number).

Adjustment of resource provision estimates

A survey of patterns of resource provision in Australia might be needed if resource provision in the studies presented in Section B reflects patterns of resource use that are different from those used and likely to be replaced in Australia (e.g. if they reflect overseas health care systems or the requirements of the study protocol) or were incompletely measured. This survey could be a cross-sectional study observing and recording patterns of resource provision in Australia. An alternative, but less preferred, option could be a survey of Australian expert opinion on the likely patterns of resource provision, either describing overall Australian practice or advising on modifying overseas patterns that are more relevant to Australia (see Appendix 2).

Justify the application of these cross-sectional data into a longitudinal model and consider any possible implicit assumptions. For example, if the clinical response to the therapeutic medical service(s) (that have been initiated as a result of conducting the proposed investigative medical service) involves returning to a less severe health state, the associated patterns of resource provision might not necessarily reflect those of an earlier health state (i.e. before the disease progression meant that the patient became eligible for the therapeutic medical service). As an extreme example of this, applying patterns of resource provision for asymptomatic patients would obviously not be reasonable if those patterns ascertained for patients with watchful waiting at an early stage of an indolent disease were related to patients achieving full symptom control on analgesics at a terminal stage of the same disease.

If any patterns of resource provision from an individual study are to be modified in a model (such as the exclusions of ‘protocol-derived’ resource provision), discuss the extent to which these resources might have affected the results of the trials in terms of health outcomes. This might raise broader applicability issues in terms of changing the circumstances of use.

C3 Results of pre-modelling studies

INFORMATION REQUESTS

- Present the results of each pre-modelling study undertaken to address each translation issue specified in Sub-section C1 (and for which a plan is presented in Sub-section C2).
- Provide:
 - copies of all sources of data in an attachment or a technical document, cross-referenced from the main body of the assessment report; and
 - electronic copies of all computer-based analyses.

Results

Where possible and appropriate, present the results of each analysis for which a plan is presented in Sub-section C2 and estimate the comparative treatment effect as results separately for:

- the proposed investigative medical service;
- its main comparator; and
- the increment with its 95% confidence interval.

Where a scenario-based valuation study has been used to transform the results of studies presented in Section B or any other health state into utility valuations, present these as disaggregated results corresponding to each health state presented as a scenario (see Appendix 4). Also include an estimate of statistical uncertainty around each result.

Discuss the implications of each analysis on the conclusions from the results of the overall clinical evaluation in Section B8. Where a cross-sectional study or expert opinion survey has been used to estimate patterns of resource provision, report that provision, where possible, on a per patient basis and on a per period of time basis.

Clear presentation of pre-modelling studies is expected to increase MSAC's confidence in the economic evaluations that rely on those translations. At all times in pre-modelling studies, it is important to maximise MSAC's confidence (in the primary inference) that substituting the proposed investigative medical service for the main comparator, as proposed, alone causes the differences in the subsequent streams of costs and outcomes. In practical terms, this means that if any stream of costs for a service is to be modified in a model, consideration should be given to any consequential impact on the corresponding stream of outcomes. Similarly, if any stream of outcomes for a service is to be modified in a model, consideration should be given to any impact on the corresponding stream of costs to ensure that the modification is plausible. Discuss these considerations whenever they are applicable to the results of a particular pre-modelling study.

Justify any results to be used in Section D of the assessment report where more than one option exists. Comment on any uncertainties in this selection, including how they will be assessed in the sensitivity analyses of the economic evaluation. Also comment on any combinations of the results of more than one analytical plan in constructing the economic evaluation and any uncertainties arising from those combinations, including how they will be assessed in the sensitivity analyses of the economic evaluation.

Original sources and electronic calculations

Separately provide copies of the original sources of all data (beyond those already presented with assessment report Section B) and reports of studies commissioned for the assessment report in an attachment or technical document. Cross-reference the extraction of data from each source to the level of the page, table or figure number of the source document.

Also, to enable independent verification of each analysis, provide an electronic copy of any computer-based calculations of the analysis (see Part I, Section 5).

C4 Relationship of each pre-modelling study to the economic evaluation

INFORMATION REQUESTS

- Discuss the results of each pre-modelling study and explain how they will be used in the economic evaluation (Section D).
- Provide a summary from Sub-section C3 and their uses in responses to Section D.

Uses of pre-modelling study results

Each pre-modelling study has the objective of providing support for one or more inputs in the economic evaluation. There might be more than one pre-modelling study to support more than one translation step between the overall clinical evaluation and the economic evaluation. When this occurs, the combination of pre-modelling studies might compound the effect of uncertainty. This might need examination in the sensitivity analysis in Sub-section D6.

Section D provides more guidance on how to present the impacts on the economic evaluation of more than one translation step.

Summary table

Table C4.1 provides a suggested format to summarise the main results of each pre-modelling study presented in Section C of the assessment report and their use in the economic evaluation presented in Section D of the assessment report, including in the sensitivity analyses presented in Sub-section D6. This will facilitate cross-referencing across the responses to information requests in the two sections and thus the transparency of the presentation of this information.

Table C4.1 Summary of results of pre-modelling studies and their uses in the economic evaluation

Pre-modelling study	Results	Use in Section D	Cross-reference	Use in Sub-section D6	Cross-reference
Applicability pre-modelling studies					
Study 1					
Etc.					
Extrapolation pre-modelling studies					
Study 2					
Etc.					
Transformation pre-modelling studies					
Study 3					
Etc.					
Other translation pre-modelling studies					
Study 4					
Etc.					

Section D

Economic evaluation for the main indication

Introduction

The purpose of Section D in the assessment report is to present an economic evaluation of substituting the proposed investigative medical service for the main comparator in the context of the listing requested. Requests are made for a full and transparent description of the economic evaluation, as well as for the presentation of sensitivity analyses to demonstrate the robustness of the economic valuation.

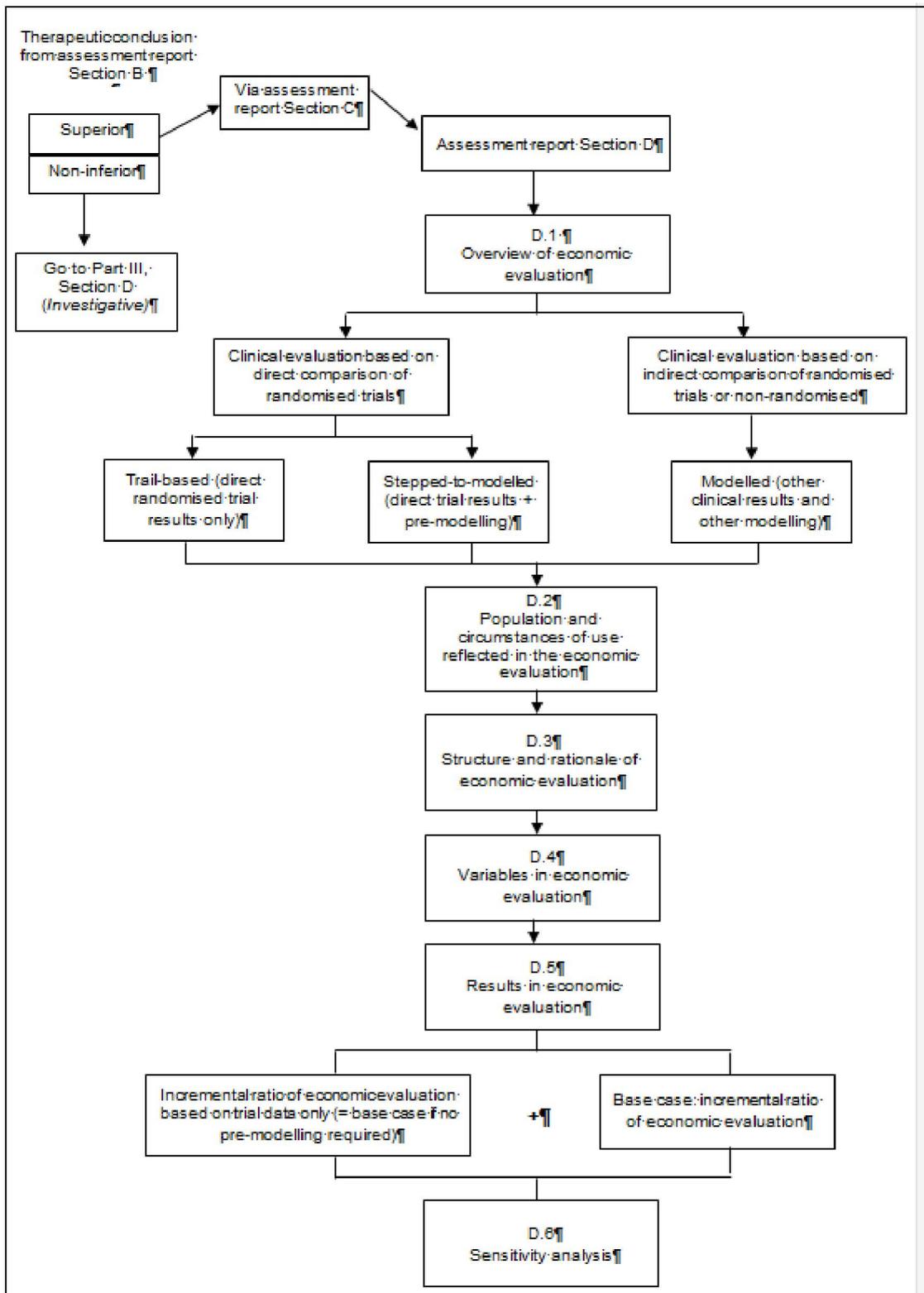
As already described in Part II, Section B and shown in Figure D.1, the economic evaluation of the proposed medical service initially depends on whether the therapeutic conclusion shows:

- the proposed investigative medical service is superior to the main comparator; or
- the proposed investigative medical service is non-inferior (equivalent) to the main comparator; or
- the proposed investigative medical service is inferior to, but significantly less expensive than, the main comparator.

This Section provides information requests for assessment reports for which there is a conclusion of superiority. Information requests for economic evaluations based on a conclusion of non-inferiority are provided in Part III, Section D(i).

Furthermore, the approach described in this Section mainly refers to assessment reports where the economic evaluation is either ‘trial based’ (i.e. based on results from direct clinical trial evidence (See Section B1) trials; or ‘stepped-to-modelled’ (i.e. study results with pre-modelling; see Section C). Thus, it is intended to maximise MSAC’s confidence in an economic evaluation based on this most preferred means of detecting and estimating incremental impact on health outcomes, resource use and cost effects relevant to the requested listing.

Figure D1 Key information requests for assessment report Section D of a standard assessment for MSAC



D1 Overview of the economic evaluation

INFORMATION REQUESTS

- State whether the base case of the economic evaluation is generated by:
 - a trial-based economic evaluation (i.e. based on direct evidence presented in Section B)
- A stepped economic evaluation (i.e. derived from evidence presented in Section B or C. State which type(s) of economic evaluation is presented.
- Provide copies of all the original sources of all data or opinion used, and cross-reference the extracted data to the source documents.

Generation of the base-case economic evaluation

The three steps described below show the approach to an economic evaluation based on a therapeutic conclusion of superiority derived from direct comparison randomised trials.

Step 1: Trial-based economic evaluation

The first step involves an economic evaluation based on the unmodified trial-based estimate of treatment effect on incremental provision of health care resources and incremental health outcomes (i.e. using the most internally valid evidence from the direct evidence presented in Section B. If the direct evidence recruited patients directly representative of those for whom listing is sought, trialled the proposed investigative medical service in the circumstances of use expected to apply to the requested proposed therapeutic medical service if MBS-listed in Australia, and directly measured and reported patient-relevant end points during an appropriate time horizon (i.e. if no pre-modelling studies are reported in Section C), the trial-based evaluation is sufficient to provide the base case of the economic evaluation, and Steps 2 and 3 are not required.

Step 2: Applying treatment effects on health care resource use if MBS-listed in Australia

Frequently, the results of the studies reported in Section B provide insufficient information on which to base a judgment about the full clinical and economic performance of the proposed medical service compared with its main comparator. In these instances, use a modelled economic evaluation to inform MSAC using the results of pre-modelling studies presented in Section C.

The first stage of the economic modelling is to examine the impact of applying the treatment effects on health care resources and health outcomes to the intended proposed medical service population and the circumstances of use identified by the requested restriction (as presented in Section C).

Step 3: Extrapolating and transforming health care resource use and health outcomes if MBS-listed in Australia

The final stage is to examine the additional impact on the modified economic evaluation from Step 2 of extrapolating the health care resource use and health outcomes to the time horizon of the economic evaluation and/or any transformation to final outcomes (also presented in Section C). This generates the stepped base case of the economic evaluation for assessment reports that present pre-modelling studies in Section C.

Justify any proposal to reverse the order of Steps 2 and 3 (i.e. to extrapolate and/or transform the treatment effect before applying it). In this case, the final step would still generate the base case of the economic evaluation.

Examples of reasons for presentation of a stepped economic evaluation rather than just a study-based analysis include:

- the study population and setting might be different from the target population and setting;
- the outcomes measured in the studies might not be the final outcomes of interest for the proposed service;
- a range of outcomes are of interest;
- the time frame of outcomes measured in the studies might be inadequate; and
- resource-use patterns measured in the studies might not fully reflect those expected in practice (e.g. some resources might not be measured in the studies, and some 'protocol-driven' resources might be included that are not relevant to the proposed provision of the service).

Type of economic evaluation

To identify the most appropriate evaluation, the assessment report should first classify the proposed service using the grid provided in Table D1.1 (also refer to Table B8.1). This classification should be based on the differential effectiveness and safety of the proposed investigative medical service under consideration compared with the appropriate comparator(s) when used in the target population and setting (i.e. the information presented in Section B). In classifying the service, it might also be necessary to consider changes in the profile of risks associated with the proposed service, compared with the main comparator(s).

In classifying a service, the quality and strength of the available evidence should be taken into consideration. MSAC has a strong preference for making decisions on the basis of data from direct evidence and will be most influenced by the results of these studies as the most rigorous source of data. However, MSAC has considered and will continue to consider all levels of evidence including linked evidence analysis simply because of the scarcity of direct evidence in relation to investigative medical services.

Where there are trade-offs between incremental effectiveness and incremental safety; that is, where there is reduced effectiveness but improved safety (see Table D1.1) or improved effectiveness but reduced safety, consideration will be required as to whether there are net clinical benefits or net harms to patients, overall. This might involve a valuation of the different effects associated with a service and/or modelling of various outcomes. Assumptions made in reaching the conclusion about whether a service has net clinical benefits should be stated explicitly.

Table D1.1 Classification of a service under MSAC consideration/Classification of the effectiveness of the proposed medical service over its main comparator and guide to the suitable type of economic evaluation.

Comparative safety	Comparative effectiveness			
	Inferior	Uncertain ^a	Non-inferior ^b	Superior
Inferior	Health forgone: need other supportive factors	Health forgone possible: need other supportive factors	Health forgone: need other supportive factors	? Likely CUA
Uncertain ^a	Health forgone possible: need other supportive factors	?	?	? Likely CEA/CUA
Non-inferior ^b	Health forgone: need other supportive factors	?	CMA	CEA/CUA
Superior	? Likely CUA	? Likely CUA	? Likely CEA/CUA	CEA/CUA

CEA = cost-effectiveness analysis; CMA = cost-minimisation analysis; CUA = cost-utility analysis

? = reflects uncertainties and any identified health trade-offs in the economic evaluation, as a minimum in a cost-consequences analysis

a 'Uncertainty' covers concepts such as inadequate minimisation of important sources of bias, lack of statistical significance in an underpowered trial, detecting clinically unimportant therapeutic differences, inconsistent results across trials, and trade-offs within the comparative effectiveness and/or the comparative safety considerations (e.g. where the safety profiles of the compared medical services differ, with some aspects worse for the proposed medical service and some aspects better for the proposed medical service).

b An adequate assessment on 'non-inferiority' is the preferred basis for demonstrating equivalence.

Non-inferior (equivalent) service

If the proposed investigative medical service has been shown to be non-inferior (equivalent) to the main comparator, a cost-minimisation analysis is appropriate (or cost analysis under limited circumstances where the proposed medical service is non-inferior to the main comparator, but has a superior safety profile that generates cost offsets from reduced use of health care resources to manage adverse reactions). Part III, Section D(i) provides the information requests associated with these evaluations.

A cost-minimisation analysis should only be presented when the proposed service has been indisputably demonstrated to be no worse than its main comparator(s) in terms of both effectiveness and safety, so the difference between the service and the appropriate comparator can be reduced to a comparison of costs. In most cases, there will be some uncertainty around such a conclusion (i.e. the conclusion is often not indisputable). Therefore, when an assessment report concludes that an intervention was no worse than a comparator, an assessment of the uncertainty around this conclusion should also be provided by presentation of cost-consequences, cost-effectiveness and/or cost-utility analyses.

Superior service

If the proposed investigative medical service has been shown to be superior to the main comparator, there are four types of economic evaluation that might apply, depending on the outcome of the clinical evidence (see Table D1.1):

- Cost-utility analysis (generally preferred)**

A cost-utility analysis presents the health outcome in terms of the life-years gained from the start of the analysis, with each life-year adjusted by a utility weight that represents society's preferences for the health outcome experiences in that life-year relative to full health. The ultimate benefit of restored health is the restoration of health-related quality of life; for example, restoration of opportunities to undertake activities of daily living. Economists have attempted to identify the value placed by individuals on different health states. The basis for this valuation is that each increment in health-related quality of life gives satisfaction (measured as the strength of preference for the restored health over the pre-treatment state of health and termed 'utility' by economists), which is the ultimate outcome of life. The denominator in a cost-utility analysis is most commonly the incremental QALY gained, which is the difference between the two profiles following the use of the proposed medical service or its main comparator, each calculated as the times spent in successive varying health states, with each period of time weighted by the strength of preference for, or the utility weight of, its respective health state (see Appendix 4 for further guidance on valuing health outcomes in utility terms).
- Cost-effectiveness analysis**

A cost-effectiveness analysis measures the incremental cost per extra unit of health outcome achieved. It differs from a cost-utility analysis in that the health outcome is reported in its natural units. If the proposed investigative medical service is demonstrated to offer more of a given health outcome than its main comparator (e.g. it achieves the desired health outcome in a higher proportion of patients), this goes beyond cost-minimisation. The outcomes reported from the clinical evaluation might need to be transformed in a modelled cost-effectiveness analysis; where this is done the choice of outcome should be justified.
- Cost-benefit analysis (supplementary option)**

A cost-benefit analysis expresses all outcomes (health and non-health) valued in monetary rather than natural or utility units. This is in contrast to other forms of economic evaluations and requires a monetary valuation of these outcomes (see Section A5.2 of Appendix 5). Cost-benefit analysis can also include both health and non-health outcomes.
- Cost-consequences analysis (if disaggregation of outcomes would be helpful)**

A cost-consequences analysis compares the incremental costs of the proposed investigative medical service over its main comparator with an array of outcomes measured in their natural units rather than a single representative outcome as presented in a cost-effectiveness analysis. It can be presented if the proposed investigative medical service is demonstrated to have a different profile of effects that are not adequately captured by a single outcome measure; there might be trade-offs between the two options in terms of the directions of the changes in effectiveness and safety (and within effectiveness and safety). As such, it is a form of disaggregated analysis of changes in patterns of health care resource provision and changes in health outcomes, and can be presented before presenting other types of aggregated economic evaluation, such as a cost-utility analysis.

Table D1.2 shows the type of economic evaluation that should be presented for each classification from Table D1.1.

Table D1.2 Type of economic evaluation that should be presented for various classifications of a service under MSAC consideration

Classification	Type of economic evaluation
The service is more effective than the appropriate comparator and is associated with improved safety.	Cost-consequences, cost-effectiveness, cost-utility, cost-benefit.
The service is more effective than the appropriate comparator and is no worse than the comparator in terms of safety.	Cost-consequences, cost-effectiveness, cost-utility, cost-benefit.
The service is more effective than the appropriate comparator but is associated with reduced safety: (i) Overall, there are net benefits to patients as the benefits from improved effectiveness outweigh the harms from reduced safety and/or changed risk profile. (ii) Overall, the service is no worse than the comparator because the benefits from improved effectiveness at least offset the harms from reduced safety and/or changed risk profile (iii) Overall, there are net harms to patients as the harms from reduced safety and/or changed risk profile outweigh the benefits from improved effectiveness.	(i) Cost-consequences, cost-effectiveness, cost-utility, cost-benefit (ii) Cost-consequences, cost-effectiveness. This may be reducible to cost-minimisation (i.e. presentation of an incremental cost-effectiveness for the base case may be inappropriate if net clinical benefits are assumed to be zero). (iii) No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service.
The service is no worse than the comparator in terms of effectiveness but is associated with improved safety.	Cost-consequences, cost-effectiveness, cost-utility, cost-benefit.
The service is indisputably demonstrated to be no worse than the comparator in terms of both effectiveness and safety.	Cost-minimisation. In the case where there is any uncertainty around the conclusion that the service is no worse than the comparator in terms of effectiveness and safety, cost-consequences, cost-effectiveness, and/or cost-utility analyses should be provided.
The service is no worse than the comparator in terms of effectiveness but is associated with reduced safety.	No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service.
The service is less effective than the comparator but is associated with improved safety: (i) Overall, there are net benefits to patients as the benefits from improved safety and/or changed risk profile outweigh the harms from reduced effectiveness. (ii) Overall, the proposed service is no worse than the comparator because the benefits from improved safety at least offset the harms from reduced effectiveness and/or changed risk profile. (iii) Overall, there are net harms to patients as the harms from reduced effectiveness outweigh the benefits from improved safety and/or changed risk profile The proposed service is less effective than the comparator and is no worse than the comparator in terms of safety.	(i) Cost-consequences, cost-effectiveness, cost-utility, cost-benefit. (ii) Cost-consequences, cost-effectiveness (which may be reducible to cost-minimisation i.e. presentation of an incremental cost-effectiveness for the base case may be inappropriate if net clinical benefits are assumed to be zero). (iii) No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service. No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service.
The proposed service is both less effective than the comparator and is associated with reduced safety compared with the comparator.	No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service.

From Table D1.2, it can be seen that an economic evaluation should be presented in all assessment reports to be considered by MSAC except when a service is indisputably demonstrated to be associated with net clinical harms to patients (as it is unlikely that MSAC will recommend government subsidy of the service).

The assessment report should state what type of economic evaluation is being presented. All analyses should explicitly consider all the advantages and disadvantages of the proposed service that are listed in the clinical balance sheet, compared with the comparator. However, there are some circumstances where simplified analyses will be appropriate and acceptable (see Sub-section D3 for further details).

An iterative approach to the classification and type of analysis might be required. For example, a valuation of the different effects associated with a service and/or modelling of various outcomes might be required before a service can be definitively classified according to Table D1.1. In these cases, the structure of the economic evaluation and the assumptions made in valuation of outcomes must be presented clearly. Adequate sensitivity analysis should also be provided to allow MSAC to gauge the robustness of the classification selected. Thus, although the service might ultimately be classified as being no worse than the comparator (e.g. where improved effectiveness is considered to offset reduced safety), such that a cost-minimisation analysis is considered appropriate, a cost-consequences and a cost-effectiveness analysis that explicitly shows the valuation of the various outcomes should also be presented. Sensitivity analyses should also be presented which examine the effect of varying assumptions in the valuation of outcomes.

Note that the various types of analyses should not be considered mutually exclusive. In many cases it will be appropriate for more than one type of analysis to be presented. As discussed in Sub-section D3, a stepped economic evaluation is requested. Such an analysis will typically start with a cost-consequences analysis and will progress, where appropriate, through various steps where various aspects of modelling are introduced such that, ultimately, a base-case cost-effectiveness or cost-utility analysis is presented. A trade-off between the most appealing outcome upon which to base the economic evaluation from a theoretical point of view and the degree of uncertainty in the estimate of incremental cost-effectiveness is often required. Extrapolation of outcomes beyond the evidence will introduce uncertainty in estimates of incremental cost-effectiveness. For example, the estimate of incremental cost-effectiveness generated by a study-based analysis (i.e. based directly on the outcome from a study) might be relatively robust. However, in moving to a cost-utility analysis (which is theoretically more appealing but where assumptions of utilities for various health states might be required), additional uncertainty might be introduced.

The common output of these evaluations is a comparison of changes in outcomes and changes in costs of achieving those outcomes across the proposed medical service and the main comparator. The objective is usually to justify a price advantage for the proposed medical service compared to its main comparator. A statistically significant improvement in effectiveness alone is not necessarily sufficient to support a conclusion of acceptable cost-effectiveness. Consideration is also given to whether the detected differences are clinically important overall and whether the extent of improvement is sufficient to justify any requested price advantage (after accounting for any justified cost offsets).

General guidance on preferred and supplementary types of economic evaluation

The various types of economic evaluation are not necessarily mutually exclusive and it might be appropriate to present more than one type (e.g. both cost-effectiveness and cost-utility analyses). Depending on the circumstances, there might be a trade-off between the most appealing approach from a theoretical point of view and the degree of uncertainty in the estimate of incremental cost-effectiveness. For example, estimating the incremental cost-effectiveness based directly on the outcome from a trial might be relatively robust. However, additional sources of uncertainty might be introduced when moving to a cost-utility analysis (a cost-utility analysis is theoretically easier to interpret and compare across assessment reports and medical conditions, but it might require assumptions of utility weights for various health states). The three steps described in the beginning of this Sub-section D1 to improve transparency for economic evaluations are designed to help make these trade-offs and their implications explicit.

Given these considerations, a cost-utility analysis is the preferred form of economic evaluation for either or both of the following situations:

- where there is a claim of incremental life-years gained in the economic evaluation — to assess the impact of quality adjusting that survival gain; and
- where relevant direct comparison randomised trials report results using a multi-attribute utility instrument (MAUI).

However, for the reasons given above, the preference for a full cost-utility analysis is less clear in other situations, even where there is a claim of quality-of-life or disability improvements, or where there are differential quality-of-life impacts arising from the therapies being compared in an assessment report to derive a common outcome across assessment reports. Therefore, in the situation of an improvement in quality of life but not in quantity of life, an assessment report should present a cost-utility analysis or justify the decision to not transform the quantified health outcomes via a utility valuation.

Cost-benefit analysis is not preferred because it is not likely to be helpful to most MSAC deliberations (further reasons are given in Appendix 5). Thus, although monetary valuation of health outcomes is allowed, it is considered to be supplementary to utility valuation presented in a cost-utility analysis. If a cost-benefit analysis is presented in the absence of a cost-utility analysis, MSAC might not consider it to have the same weight.

Similarly, the base-case economic evaluation should be focused on material incremental changes in the provision of health care resources and on material incremental changes in health outcomes. Supplementary analyses can be used to present any material incremental changes in the provision of non-health care resources and/or in non-health outcomes.

Sources of information

Separately provide copies of the original sources of all data (beyond those already presented in Sections B and C) or expert opinion used in the model in an attachment or technical document. Cross-reference the extraction of data from each source to the level of the page, table or figure number of the source document.

D2 Population and circumstances of use reflected in the economic evaluation

INFORMATION REQUESTS

- Describe and justify the demographic and patient characteristics of the population included in the economic evaluation.
- Describe and justify the circumstances in which the proposed medical service and main comparator are used in the economic evaluation.
- Assess the consistency of the demographic and patient characteristics and of the specified circumstances of use across the study populations, the population in the economic evaluation and the population for whom listing is sought.

In this Section, analysts should provide information to allow MSAC to assess whether the evidence presented is applicable and generalisable to the population and circumstances of use for which the service is proposed (see Table D2.1).

Table D2.1 Definitions for populations and circumstance of use that should be taken into account in the evaluation

Term	Description
Target population and circumstances of use	Population and setting for which government subsidy of the service is being requested
Study population and circumstances of use	Population and setting for which evidence of efficacy and safety has been presented in assessment reports Sections B and/or C
Wider population and circumstances of use	Broader population and setting in which the service is likely to be used if MBS-listed

Population (demographic and patient characteristics)

Use summary statistics (where appropriate) to describe the demographic and clinical characteristics for the population entering the economic evaluation. Include information about the distribution around means where appropriate.

Examples of patient characteristics are provided in Section A.

Use cross-references, as appropriate, to Section A when justifying the definition of each characteristic of the population in the economic evaluation in relation to the population for whom listing is sought. Also highlight any difference in relation to the study populations for whom evidence of effectiveness and safety are presented (using cross-references, as appropriate (Sub-section C4) if pre-modelling studies are presented to apply these results).

Circumstances of use

Use cross-references, as appropriate, to Section A when describing and justifying the definition of each circumstance of use (setting) assumed in the economic evaluation in relation to the medical condition under which listing is sought. Also highlight any difference in relation to each circumstance for which evidence of effectiveness and safety is presented from the studies (Sub-section C4) if pre-modelling studies are presented to apply these results).

Examples of types of circumstance are provided in Section A.

The application or assessment report should describe the setting in which the service and its main comparator(s) are assumed to be used in the economic evaluation. Examples of elements of settings that could be detailed 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 services; for example, in a 24-hour or in a 12 or 24-month period;
- any required co-delivered medical services or treatments (including any additional diagnostic tests required);
- any contra-indicated medical services or treatments;
- any unique characteristics of the referrer or provider (e.g. specific qualifications or training); and
- any specific requirements in terms of geography, facilities or location of delivery of service (e.g. limited to hospital setting or to approved laboratories; specification of any specific equipment or facilities that need to be available).

Consistency across characteristics

Assess the degree of consistency of the demographic and patient characteristics and of the specified circumstances of use across:

- the study populations and circumstances of use described in Sub-section C4 if pre-modelling studies are presented to apply the results of these trials);
- the target population and circumstances of use, which should reflect the clinical management algorithms presented in Section A; and
- the wider population and circumstances.

The population for whom funding is being examined might be less well defined than the other two groups. However, its inclusion captures the potential for use of the proposed medical service in a broader population and/or broader circumstances than the target population and circumstances if the proposed medical service were MBS-listed in Australia. Including the population might also be useful for capturing any limitations of the economic evaluation in truly replicating the target population and circumstances. The importance of examining the incremental cost-effectiveness of the proposed medical service in this population increases with increasing risk of substantial use of the proposed medical service beyond the intention of the requested restriction (see also Sub-section D6).

Table D2.2 suggests a format that will summarise these characteristics and circumstances for which sensitivity analysis shows that the variable is important.

Table D2.2 Comparison of characteristics of trial and requested populations and circumstances of use

Population and circumstance ^a	As defined in trial(s) using ITT population	As defined by the requested restriction	If use beyond the requested restriction might arise
Medical condition of the population			
Age of the population			
Restriction criteria (including any limitations on disease severity, preconditions or previous treatments, or continuation rules)			
Limitations on response or surgical experience considerations of use of proposed medical service			
Repeat for each other variable that varies across these populations and circumstances, and for which sensitivity analysis shows the variable is important			

ITT = intention to treat

a For each identified population characteristic and circumstance of use, provide a footnote explaining any differences between these populations and relate this to any pre-modelling study presented in Section C to apply the evidence from the overview of the trial(s) to the requested restriction.

Justifying restrictions

In the case where it is proposed that eligibility for a service be restricted to a subgroup of patients with a clinical condition, the proposed restriction should be justified as follows:

- The intention of the requested restriction should be indicated in the assessment report.
- To help minimise usage beyond the intention of the requested restriction, for each population or setting element included in the wording of a restriction, the assessment report should:
 - identify and define the element unambiguously for example:
 - risk factors associated with the medical condition;
 - markers of severity or progression of the medical condition; and
 - name of service and duration criteria for previous medical services, as appropriate;
 - specify objective criteria in preference to subjective criteria in identifying the element;
 - justify any thresholds within these criteria (these thresholds and justifications should be consistent with study eligibility criteria and subgroup stratification criteria as appropriate); and
 - resolve copyright issues about any proposed medical service before proposing its use as part of a restriction.

The assessment report should present a discussion addressing the trade-offs between the clinical preference for simple, unambiguous listings versus increasingly complex restrictions designed to limit new services to those relatively few patients for whom the proposed service might be justified as being acceptably cost-effective at the price requested.

The further the eligibility criteria specified in a restriction shift practice away from otherwise uninfluenced practice, the more incentive there is for referrers/providers and patients to seek subsidy despite the restriction. The approach listed above (identifying and justifying any restrictions) is intended to help justify the choice of restriction from the alternative options that might apply. This approach becomes more important as the restriction becomes more complex or more expensive for the relevant body to administer.

If the proposal is for eligibility for a service to be restricted to a subgroup of patients with a clinical condition, the potential for use of the service in a wider population or setting than the target population and setting, if government subsidy of the service is recommended, should also be assessed.

Presenting the information

Table D2.3 shows a hypothetical example where it is proposed that a new treatment be made available as a second-line agent for the management of adults with hyperthyroidism to provide a suggested format for presentation of information about the target, study and wider populations and settings.

Where there are differences, or potential differences, between any of the groups, economic analyses should be presented for each of the scenarios.

When presenting economic evaluations for different populations, the assessment report should consider whether changes in the population have implications for the cost associated with the proposed service (e.g. if economies of scale might be captured by using a service in a wider population). Further advice is provided in Sub-section D4.

Table D2.3 Example of a comparison of the characteristics of target, study and wider populations and settings

Population	Target	Study	Wider
Clinical condition	Hyperthyroidism due to any cause	Hyperthyroidism due to Graves' disease	Hyperthyroidism due to any cause
Comment	Only patients with hyperthyroidism due to Graves' disease were recruited to the only direct comparison randomised trial comparing service A with service B (Jones et al 2000), but subsidy is requested for all patients with hyperthyroidism, regardless of aetiology. Smaller, non-comparative studies (Brown et al 1995, Smith et al 1997) have examined the efficacy and safety of Service A in patients with hyperthyroidism due to other causes. The effect size observed in these studies was similar to that observed in Jones et al, 2000; however, it is acknowledged that <etc.>.		
Age	Adults	18–75 years	Adults
Comment	Although only patients aged up to 75 years were eligible for entry to the direct comparison randomised trial comparing service A with service B (Jones et al 2000), service A has been used in patients over the age of 75 with similar effects as in other adult populations (Smith et al 1990) <etc.>.		
Gender	70% females 30% males	50% females 50% males	70% females 30% males
Comment	Although the proportion of females with condition X recruited to the trial reported by Jones et al 2000 was lower than the proportion of females with hyperthyroidism in the Australia, a test for interaction did not demonstrate gender to be a treatment effect modifier <etc.>.		
Initiation criteria	Serum TSH < 70% x Serum T₃ > 120% y	Serum TSH < x Serum T₃ > y	Serum TSH < 85% x Serum T₃ > 110% y
Comment	Subsidy of service A is requested for a more severely affected population than recruited to the trial reported by Jones et al (2000). Subgroup analysis demonstrates serum TSH and T ₃ levels at baseline to be a treatment effect modifier, with a greater relative response rate to service A in patients with levels of serum TSH below 70% x and levels of serum T ₃ greater than 120% y. It is acknowledged that there might be some use beyond the population for whom subsidy of service A is sought. Thus, sensitivity analyses are presented examining the effect on incremental cost-effectiveness and financial implications of use of the service beyond the population for whom subsidy is sought.		
Position in management algorithm	Second line	Second line	Second line but some first-line use
Comment	Consistent with the direct comparison randomised trial (Jones et al 2000) comparing service A with service B, subsidy is proposed for use of service A only in patients failing to respond to service C. However, it is acknowledged that there might be some use of service A in the first-line management of hyperthyroidism (i.e. as a substitute for service C instead of service B). Thus, cost-effectiveness analysis is also presented versus service C.		
Limitations on frequency of use	Patients will be permitted to receive service A as a subsidised service on two separate occasions	Patients were permitted to receive service A on two separate occasions	Patients will be permitted to receive service A as a subsidised service on two separate occasions
Comment	The number of times the service might be delivered to the patients on a subsidised basis is consistent with the number of times the patients were able to receive the service in the clinical trial reported by Jones et al (2000).		

D3 Structure and rationale of the economic evaluation

INFORMATION REQUESTS

- Review the relevant economic literature and present the results.
- Specify any software used to conduct the economic evaluation.
- Ensure that all variables in the electronic copy of the economic evaluation can be changed independently during the evaluation, including allowing the base case of the economic evaluation to be completely respecified and allowing a new set of sensitivity analyses to be conducted with each respecified base case.
- Describe the structure of the economic evaluation.
- Justify the appropriateness of the structure in reflecting the context of use of the compared alternatives and the outcomes of their use.
- Define and justify the time horizon and nature of the outcomes used in the economic evaluation.
- Describe the methods used to calculate the results of the economic evaluation (e.g. cohort expected value analysis, Monte Carlo simulation).
- Provide copies of identified papers in an appropriately labelled attachment separate from the main body of the assessment report.

By definition, the economic evaluation is intended to inform a decision. Therefore, the structure of the evaluation allows the comparison of the streams of outcomes and resources following the use of either the proposed medical service or its main comparator to calculate incremental outcomes and costs of these streams. MSAC has a preference for a decision-analytical framework that clarifies the comparison of these streams of outcomes and resources.

Literature review

Applicants should search the literature for published cost-effectiveness analyses of the proposed service. A list of all of the published reports that are retrieved by the search should be provided in the application or assessment report.

The economic analyses that are directly relevant to MSAC's considerations (i.e. economic evaluations performed for the same population and setting in which the service will be used) should be identified using a tight set of inclusion and exclusion criteria, which should be detailed in the assessment report. The assessment report should also provide a critical review of the included studies.

An independent economic evaluation might not be required if there is already a high-quality economic evaluation in the public domain that provides an estimate of incremental cost-effectiveness for the proposed service in a population and setting that is similar to the proposed Australian population and setting. Such an evaluation needs to be based on the appropriate:

- therapeutic and management setting;
- patient population; and
- input variables.

In these circumstances, an assessment of the most appropriate publicly available evaluation should be presented in the report for MSAC according to the requirements of these Guidelines. That is, the evaluation available in the public domain should be assessed according to this Sub-section, and Sub-sections D4 and D5. All details requested in these Sections should be provided in the assessment report.

Where a model in the public domain is considered to have an appropriate structure, but is populated with values for variables that do not correspond to the values that would apply in the Australian population and setting or proposed by the Final PICO Confirmation, it might be appropriate to use the model but to update values for the variables to values that would apply in the Australian context. Again, the model should be assessed according to this Sub-section, and Sub-sections D4 and D5. All of these Guidelines and all details requested in these Sections should be provided in the assessment report.

If a search of the literature fails to identify any directly relevant economic evaluations, an independent economic evaluation should be conducted. This Sub-section, and Sub-sections D4 and D5 of these Guidelines describe the information required and how the economic evaluation should be presented.

Present the results of a search of the literature for reports of economic evaluations of similar decision analyses (in terms of similarity to the treatment algorithm and/or the proposed and similar medical services). Where the assessment report's model is different from the literature-sourced models, explain the basis for the selection of the assessment report's approach.

Software package

Specify the name and version of any software package used to conduct the economic evaluation. Software packages that support decision analyses and can be readily critiqued currently consist of:

- TreeAge Pro Suite[®];
- Excel 2010-2013[®], including @RISK[®], but not necessarily including all advanced features and plug-ins (e.g. Crystal Ball[®] and customised macros developed using Visual Basic); and
- STATA.

Economic evaluations constructed using any of these may be submitted without earlier arrangement with the HTA Team. For further information, please refer to Section 5 of the Guidelines.

Fully accessible electronic copy of the economic evaluation

Ensure that all variables in the electronic copy of the economic evaluation can be changed independently, including allowing the base case of the economic evaluation to be completely respecified and allowing a new set of sensitivity analyses to be conducted with each respecified base case.

Structure of the economic evaluation

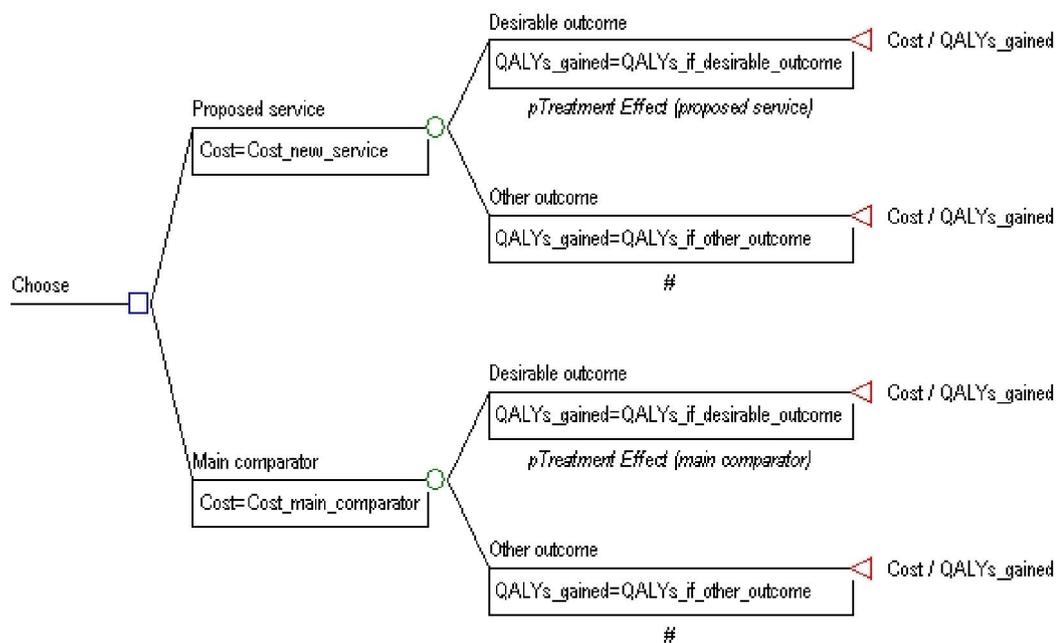
The description of the economic evaluation should include:

- a statement defining in detail the medical service options for which costs and outcomes are estimated in the economic evaluation;
- a description of each of the types of event and health states possible in the economic evaluation, together with a justification of the selection of each health state for inclusion in the evaluation and a justification for those that were considered potentially suitable but that were excluded to avoid excessive complexity;
- a description of the relationships and interactions between the various events and health states possible in the economic evaluation (including, where relevant for a state transition model, a detailed description of all possible transitions between the health states, see below);
- a description of all assumptions made in the construction of the economic evaluation; and
- a decision tree diagram summarising the structure of the economic evaluation.

Study-based evaluation

If the study population and setting are the same as the target population for the proposed service and outcomes have been reported for all patient-relevant endpoints, it might be appropriate to present a simple economic evaluation based directly on the results of the included studies. The structure of a basic economic evaluation is shown in Figure D3.1.

Figure D3.1 Example of the structure of a basic economic evaluation



Stepped economic evaluation

Frequently, the results of the available studies provide insufficient information on which to base a judgement about the clinical and economic performance of the proposed service relative to that of the comparator. In these circumstances (which are a matter of judgement), a stepped economic evaluation (which introduces the various aspects of modelling in separate steps) will be useful to MSAC. Examples of reasons for presentation of a stepped economic evaluation rather than just a study-based analysis include:

- the study population and setting might be different to the target population and setting;
- the outcomes measured in the studies might not be the final outcomes of interest for the proposed service;
- a range of outcomes are of interest;
- the time frame of outcomes measured in the studies might be inadequate; and
- resource-use patterns measured in the studies might not fully reflect those expected in practice (e.g. some resources might not be measured in the studies, and some resources might be included that are not relevant to the proposed provision of the service).

Presenting a stepped evaluation

To ensure that the manner in which available information is incorporated into the economic evaluation is transparent, MSAC requires the presentation of a stepped economic evaluation that starts with a study-based cost-consequences analysis and progresses through various steps of the modelling in turn (population and setting, outcome, time horizon, resource use, etc). These steps might require the presentation of additional evidence. Guidance for the presentation of this evidence is provided in Section C.

MSAC recognises that the conduct of a complex economic evaluation for a service might be associated with costs that could exceed the costs of actually providing that service. Therefore, a simple economic evaluation, such as a study-based economic evaluation or a simplified model, is acceptable if the following criteria are both met:

- the service is likely to be used by small numbers of patients; and
- the total government expenditure on the service is likely to be small.

To ensure consistency across economic analyses considered by MSAC, the preferred elements of a base-case economic evaluation are summarised in Table D3.1.

Table D3.1 Key elements of the base-case economic evaluation

Element of economic evaluation	MSAC's preference for the base-case analysis	Section providing further details
Perspective	Societal perspective. However, costs and benefits should be presented aggregated to the following three levels: <ul style="list-style-type: none"> • taking an MBS (or other relevant government program) perspective (i.e. including costs and benefits incurred by the MBS); • taking a health care perspective (i.e. including only costs related to provision of health care resources regardless of who incurs them, and including only health outcomes); and • taking a societal perspective (i.e. including all costs and benefits). 	Section D
Comparator	Currently available service that is most likely to be replaced by the new service	Section A
Type of economic evaluation	Cost-effectiveness analysis	Section D
Source of evidence	Systematic review	Section B
Values of parameters	Unbiased, plausible estimates. Where there is room for judgement and considerable uncertainty around the value of a parameter, a conservative approach to the valuation of that parameter should be adopted.	Section D3
Outcome on which evaluation should be based	The outcome measure that most closely and validly estimates the final health outcome from a patient perspective. Health-related QALYs should be used where feasible.	Section D3
Discount rate	An annual rate of 5% for both costs and benefits	Section D4

A description of the structure of each step of the economic evaluation should be provided, and include:

- an explicit statement of the options for which costs and benefits are being estimated in the economic evaluation, and the justification for the selection of options included in the evaluation;
- a description of each of the events and health states possible in the economic evaluation;
- justification of the selection of health states for inclusion in the economic evaluation (and those excluded to avoid excessive complexity);
- a description of the relationships and interactions between the various events and health states possible in the economic evaluation (including detail of the transitions possible between the health states);
- a description of assumptions (both implicit and explicit) made in the construct of the economic evaluation; and
- a decision-tree diagram summarising the structure of the economic evaluation.

The assessment report should present a justification for the overall structure of the base-case economic evaluation, particularly in relation to:

- the natural history of the condition being managed, prevented or diagnosed;
- the management algorithm that applies currently and the management algorithm that will apply should the service be MBS-listed;
- the management algorithm that applied in the studies used as evidence to demonstrate the safety and effectiveness of the proposed service; and
- the structure of other relevant models reported in the public domain.

The report should also identify and consider assumptions built into the structure of the economic evaluation and comment as appropriate.

Defining and justifying the time horizon

The time horizon over which costs and benefits of a service and its comparator are measured in each step of the evaluation should be defined and justified. The assessment report should define and justify the time points at which events are assumed to occur and the duration of time spent in health states (include details of cycle length for Markov models). The appropriate time horizon for follow-up will relate to the natural history of the disease, the treatment pattern and the time period over which outcomes from the service or main comparator could be expected to occur. For example, the time horizon over which costs and health benefits of a diagnostic test for an acute event (e.g. a nonlife-threatening infection) might be relatively short, whereas the appropriate time horizon to consider for a treatment for a chronic illness will be longer.

Discounting

Where costs and benefits of a service and/or its comparator are presumed to be borne over more than one year, the present value of future costs and benefits should be used in the economic evaluation. This means that discounting should be applied to both costs and benefits sustained in the period beyond the first year. Costs and benefits should be discounted at an annual rate of 5%. As discussed in Section D6, a sensitivity analysis examining the impact of discounting should be performed.

Describing the methods used

The methods used to generate results of the economic evaluation should be described; for example:

- expected value analysis (or cohort analysis);
- Monte Carlo simulation (the assessment report should specify whether first-order and/or second-order distributions are sampled); and
- Markov models (the assessment report should specify whether a half-cycle correction has been included or justify its exclusion).

Dealing with uncertainty

The value of information from a complex economic evaluation diminishes as greater uncertainties are introduced through the process of modelling. The application or assessment report should consider the extent to which the value of more extensive analysis will be limited by the quality of the underlying data and the extent to which uncertainties in the clinical evidence will be amplified by modelling. Progression through modelling steps should continue only as long as the results generated are likely to be of value and informative to MSAC.

The type of presentation that is likely to be of greatest value to MSAC might vary with the level of evidence available. For example, in some circumstances, the evidence base might be extremely weak (e.g. where a claim that a service is safe and ‘promising’ in terms of 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, and that essentially seeks to determine the minimum extent of benefit that would be required for the service to be considered acceptably cost-effective, might be more informative than reporting of an incremental cost-effectiveness ratio based on a single point-estimate of incremental effectiveness.

The objective of cost-effectiveness analysis should be to provide an unbiased, plausible estimate of the incremental cost-effectiveness of the medical services being compared. Where an element of judgement is required, and where there is considerable uncertainty around the value of a parameter, a conservative approach to the assignment of a value to that parameter should be adopted for inclusion in the base case.

Justification of the structure

Justify the overall structure of the economic evaluation in relation to the current and proposed clinical management algorithms (and the requested restriction, as appropriate) presented in Section A of the assessment report, and the treatment algorithms used in the studies presented (using cross-references, as appropriate, to Sections B and C of the assessment report). When justifying the overall structure of the economic evaluation in relation to the current and proposed clinical management algorithms, discuss the consistency across:

- the alternative treatment options examined in the economic evaluation and those considered appropriate in response to Sub-section A5;
- the clinical management algorithms assumed in the structure of the economic evaluation before and after the implementation of the requested listing and the algorithms presented in response to Sub-section A6; and
- the clinical management algorithms assumed in the structure of the economic evaluation and the clinical management algorithms for which clinical evidence is presented in Sections B and C.

Identify and consider implicit assumptions built into the structure of the economic evaluation and comment as appropriate.

Time horizon and outcomes used in the evaluation

Time horizon

Define and justify the time horizon over which the costs and outcomes of the proposed medical service and its main comparator are estimated in the economic evaluation. The appropriate time horizon for follow-up relates to the natural history of the medical condition, the treatment patterns, and an estimation of the time period(s) over which outcomes from the two therapies would be expected to occur. For example, a relatively short time horizon could apply when treating an acute event, whereas a longer time horizon would be required for a chronic illness.

Outcomes

Indicate whether the outcomes generated by the economic evaluation represent the final outcomes of treatment. Where the economic modelling structure is used (rather than a separate pre-modelling study, see the *Therapeutic Guidelines* to transform a quantified treatment effect measured on a surrogate outcome in the trials to predict a subsequent quantified treatment effect on the intended final outcome, explain and justify the method of this transformation, including a justification for how the relationship might vary over time. Use a pre-modelling study to show that a systematic approach has been taken to select and justify the modelling approach taken to estimate the final outcomes.

Methods used to generate the results

Describe the methods used to calculate the results of the economic evaluation (e.g. directly trial-based, cohort expected value analysis, Monte Carlo simulation).

If the economic evaluation is directly based on individual patient data on costs and outcomes from a relevant, direct comparison randomised trial, indicate whether a probabilistic sensitivity analysis has also been conducted. If so, indicate whether the sensitivity analysis has been calculated parametrically (e.g. Fiellers method) or non-parametrically (e.g. bootstrapping), and justify the choice of method.

Where quantified estimates of outcomes are generated over time, explain the underlying assumptions and rationale, for instance, in sufferers of COPD, the incidence of severe exacerbation events requiring treatment become more frequent and severe as the disease progresses. In other medical conditions, assuming a linear relationship between outcomes and time might be clinically plausible, identify and consider inferential assumptions built into the structure of the economic evaluation and comment as appropriate. Show that a systematic approach has been taken to select and justify the assumptions made to quantify the outcomes over time; for example, reference the literature search for similar economic evaluations and/or using a pre-modelling study to present the search for studies of the natural history of the condition.

State transition models

For models involving more than one time period (e.g. state transition models) present the transition diagram (or matrix). This complements the decision-tree diagram by identifying the health states possible in the economic evaluation, indicating the presence and direction of transitional paths between health states, and defining the type of each health state as appropriate (e.g. temporary, absorbing).

Describe the model mechanics: define and justify the cycle length and the follow-up time, and comment as necessary. Define and justify the time points at which events are assumed to occur and the duration of time spent in health states. For a Markov model, specify whether a half-cycle correction has been included or justify its exclusion.

Clearly link each patient-relevant outcome and resource item in the model to its relevant health state(s).

Comment as appropriate on the impact of implicit assumptions inherent in the method chosen. For example, for an economic evaluation that includes Markov components, it is relevant to check the following assumptions:

- Is the memorylessness assumption of the model valid in this case (i.e. is it correct to assume no memory for previous states, such that transition probabilities are independent of previous states)?
- Are there constant or non-constant transition probabilities? If the transition probabilities are constant or homogenous across cycles in the model, they are assumed to be independent of time and thus independent of time-related probabilities, such as ageing of the population and variation in competing risks of the population over time. Allowing for ageing and variation in competing risks of the population over time requires transition probabilities that can vary (i.e. are non-homogenous) across time (number of cycles) in the model.

Describe how the model is calculated (e.g. hypothetical cohort or Monte Carlo simulation). If a Monte Carlo simulation is used, then also:

- specify the number of iterations used per simulation and justify this selection in terms of whether it samples the distribution(s) adequately;
- specify the number of simulations per analysis and justify this selection; and
- indicate whether second-order (or parameter) uncertainty has been simulated and hence whether probabilistic sensitivity analysis is enabled.

Sources of information

Papers identified from the literature review are a useful resource for assumptions relating to the structure and variables in the economic evaluation. Provide copies of all identified papers used in the evaluation in an appropriately labelled attachment separate from the main body of the assessment report.

D4 Variables in the economic evaluation

INFORMATION REQUESTS

- Present, as a minimum, the following information for each variable used in the economic evaluation:
 - name (and definition, as necessary);
 - quantity in natural units (as appropriate; for example, this is not applicable for unit costs); and
 - source.
- Identify and list the direct health care resource items for which there would be a change in use associated with substituting the proposed medical service for the main comparator and define each in terms of natural units.
- Estimate the present value of direct health care resource costs and health outcomes.
- Discuss the implications for the economic evaluation of any important deficiencies in the available evidence base.
- Summarise this information in a table for each type of variable and provide further details of calculations, as necessary.

Variables used in the evaluation

Variables used in the economic evaluation might include:

- health care resource items provided (unit costs should be presented and sourced, quantities should be provided as appropriate);
- outcomes (presented in such a way as to allow the three steps to increase transparency to be distinguished);
- probabilities within each branch of a decision analysis (including transition probabilities or rates in a state transition decision analysis); and
- the discount rate applied to costs and outcomes (discount costs and outcomes incurred beyond the first year at a rate of 5% per year).

The names and definitions of variables should be sufficiently precise to permit verification and replication of the economic evaluation. For example, an Australian Refined Diagnosis Related Group (AR-DRG) item number is more precise than an episode of hospitalisation. For each source, provide full citation details, including item number or page number as appropriate. It might be necessary to cite more than one source for some variables (e.g. the quantity and unit cost of a resource item).

Each economic evaluation should consider explicitly all material differential effects between the proposed medical service and its main comparator (i.e. include all advantages and disadvantages in the analysis). To help demonstrate this, Sub-section D5 requests the presentation of the results of the economic evaluation first in disaggregated form (i.e. as an array of all material costs and consequences; see the definition of a cost-consequences analysis in Sub-section D1).

For the results of trials and pre-modelling studies conducted to provide variables for the economic evaluation, cross-refer to relevant responses to Sections B and C of the Guidelines as appropriate.

Justify and assess the impact of any change in the source of information for a variable used in the evaluation from that given or recommended elsewhere. For some variables where there is no recommended source and several different options are available (e.g. rates of progression of a chronic medical condition), it might be important to show that a systematic approach has been taken to select and justify the option used in the economic evaluation (e.g. using a pre-modelling study). The judgment of this importance should be influenced by the sensitivity of the results of the economic evaluation to substituting the different options for the selected option.

Discuss the implications for the economic evaluation of any important deficiencies in the available evidence base. For example, some variables might be estimated imprecisely, or evidence might have been gathered in different populations and circumstances of use or in other health care systems (which is arguably more important for costs). In such cases, explain the limitations of the data and provide details of any attempts to overcome those limitations. Assess the implications using a sensitivity analyses (see also Sub-section D6).

Adverse reactions

Including information on adverse reactions in an economic evaluation can be difficult. Adverse reactions have two main impacts on an economic evaluation: they affect the health outcomes of proposed medical service treatment, and they contribute to the total cost of treatment. Avoidance of an adverse reaction typically associated with the use of the main comparator might be an important and intended outcome of treatment with the proposed medical service. Adverse reactions might affect quality of life, particularly if they have to be tolerated over long periods. Adverse reactions might also lead to discontinuation of the medical service and subsequent substitution of another medical service. A comparative analysis of time-to-treatment cessation of the proposed medical service and the main comparator on the basis of intention-to-treat is useful in this situation. Adverse reactions can contribute to costs through unintended hospitalisations, and additional procedures and investigations. Deal appropriately with these impacts to avoid double-counting in the economic evaluation. Generally, the preferred approach is to include them in a full economic evaluation. However, in some circumstances, presenting a cost analysis might suffice (see Part III, Section D(i)).

Direct health care resources

The health care resource items for which there would be a change in use associated with substituting the proposed medical service for the main comparator need to be identified.

The following should be considered where appropriate:

- proposed therapeutic medical services (direct costs of treatment and proposed therapeutic medical services used to treat adverse reactions);
- medicines, including pharmaceutical benefits;
- hospital services;
- diagnostic and investigational services;

- community-based services; and
- any other direct medical costs.

Define the natural units, such as number of general practitioner consultations or admissions per diagnosis-related group, used to measure the change in the amount of each resource item.

Present value of direct health care resource costs

For each type of health care resource, quantify the number of natural units provided for each alternative (e.g. number of general practitioner consultations, allied health care practitioners, surgery assistants, anaesthetists, number of episodes of hospital admissions). The relevant economic measure is the amount of resource provided, rather than the amount of resource consumed.

Describe and justify the basis for these estimates, specifying the source of the information. The pattern of provision of resources might be measured prospectively in the course of a clinical study by retrospective review of relevant records, by administration of a questionnaire or survey, or through the use of 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.

Section D adopts a broad perspective for the valuation of health care resources, so all contributions to the costs of health care resources - including those paid for by, governments, health insurance agencies and any other part of society including potential costs for patients where relevant - should be considered for inclusion in the economic evaluation. In contrast, Section E primarily considers contributions to resources paid for if publicly funded in Australia only and by government health budgets only.

It might be reasonable to exclude types of resources that have such a small impact on incremental costs that they would not have a material influence on the conclusion of the economic evaluation.

The unit prices should be as current as possible at the date of the submission of the assessment report. If there are particularly pressing reasons to use different unit prices, justify each and supply its source or describe its generation. Ensure that any different unit price is consistent with the broad perspective of including all contributions to the costs of health care resources, in keeping with the rest of this document. To permit MSAC to gauge the effect of using the alternative unit costs, present the results of the economic evaluation using first the unit costs recommended by the manual and then the alternative unit costs.

A format for summarising the minimum dataset of resource items and their associated unit costs relevant to the economic evaluation is suggested in Table D4.1. Some rows have been completed to clarify the suggested format. These are samples for each identified category, which are consistent with the manual, but are not comprehensive of all types of health care resource items, natural units of measurement, or sources of unit costs.

Table D4.1 List of health care resource items and unit costs included in the economic evaluation

Resource item	Unit of measurement	Unit cost	Bearer of cost	Source of unit cost
Medical services				
GP attendance	Visit	\$x	\$y (MBS)	MBS item 23 ^a
Initial specialist attendance	Visit	\$x	\$y (MBS)	MBS item 104 ^a
Subsequent specialist attendance	Visit	\$x	\$y (MBS)	MBS item 105 ^a
Hospital services				
Hospitalisation for retinal procedure	Hospital stay	\$x	\$x (government)	Average cost per DRG according to AR-DRG Public Sector Estimated Cost Weights Round 7 — Item C03Z
Diagnostic and investigational services				
Ultrasound of orbital contents	Visit	\$x	\$y (MBS)	MBS item 55030 ^a
Pharmaceuticals				
Tobramycin eye drops	Bottle (5 mL)	\$x	\$y (PBS) \$z (patient)	PBS item 2328 M — average co-payment estimated assuming a percentage of patients are general and remainder are concessional
Tobramycin eye ointment	Tube (3.5 mL)	\$x	\$y (PBS) \$z (patient)	PBS item 2329N — average co-payment estimated assuming a percentage of patients are general and remainder are concessional

^a Costs can be obtained from MBS Online.

All steps taken to calculate costs in the economic evaluation should be presented in a way that allows independent verification of the calculations. If a complete presentation is likely to make the main body of the assessment report too bulky, the calculations should be presented in a technical document (see Section 4.3 in Part I). Provide clear cross-references between the calculations and the main body of the assessment report. Include an electronic version of the detailed calculations.

Value future costs at current prices. This is consistent with using constant prices in the economic evaluation. Accordingly, no allowance for future inflation should be included in the calculations.

The present value of future costs should also be estimated. This means that where costs extend over a number of time periods (beyond one year), they should be discounted. Discounting of future costs and benefits is a standard feature of an economic evaluation. Costs or benefits are discounted at an annual rate of 5%. If discounting is important in an economic evaluation, this can be examined in sensitivity analyses using different discount rates (see Sub-section D6).

Present value of health outcomes

Nominate and justify the outcome that is considered to best reflect the comparative clinical management algorithm performance of the medical services being compared. This should generally be based on the outcome measure that most closely and validly estimates the final health outcome from a patient perspective. The outcome on which the economic evaluation is based might need to reflect more than one type of intermediate outcome (e.g. where desired and adverse outcomes need to be considered). Justify the choice of any other outcome measure included in the economic evaluation.

For each relevant outcome, quantify the effect of the proposed medical service on the course of the medical condition being managed, either in terms of direct increments, or as streams of effects for the proposed medical service and main comparator in separate arms of the decision analysis, with the increments determined across the arms. Where possible and appropriate, quantify this effect in terms of the patient's health-related quality of life, distributed across different health states over time. Where utility weights were not elicited via a MAUI in the direct comparison randomised trials, this might form a basis for valuing these effects in a manner that reflects the preferences of the general population (see Section C and Appendix 5). Describe and justify the basis for these estimates, specifying the source of the information, including by reference to the data presented in Sections B or C of the assessment report. Distinguish between data on outcomes that are directly derived from the primary evidence, and extrapolations or modelling of outcomes beyond that available from the primary evidence.

List and document all variables influencing the estimate of outcomes in a table. In the table, highlight the variables that generate the incremental treatment effect on the final outcome estimated in the economic evaluation. These variables include the health states representing the patient-relevant outcomes and the probabilities in each branch of the decision analysis that together simulate a treatment effect by differing between the two arms (representing the proposed medical service and its main comparator) of the economic evaluation. Explain the mechanics of this simulation, because it is usually an important driver of an economic evaluation, and assess the resulting estimate of incremental treatment effect in the context of the analyses presented in Sections B or C of the assessment report.

The present value of future health outcomes measured from the trials or estimated from the model should also be calculated using the approach described above for costs.

If health-related quality of life is not measured directly in the direct comparison randomised trials using a MAUI, which allows direct translation to utility weights via the associated preference-based scoring algorithm, the economic evaluation might include scenario-based utility weights to transform the outcomes measured in those trials into a cost-utility analysis (see Sub-section D1 and Appendix 4).

Transition variables can affect both the streams of costs and outcomes. It is usually easier to discuss them alongside the outcome variables.

State transition models

Present the transition probabilities of the model, preferably in a matrix. Provide the source of each transition probability and justify the estimate used. Pay particular attention to the transition probabilities that simulate a treatment effect by differing between the proposed medical service and its main comparator. For each transition probability, and for any other time or age-dependent variable, indicate whether it is assumed to be constant or to vary over time, and justify the assumption. If a transition probability is modelled as varying according to time or age, describe how this is achieved in the model.

Where probabilistic cost-effectiveness modelling is presented, list the probability distribution around each variable and justify the selection of each type. For example, gamma or log-normal distributions (i.e. non-negative) could be used for cost parameters, beta distributions for transition probabilities in a control arm, and log-normal distributions for relative risks. For a modelled estimate of incremental effectiveness derived from direct comparison randomised trial evidence, explain how the assumed distribution of the variable reflects the 95% confidence interval around the estimate reported in the trial(s). For each other variable, explain and justify how the selected distribution reflects the extent of statistical imprecision associated with the variable. Also explain and justify each assumed correlation (or lack of correlation) of distributions across the variables.

Additional considerations relating to necessary diagnostic criteria

A number of issues arise when an economic evaluation needs to reflect the impact of requesting that diagnostic tests and/or criteria be specifically used to determine eligibility to start or continue MBS-listed medical service.

Ensure that the costs of conducting tests and/or implementing criteria are included in the economic evaluation and are generated for the population tested, not just the population with positive results. The costs should include assessments that demonstrate that certain individuals do not meet the eligibility criteria and for repeat assessments of these individuals.

Also examine the overall impact of false positive and false negative results on the identification of eligible patients, and/or treatment response on the application of the trial results for the economic evaluation, particularly if the latter are used in any proposed continuation criteria in the requested restriction. This examination of predictive value typically requires a separate presentation of additional information on the reliability, sensitivity and specificity of the relevant tests and/or criteria, both across all trials presented and in regular Australian practice. Because predictive value also varies by varying prevalence, evidence of varying prevalence should also be provided. False positives and false negatives both tend to diminish the ability of the tests and/or criteria to make the incremental cost-effectiveness ratio more favourable than an analysis that does not include the tests and/or criteria that the costs of the diagnostic work-up alone make the ratio less favourable.

When considering the impacts of diagnostic tests, distinguish between health outcomes and non-health outcomes. Affected health outcomes include a risk of harm to individuals examined for the diagnostic test, or a risk of harm that arises from changes in treatment that result from the diagnostic test. Include health outcomes only in the base-case analysis. Consider including any non-health-related impacts in a supplementary analysis.

D5 Results of the economic evaluation

INFORMATION REQUESTS

- Present the cost per course of treatment if the proposed medical service is associated with acute or self-limited treatment, or the cost per year if the proposed medical service is associated with chronic or continuing treatment.
- Present the remaining results of the economic evaluation first in a disaggregated form, then in increasingly aggregated forms. Use discounting as appropriate.
- Present the appropriately aggregated and discounted results separately for outcomes and costs, and separately for the proposed medical service and its main comparator.
- Present separate estimates of the incremental cost and the incremental effectiveness of substituting the proposed medical service for the main comparator.
- For cost-effectiveness and cost-utility analyses, present the incremental cost-effectiveness ratio as the incremental cost of achieving each extra unit of outcome with the proposed medical service substituted for the main comparator (the base case of the economic evaluation).
- Draw a conclusion from the base-case economic evaluation that reflects the degree of uncertainty around the presented incremental cost-effectiveness ratios.

ADDITIONAL INFORMATION REQUESTS IF THE EVALUATION INCLUDES VARIABLES REPORTED IN SECTION C

- Present the results of the three steps described in Sub-section D1 to derive a stepped base-case economic evaluation.
- Identify components of the evaluation that have more important impacts on the incremental cost-effectiveness ratio.
- Assess the strength of the evidence that supports the components with the more important impacts and as the basis for identifying matters for the sensitivity analyses.

The presentation of disaggregated results depends on the methods used to generate the results of the economic evaluation. For example, where possible, present the quantity of each type of resource provided in its natural units as well as its cost valued in dollar terms; and/or present the costs and outcomes associated with each branch in the tree of the decision analysis; and/or each health state where the economic evaluation involves a state transition model.

Health care resource costs

Present the estimated health care resource costs in disaggregated form (i.e. separately for each type of resource provided). The nature of this disaggregation is likely to vary across types of economic evaluations.

For a decision analysis that does not calculate costs and outcomes over multiple intermediary time periods (e.g. a decision analysis that is not a state transition model), estimate and present the number of each type of resource item provided in its natural units at each stage in each branch of each arm of the economic evaluation. Then sum the numbers of each type of resource item in each arm before multiplying by the appropriate unit cost for the resource item. In this circumstance, it is helpful to present a table similar to Table D5.1.

For a comparison across state transition models that calculate costs and outcomes over multiple intermediary time periods (e.g. Markov models), two tables (see Tables D5.2 and D5.3) are needed to summarise this type of information.

First, present in a table the number of each type of resource item provided in their natural units for each health state of the models calculated over the duration of one cycle (this should be constant over any cycle in each model each time the health state is entered). Then multiply by the appropriate unit cost for the resource item before summing to estimate the costs for the health state (see Table D5.2).

Second, present a table that partitions the costs according to their health states across all cycles of the models (see Table D5.3).

Table D5.1 List of health care resource items and summary of cost impacts in the economic evaluation

Type of resource item	Cost for proposed therapeutic medical service	Cost for main comparator	Incremental cost	% of total incremental cost
<i>Pharmaceutical products</i>				
PBS drug form and strength	\$x	\$y	\$x – \$y	z%
Non-PBS drug form and strength	\$x	\$y	\$x – \$y	z%
<i>Medical services</i>				
Type of medical practitioner attendance	\$x	\$y	\$x – \$y	z%
<i>Hospital services</i>				
Hospitalisation admission	\$x	\$y	\$x – \$y	z%
Outpatient clinic	\$x	\$y	\$x – \$y	z%
Emergency department	\$x	\$y	\$x – \$y	z%
<i>Diagnostic and investigational services</i>				
Type of service	\$x	\$y	\$x – \$y	z%
<i>Allied health care services</i>				
Type of allied health consultation	\$x	\$y	\$x – \$y	z%
Total	\$x	\$y	\$x – \$y	100%

Note: For a decision analysis that does not calculate costs and outcomes over multiple intermediary time periods.

Table D5.2 List of health care resource items and summary of cost impacts for each health state in a state transition model

Type of resource item	Number of items in natural unit of measurement	Unit cost	Total cost
Health state 1			
Resource type 1		\$x	\$x
Resource type 2		\$x	\$x
Etc.		\$x	\$x
Total for health state 1			\$x
Health state 2			
Etc.		\$x	\$x

Table D5.3 List of health states and summary of cost impacts included in the economic evaluation

Health state in model	Cost for proposed medical service	Cost for main comparator	Incremental cost	% of total incremental cost
Health state 1	\$x ₁	\$y ₁	\$x ₁ – \$y ₁	z ₁ %
Health state 2	\$x ₂	\$y ₂	\$x ₂ – \$y ₂	z ₂ %
Etc.	\$x _{etc.}	\$y _{etc.}	\$x _{etc.} – \$y _{etc.}	z _{etc.} %
Total	\$x	\$y	\$x – \$y	100%

Calculate and present the present value of the direct health care resource costs for each treatment (i.e. separately for the proposed medical service and its main comparator).

Calculate and present the incremental direct health care resource costs by subtracting the present value of direct health care resource costs of the main comparator from those of the proposed medical service. The incremental costs are therefore the costs of any increase in resource provision minus offsets resulting from any improvement in outcome.

Health outcomes

Provide the estimated present value of the health outcomes in disaggregated form (i.e. separately for the proposed medical service and its main comparator).

Calculate and present the incremental health outcomes by subtracting the present value of the health outcomes of the main comparator from those of the proposed medical service.

For a comparison across state transition models that calculate costs and outcomes over multiple intermediary time periods (e.g. Markov models), also present a table that partitions the outcomes in the models according to their health states (see Table D5.4).

Table D5.4 List of health states and summary of health outcomes included in the economic evaluation

Health state in model	Outcome for proposed medical service	Outcome for main comparator	Incremental outcome	% of total incremental outcome
Health state 1	X ₁	y ₁	X ₁ – y ₁	Z ₁ %
Health state 2	X ₂	y ₂	X ₂ – y ₂	Z ₂ %
Etc.	X etc.	y etc.	X etc. – y etc.	Z etc. %
Total	x	y	x – y	100%

Additional disaggregations of state transition models

Where the economic evaluation involves a state transition model, present model traces (e.g. Markov traces) that plot key outputs on a graph with time on the x-axis against the changing outputs on the y-axis in tabulated or graphical form, or, preferably, both forms. For some state transition models, such as those calculated by Monte Carlo simulations, tracker variables could be used to record the information necessary to construct the model traces. Comment on whether each of the model traces makes sense.

For each arm (i.e. for the proposed medical service and its main comparator) and after each cycle, present model traces that:

- identify the proportions of the cohorts in each health state (both for the increment of each cycle over the previous cycle and as cumulative results);
- correspond to observed data (e.g. a model of a medical service used in oncology that generates life-years gained from disease-free survival can be compared with a Kaplan–Meier curve of overall survival, or a model of a medical condition that generates clinical events can be compared with observed data on the natural history of the medical condition, or a genetic test leading to reduced costs in the next generation); and
- sum the outcomes (e.g. QALYs) and the costs (both for the increment of each cycle over the previous cycle and as cumulative results), discounted as appropriate.

For the increment of the proposed medical service over its main comparator after each cycle, present model traces that calculate the incremental costs, incremental outcomes and incremental cost-effectiveness, each discounted as appropriate. For each of these, present model traces both for the increment of each cycle over the previous cycle and as cumulative results.

Where possible, compare those model traces that correspond with observed or empirical data (e.g. overall survival or partitioned survival) as a means of validating the model. Comment on and explain any differences indicated by this comparison to help validate the model (see below).

Incremental costs and effectiveness

Present the base-case incremental cost-effectiveness ratio calculated as the incremental costs divided by the incremental health outcomes.

If the outcome in the denominator of the incremental cost-effectiveness ratio does not include time as part of the units of measurement (e.g. the outcome is expressed on a per-patient or on a per-event basis rather than a per life-year gained basis or a per QALY

gained basis), then also specify the duration of the economic evaluation when presenting these results (e.g. ‘per extra responder at six months’). This helps in the interpretation of the ratio, because — except when limited to a defined course of treatment — the cost of treatment per patient usually increases over time.

Reflect the degree of uncertainty (see Sub-section D6) around the incremental cost-effectiveness ratios from the presented results when drawing conclusions from the economic evaluation. Avoid terms such as ‘dominant’ and ‘dominated’ except in situations where one alternative both costs less and is more effective than the other under a wide range of plausible assumptions.

Where probabilistic cost-effectiveness modelling is undertaken or a probabilistic cost-effectiveness analysis is based directly on a direct comparison randomised trial, present the distribution of overall results both in a scatter plot on the cost-effectiveness plane and in a tabulated format, including the percentages of the distribution of the results in each quadrant of the cost-effectiveness plane. Also present cost-effectiveness acceptability curves. Avoid over-interpreting these results. For example, unless the data contributing to this analysis are derived directly from individual patient data collected in the context of a direct comparison randomised trial, important sources of non-statistical uncertainty also need to be examined separately from this analysis.

If the incremental cost-effectiveness ratio is based on a disease-specific outcome (i.e. other than extra life-years gained or extra QALYs gained), consider whether this ratio can be compared to a similar ratio known to the applicant that might be related to one or more previous MSAC deliberations. Such previous deliberations might provide a narrower benchmark or frame of reference than the more widely conceptualised ‘league table’ based on the two more widely comparable outcomes above. The precedence value is not necessarily determinative because it is indirect at best and might not capture all elements of an overall comparative cost-effectiveness assessment, let alone the influence of other relevant factors (such as disease severity; see Section F for an opportunity to identify and comment on these). However, a proposed medical service with a less favourable incremental cost-effectiveness ratio in a particular restriction than another comparable medical service and restriction previously rejected is unlikely to be recommended.

On the other hand, a proposed medical service with a more favourable incremental cost-effectiveness ratio in a particular restriction than another comparable medical service and restriction previously recommended, is likely to be recommended. Examples of listed medical services that might provide possible benchmarks include:

- a pathology test that is not widely used due to its perceived disadvantages compared to the proposed test or not being the gold standard (and so the appropriate main comparator for the proposed test is no test); and
- a listed test that has a restriction that is similar to the requested restriction of population for the proposed test (e.g. there might be different thresholds determining eligibility according to risk factors that are specified in both restrictions; see also Section A).

If a claim is made for a change in non-health care resource costs or a change in non-health outcomes such as production changes, present a supplementary analysis with these included (see Appendix 6 for a rationale).

Validating the incremental cost-effectiveness ratio

Consider developing and presenting any approaches to validate the results of a modelled economic evaluation. The comparison of model traces with observed or empirical data (see above) is one such approach where the economic evaluation involves a state transition model. Comment on and explain any differences indicated by this comparison to help validate the model.

Related approaches might compare the output of the model assuming no medical service, with any epidemiological data on the natural history of the medical condition being modelled. Related approaches might also compare the output of the model assuming a particular medical service, with any available long-term longitudinal observational data on that medical service.

Where a model relies on one estimate of treatment effect (e.g. a treatment effect used to transform a surrogate outcome to a final outcome, or a treatment effect on one component of a composite outcome) and there is a comparable estimate of treatment effect on another outcome generated by the model (e.g. the final outcome or another component in the composite outcome), consider using this as a basis to validate the results of the model.

Stepped economic evaluation (requested if the evaluation includes variables derived from Section C)

As explained in Sub-section D1 of these Guidelines, if pre-modelling studies are presented in Section C, a stepped approach is requested to help MSAC gauge the impact of making these modifications on an unmodified trial-based economic evaluation. See Tables D5.5 and D5.6 for further advice on presenting this analysis.

The preferred order of considering the translation of the trial-based economic evaluation (Step 1) is to consider next the impact of applying the treatment effect (Step 2), where applicable. To facilitate this consideration, the structure of Table D5.5 is aligned to the structure of Table D2.1. More flexibility is warranted in considering the impact of extrapolating and transforming the treatment effect (Step 3). Table D5.6 therefore suggests three alternative next steps to combine the results of Step 2 with either an extrapolation step or a transformation step (Step 3a). Each of these represents the incorporation of a possible pre-modelling study; an assessment report need only report the option for Step 3a that is relevant to its economic evaluation. The final row of Table D5.6 incorporates all pre-modelling studies to complete the impacts of translation (application, extrapolation and transformation) of the trial-based economic evaluation into a modelled economic evaluation. The incremental cost-effectiveness ratio should therefore correspond to the base case of a stepped economic evaluation presented in an assessment report.

If it would further clarify the impacts of translation of the clinical evaluation to the economic evaluation, present more steps and/or more detail of each step (e.g. costs for the proposed medical service and the main comparator, as well as the incremental costs).

The three steps also help identify assumptions and approaches to be examined in more detail in the sensitivity analyses. For example, if the main impact is achieved by extrapolating the final outcome over time, discuss the rationale for the important underlying assumptions for the extrapolation, such as an assumption about the duration of treatment effect (continued divergence of survival curves) or an assumption that a difference generated by one point in time is maintained (at which point the survival

curves remain parallel), rather than the more biologically plausible assumption of eventual convergence of survival curves. In this example, it is therefore important that the biological plausibility and validity of the extrapolations are considered (e.g. an assumption of a linear relationship between outcomes and time might not be clinically plausible for many medical conditions).

Consider also the compounding impact on uncertainty of combining these steps to estimate the overall treatment effect on the final outcome in the economic evaluation.

Table D5.5 Assessment of the implications for the economic evaluation of applying the clinical evaluation (Step 1 then Step 2)

Population and circumstances of use	As defined in trial(s) using ITT population	As defined by the requested restriction ^a
Costs		
Costs of treatment involving the proposed medical service	(Trial-based)	(Trial-based) ^b
Costs of treatment involving the main comparator	(Trial-based)	(Trial-based) ^b
Incremental costs	(Trial-based)	(Trial-based) ^b
For each trial-based outcome relied on in the economic evaluation before any extrapolation and/or transformation		
Extent of outcomes with the proposed medical service		
Extent of outcomes with the main comparator		
Incremental effectiveness (with 95% CI)	(From Section B)	
ICER	(Step 1)	(Step 2)
Sensitivity analysis of ICER substituting the upper 95% confidence limit of the difference in outcomes achieved		
Sensitivity analysis of ICER substituting the lower 95% confidence limit of the difference in outcomes achieved		

CI = confidence interval; ICER = incremental cost-effectiveness ratio; ITT = intention to treat

a If there is no need to apply the results of the clinical evaluation, the data in this column should be identical to the data in the adjacent column reporting the incremental impacts using the results for the ITT population.

b Justify any variation in estimate of incremental costs from the trial-based costing.

Table D5.6 Assessment of the implications for the economic evaluation of extrapolating and transforming the clinical evaluation (Step 3)

	Incremental costs	Incremental effectiveness	Incremental cost-effectiveness ^a
For each trial-based outcome relied on in the economic evaluation without further modification	(From corresponding row of Step 2 in Table D5.5)	(From corresponding row of Step 2 in Table D5.5)	(From corresponding row of Step 2 in Table D5.5)
For any trial-based outcome relied on in the economic evaluation <i>with any extrapolation from the time horizon of the trial(s) only</i> ^b	(Based on corresponding extrapolation of duration of treatment, if any)	(If extrapolation is required)	(Alternative Step 3a)
For any important outcome generated for or by the economic evaluation from the trial-based outcome(s) (<i>'transformation of nature of outcome' only</i>) ^c	(Include here any modelled increases in the provision of some resources and any modelled offsetting decreases of others)	(If this approach is used, explain why a presentation here is not possible)	(Alternative Step 3a)
For the final outcome relied on in the economic evaluation generated as a valuation of the trial-based outcome(s) (<i>'value transformation' only</i>)	(Should not change from Step 2 because nature of outcome does not change)	(If this approach is used, explain why a presentation here is not possible)	(Alternative Step 3a)
For the final outcome relied on in the economic evaluation <i>combining any extrapolation from the time horizon of the trial(s) with any transformation of the trial-based outcome(s)</i>			(Completed Step 3 and expected base case) ^d

a With sensitivity analyses substituting the upper and lower 95% confidence limits of the difference in outcomes achieved.

b Justify and explain the methods of the approach taken to align the changes in the incremental costs (or incremental effectiveness) to correspond to the changes in incremental effectiveness (or incremental costs) reported by any pre-modelling study to extrapolate the evidence from the trial(s) to the time horizon of the economic evaluation.

c Where the approach to transforming the nature of the outcome also involves extending the time horizon of the analysis, justify and explain the methods of the approach taken to align the changes in the incremental costs to correspond to the changes in incremental effectiveness reported by any pre-modelling study

d Justify if claiming a different base-case analysis from that defined above.

D6 Sensitivity analyses

INFORMATION REQUESTS

- Present univariate (one-way) sensitivity analyses on all variables using plausible extremes of values, and justify the selection of those extreme values.
- Tabulate all univariate sensitivity analyses alongside the base case.
- Present multivariate sensitivity analyses combining variables shown to be sensitive in the univariate analyses.
- Examine and present the sensitivity of the results of the economic analysis to any changes in assumptions concerning the structure of the modelled economic evaluation that are important but uncertain.

The purpose of a sensitivity analysis is to examine the effect of uncertainty around estimates and assumptions included in the economic evaluation on the results of the base-case economic evaluation. Statistical (probabilistic) uncertainty involves random error and can be reduced by increasing sample size. The many other sources of uncertainty involve systematic error, are harder to identify and cannot be reduced by increasing sample size. For example, they arise in the selection and measurement of information, the specification of the structure of a model, and the plausibility of the implicit and explicit assumptions relied on for the model, particularly in aggregating across the various sources of information.

Univariate sensitivity analyses

The univariate (one-way) sensitivity analyses on all variables should use plausible extremes of values. Justify the selection of the plausible extreme values of each variable. For example, the upper and lower 95% confidence limits of the relevant incremental treatment effect variables reported in direct comparison randomised trials, the considerations summarised or the range of estimates from the available studies of the natural history of a medical condition.

Tabulate all univariate sensitivity analyses alongside the base case. A tornado diagram with incremental cost-effectiveness on the x-axis can be used, where possible, as an efficient and informative way of summarising the results of the univariate sensitivity analyses.

Use the univariate sensitivity analyses to highlight the variables that are important drivers of the economic evaluation. Consider providing a matrix with the effects of variables on various outcomes that differ across the two arms (e.g. in terms of health outcomes, mortality and utility).

The three steps to improve the transparency of the economic evaluation are intended to help identify the basis of plausible extreme values of variables for further examination. For example, when curves have been fitted to time-to-event data to extrapolate the results beyond the duration of observed follow-up, the sensitivity analysis should examine both the uncertainty in fitting the curves for the extrapolation, and the upper and lower 95% confidence limits of the time-to-event results measured within the direct comparison randomised trials.

Multivariate sensitivity analyses

The multivariate sensitivity analyses should combine variables shown to be sensitive in the univariate analyses. Explain the selection of these variables and their combination; for example, varying more than one of the steps to improve transparency at the same time. Present the analyses in tabular and graphical format.

Where a probabilistic sensitivity analysis is provided, also examine the sensitivity of base case estimates of incremental cost, incremental effect and incremental cost-effectiveness to changes in one variable at a time as univariate sensitivity analyses conducted on each variable using plausible distributions.

Sensitivity of the results to changes in the modelled economic evaluation

Examine assumptions concerning the structure of the modelled economic evaluation that are uncertain to assess their importance by the extent to which they affect the results of the evaluation. The three steps to improve the transparency of the economic evaluation might help identify structural issues for further examination.

Similarly, if there is a risk of substantial usage beyond the intended population and circumstances of use defined in the requested restriction, examine the sensitivity of the results to the assumption of usage within these intentions. As discussed in Sub-section D2, this wider population and circumstances would be expected to have demographic and patient characteristics and circumstances that differ from the target population and circumstances. If the intention of the restriction is to limit usage to the population for which the proposed medical service is most cost-effective, these sensitivity analyses should examine the extent to which the incremental cost-effectiveness ratio would become less favourable with increasing usage beyond the restriction. Table D6.1 gives advice on presenting this analysis in a format that is comparable to Tables D2.1 and D5.5.

Table D6.1 Analyses of the implications for the economic evaluation of usage beyond the requested populations and circumstances of use

Population and circumstances of use	As defined by the requested restriction	If use beyond the requested restriction might arise
Incremental costs		
Incremental effectiveness		
Incremental cost-effectiveness ratio		

If a cost-utility analysis is presented, also present the results of the economic evaluation with the utility in all health states set to one to generate the incremental cost per extra life-year gained. This helps identify the contribution of any life extension component to the incremental effectiveness claim.

If discounting has been necessary, the robustness of the results to different discount rates (including a zero discount rate on nonmonetary outcomes alone and on both costs and outcomes) should be tested.

Section E

Estimated extent of use and financial implications

Introduction

The purpose of this Section is to generate the most likely utilisation and financial estimates by requesting a set of budget impact analyses. These analyses are relevant to both the MSAC and the Australian Government. In the event of a positive recommendation by MSAC, the Australian Government needs utilisation and financial estimates to help provide the necessary funds.

Figure E1 shows the epidemiological approach for developing utilisation and financial estimates for a medical service. A market share approach, as used in some cases for applications to PBAC is not applicable for medical services. As the flowchart shows, these are not mutually exclusive. It also helps explain the logic behind the steps that build on the epidemiological basis and that support the preferred format of calculating and presenting these estimates using the utilisation and cost model spreadsheets supplied alongside these Guidelines, based on a standardised Excel 2010-2013 or STATA or Triage or Reference Manager/RevMan workbook. Together with this Section, this preferred workbook format is primarily designed to present the necessary calculations using the epidemiological basis consistently across assessment reports.

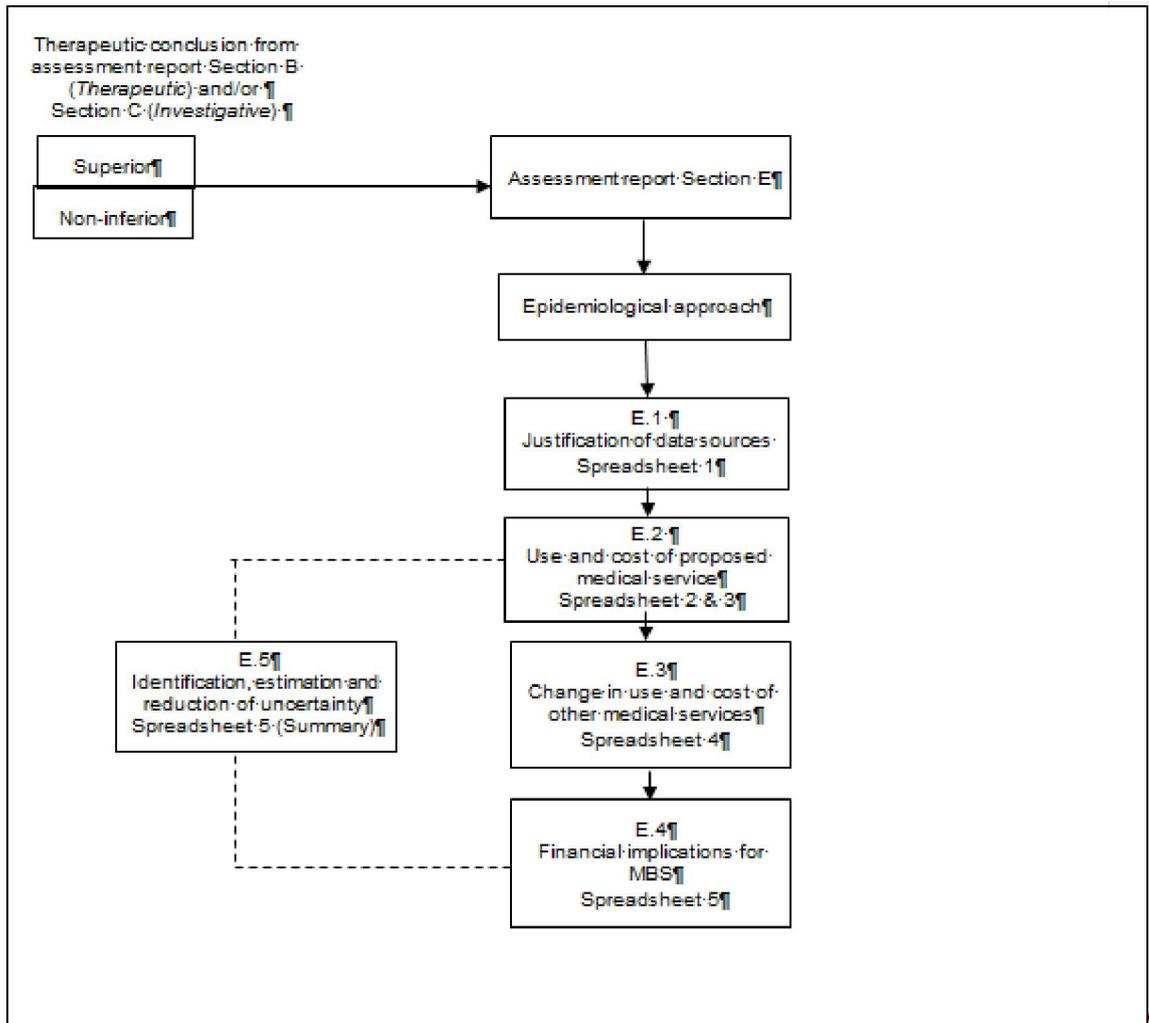
An epidemiological base is usually preferred for generating utilisation and financial estimates if in the prepared assessment report it concludes that, overall, the proposed medical service has an advantage over its main comparator(s). This decision parallels the cost-effectiveness approach that would be taken in Section D of the assessment report. The epidemiological approach first estimates the number of people with the medical condition and then uses several steps to estimate the use of the proposed medical service (see Sub-section E2) and of other medical services in the context of the main indication (see Sub-section E3).

Section E of these Guidelines focuses on the presentation of estimates adopting an epidemiological basis. This approach is informative for some assessment reports prepared - for example, where there is uncertainty in the investigative conclusion or where there is large uncertainty in the expected utilisation. (see Sub-section E5).

Sub-sections E2–E4 request financial analyses relevant to the funding program (e.g. MBS-listing budgets) by only considering health care resources subsidised through those programs. In contrast to the economic evaluation presented in Section D of the prepared assessment report, these financial analyses exclude health outcomes, scale up estimates to assess the impact for the program overall, do not use discounting, and exclude any resource item.

Figure E1

Key information requests for assessment report Section E of a standard assessment for MSAC



The following Sections lay out a preferred stepwise process to generate utilisation and financial estimates. Whenever it is thought appropriate to include an approach that is not requested below, justify the approach in the main body of the assessment report. Whenever it is thought appropriate not to take an approach that is requested below, a particularly strong justification should be provided and, where possible, the alternative approach should be presented separately and in addition to the requested approach.

Where an assessment report seeks listing for more than one indication, present a separate standardised Excel workbook for each indication (refer to Section 5 of the Guidelines). As a final step in each of Sub-sections E4 and E5, these results can be aggregated across the indications.

E1 Justification of the selection of sources of data

INFORMATION REQUESTS

- Where data are available (published or unpublished) from one or more types of data sources:
 - summarise the methods used to obtain the data;
 - present the relevant main results;
 - interpret the findings; and
 - discuss the limitations (including the representativeness of the results) and biases of the method adopted.
- Where data are obtained via one or more studies commissioned for the assessment report:
 - describe the gap in the information to be addressed by the commissioned analysis;
 - summarise the methods used to obtain and analyse the data;
 - present the relevant main results;
 - interpret the findings; and
 - discuss the limitations (including the representativeness of the results) and biases of the method adopted.
- Use Spreadsheet 1 of the standardised Excel workbook to summarise all the background information, primary (non-calculated) variables and assumptions essential to the calculation of results presented in this Section.
- Provide a copy of the data from each published and commissioned study with the attachments to the assessment report. Include the correspondence that requested the data for a commissioned study.

Published data sources

Data sources suitable to the approach taken should be stated and discussed in the assessment report. Data availability for prevalence and incidence is variable, but the best available data should be justified and used where possible. Data sources fall under the broad headings listed in Table E1.1, however, there might be other suitable data sources.⁵ In each case, the methods used should be summarised and the results presented and interpreted, including a discussion of the limitations and biases of the method used.

⁵ See *Sources of Epidemiological Data for Use in Generating Utilisation Estimates* for suggested sources of data that might be suitable for the medical condition, relevant to the assessment report.
<http://www.pbs.gov.au/info/industry/useful-resources/sources>

Sources include data from Australia or overseas, such as MBS or Casemix data, for equivalent medical services that are already listed, and overseas data on the use (in markets similar to Australia) of a proposed medical service that has no comparator that is publically funded in Australia. Where there are multiple sources of data, assess the validity and applicability of both the source and the data in relation to their use in the assessment report's calculations. The demonstration of concordance across multiple data sources of similar validity and applicability is encouraged to reduce uncertainty. Present sensitivity analyses reflecting the variation in the estimates from the available data.

Table E.1.1 Categories of data sources

Disease epidemiological data (provide estimates of prevalence or incidence in the population)
<ul style="list-style-type: none"> • Australian case or mortality registers estimate the incidence or prevalence of a disease • Large, well-designed Australian studies estimate the incidence or prevalence of a disease • Australian national health surveys estimate the prevalence of a disease
Treatment epidemiological data (provide estimates of treated prevalence)
<ul style="list-style-type: none"> • Surveys of the treated prevalence of the disease in Australia • Studies using utilisation databases, including MBS data

Studies commissioned for the assessment report may include data requests to disease registries, established epidemiological studies or ongoing utilisation studies seeking specific analyses. In each case, the information gap to be filled should be clearly described, and the results presented and interpreted, including a discussion of the limitations and biases of the method used.

In the absence of Australian observed data, a range of observed data from overseas sources could be used. When presenting these data, also discuss the applicability of the estimates from an overseas source to the Australian population. In the case of prevalence data, this discussion should further assess the impact of any variations in the subsidy arrangements between overseas health care systems and those in Australia.

Where multiple sources of data are available to address a single assumption or estimate, compare the results, assess their concordance or lack of concordance, and justify the selection of the base-case estimate and the estimates used in the sensitivity analyses. Present a summary table where multiple sources or multiple variables are being compared.

In the absence of observed data, expert opinion might be required (see Appendix 2). A commissioned evaluation of recent usage practice has many similarities with a survey of expert opinion; a distinguishing characteristic might be that a usage evaluation measures what was done, whereas experts are asked to report what they would do now or in the future.

Each time an assumption is required in the absence of data, state the assumption concisely and explain its basis. Describe the nature and likely magnitude of uncertainty for each assumption (see Sub-section E5). Present an examination of the impact of each assumption by altering it in sensitivity analyses.

Spreadsheet 1 ('Background and assumptions')

When using Spreadsheet 1 of the standardised Excel workbook to summarise the data sources, background information, primary (non-calculated) variables and assumptions, it might be helpful, if the analyses are complex, to add one or more other supporting spreadsheets in the workbook to provide more detail, such as identifying the sources of variables relied on and supporting the assumptions made. The remaining spreadsheets, which calculate the estimates (see below), should be fully integrated so that changes to any variable for the purposes of sensitivity analyses flow on appropriately through succeeding calculations to all results.

Copies of data

To allow independent assessment of the data, include copies of the data used (published, unpublished and commissioned) in an attachment to the assessment report. Ensure that the responses in Section E of the assessment report and Spreadsheet 1 provide adequate cross-references of the extraction of all data used to generate the estimates in these analyses from each attached data source (to the level of the page, table or figure number of each source document).

E2 Estimation of use and costs of the proposed medical service

INFORMATION REQUESTS

- Estimate the number of patients with the medical condition targeted by the proposed medical service, the number who would be eligible for the requested restriction and the number of patients likely to use the proposed medical service.
- Use Spreadsheet 2 of the standardised Excel 2003 workbook to calculate the results presented in this part of the Section.
- Estimate the number of times the proposed medical service is delivered in each year over five years (disaggregated into proportions for MBS-listing, and by beneficiary type).
- Estimate the costs for each form of the proposed medical service in each year over five years, multiplying by the relevant unit costs.
- Aggregate these cost calculations for the proposed medical service overall in each year over five years.
- Use Spreadsheet 3 of the standardised Excel workbook to calculate the results presented in this part of the Section.

Numbers of patients

Use of incidence or prevalence data

The choice between using incidence and prevalence data is important in estimating the likely number of patients eligible for the proposed medical service in any one year. This choice depends on the nature of the medical condition and its treatment.

In general, an incidence-based approach is preferred for a treatment of short duration, with 12 months being a suggested upper limit, because estimates should be presented in periods of one year (see below). Examples include an acute self-limiting medical condition, each episode of which is treated with a single course of treatment, and a medical condition that is managed by a single course of treatment given once in a lifetime. Incidence should be estimated on a 12-month basis.

In general, a prevalence-based approach is preferred for a treatment that is to be used for long periods, with 12 months being a suggested lower limit; for example, chronic medical conditions for which treatment is delivered regularly (i.e. without breaks in the standard treatment regimen).

For some treatments, a combination of incidence and prevalence bases might be informative. Examples include intermittent treatment of a series of acute episodes of a chronic medical condition, treatment for which is restricted to each episode and in which the proposed medical service is expected to prolong the duration of disease, including by an extension of expected overall survival.

The first example (regular treatment for chronic medical conditions) is complex, because although the number of patients who have the condition might be determined using an epidemiological approach, the number of presentations for treatment can be more difficult to determine. In the second example (intermittent treatment), allowance for an increase in prevalence might be necessary. If disease duration or life expectancy is expected to increase from fewer than five years in the current situation before the listing of the proposed medical service, it would generally be appropriate to increase the initial prevalence pool estimate on an annual basis by the difference in the 12-month incidence of new patients and the 12-month incidence of cured patients or of deaths. This should be continued either until a new steady state is achieved, with constant rather than increasing prevalence, or until the five-year horizon of the analyses is reached.

Expert epidemiological advice should be sought when estimating prevalence from incidence data or estimating incidence from prevalence data, particularly where there is doubt that the duration of disease has not remained constant over time or where it is not expected to remain constant after the listing of the proposed medical service.

Estimate the number of patients with the medical condition

Estimate the likely number of patients in the current year and in the first year of listing using one of the bases above (incidence or prevalence). These estimates should also incorporate the most probable estimates of patients who are misdiagnosed (i.e. where there might be pressure to diagnose the patient as having the medical condition to be eligible for the proposed medical service and where the differential diagnosis is unclear). Then project the numbers of patients on an annual basis for a total of five years, accounting for population growth and expected changes in prevalence and/or incidence of the condition. If appropriate, more frequent periods (e.g. monthly or three-monthly) could be calculated in the supporting spreadsheets. If so, summarise the presentation of these aggregated data as annual aliquots for a total of five years from listing (Year 1, Year 2, Year 3, Year 4 and Year 5).

Estimate the number of patients eligible for the proposed medical service

Using these annual numbers of patients with the medical condition for Years 1–5, estimate the proportions that would be expected to be eligible to receive the proposed medical service. These estimates should also include the most probable estimate of patients who are misclassified.

Estimate the number of patients likely to use the proposed medical service

Using these annual numbers of eligible patients, estimate the proportions likely to use the proposed medical service in each of the five years. The resulting estimates should reflect the likely share of the proposed medical service compared with the other treatment options currently used for eligible patients.

Spreadsheet 2 – Epidemiology of the disease and patient numbers'

Calculate the above three sets of estimates of patient numbers in Spreadsheet 2 ('Epidemiology of the disease and patient numbers') of the standardised Excel workbook.

Number of times the proposed medical service is delivered

Three elements are involved in translating the numbers of patients likely to be treated to the number of times the proposed medical service is delivered. There is no basis to suggest a preferred order in which they should contribute to the calculations.

The first element is the rate of uptake of the proposed medical service across the five years from listing. If appropriate, shorter periods (e.g. monthly or three-monthly) could be calculated in the supporting spreadsheets. If so, summarise the presentation of these aggregated data as annual aliquots for a total of five years from listing.

The second element is the frequency and duration of treatment involving the proposed medical service. Duration of treatment might be affected by adherence to treatment and rates of discontinuation (e.g. due to poor tolerance or disease progression). Consistent with the information requests in Section D, the estimates should be in terms of the quantities of treatment actually delivered rather than planned. In determining the impact of this element, the variation in duration of treatment between the context of the available randomised trials and probable use of the proposed medical service once listed for MBS funding should be considered. Aspects of this include patient preferences, physician's preferences, switching of proposed medical service, comorbidity in the patients and co-administration of other treatments. Determining estimates of treatment use for the MBS context is therefore based on a number of assumptions and uncertainties that are difficult to quantify; therefore, they should be justified and subjected to sensitivity analyses.

The third element is the mix of forms of the proposed medical service. Where more than one form is specified in response to Section A, there will be more than one product or item listed for MBS funding in Australia to distinguish between these forms, strengths and quantities. The estimates should be disaggregated to the level of the proportions of use of each of these types of the proposed medical service.

Estimate the number of times the proposed medical service is delivered each year over five years by applying these three elements to the patient number estimates from Spreadsheet 2.

Aggregated cost calculations

Estimate the costs to the MBS of the proposed medical service in each year over five years by applying these breakdowns and unit costs and then aggregating each set of cost estimates.

Spreadsheet 3 – Cost of the proposed medical service to the MBS

Calculate the above sets of estimates of administrations of the proposed medical service and costs in Spreadsheet 3 ('Cost of the proposed medical service to the MBS') of the standardised Excel workbook.

E3 Estimation of changes in use and cost of other medical services

INFORMATION REQUESTS

- Identify the other MBS-listed medical services that are likely to be affected by listing the proposed medical service.
- For each proposed medical service, estimate the extent of change in the number of times the proposed medical service is delivered each year over five years (disaggregated into proportions for the MBS and by beneficiary type).
- Aggregate both these cost calculations for the other affected medical services in each year over five years.
- Use Spreadsheet 4 of the standardised Excel workbook to calculate the results presented in this Section.

Medical services likely to be affected by the listing of the proposed medical service

MBS-listed medical services likely to be affected by the listing of the proposed medical service include:

- MBS-listed medical services substituted by the proposed medical service;
- other MBS-listed medical services with decreased usage; and/or
- other MBS-listed medical services with increased usage.

As an initial step, identify and list all MBS-listed medical services that fall into each of these three categories. The list should include those MBS-listed medical services identified in Section A.

Of the three categories, substituted medical services usually have the largest impact on the financial implications of listing the proposed medical service. There would be no substituted medical services if the proposed medical service has no comparators or if it is designed to replace a medical procedure. Where all substituted MBS-listed medical services come from a single group of medical services listed on a cost-minimisation basis, the cost differential of each against the proposed medical service should be similar. However, where the cost differential is expected to vary to an important extent across the substituted medical service, also estimate the breakdown of the proportions of the overall substitution to capture the cost implications of the variation.

Proposed medical services that are listed for MBS funding in Australia, with expected decreased usage after being listed, include those that are co-delivered with substituted medical services, those used to treat adverse outcomes to substituted medical services, and those used to treat the clinical end points that might be reduced after treatment involving the proposed medical service.

Medical services that are expected to have an increased usage after being listed for MBS funding in Australia include those that are co-delivered with the proposed medical service, and those used to treat adverse effects caused by, or outcomes of, the proposed medical service.

The impact of adverse outcomes might have less weight if the information provided in Sub-section B7 shows that they are of insufficient clinical importance to require management with MBS-listed medical services, or if they are similar for the proposed medical service and its major comparators. If there is insufficient information available to include the impact of adverse reactions on MBS expenditure, this should be noted.

Number of times the proposed medical service is delivered

Justify the approach adopted for estimating the extent of change for the forms of each affected medical service, where the approach and calculations involve uncertainty. Use the information provided in Section A and Sub-section E2. Identify and justify any inconsistency between Sections D and E of the assessment report in the identification of MBS-listed medical services that would change as a result of listing the proposed medical service, and the extent of change per patient in the first five years of listing.

Disaggregation of estimates

Disaggregation into proportions for the MBS and by beneficiary type should usually be based on the most recent 12 months of usage data from MBS Online. An exception could be where the expected substitution is for a distinctive subgroup of current use of the substituted medical service(s), in which case the disaggregation should be based on the subgroup.

Costs over five years

Estimate the costs in each year over five years of each of the forms of each of these medical service substituted, decreased and increased on the basis of each of the estimated utilisation changes. For these calculations, use constant prices, make no allowance for inflation and use a zero discount rate.

Aggregated cost calculations

Estimate the cost offsets to the MBS of the other affected medical services in each year over five years by applying these breakdowns and unit costs, and then aggregating each set of estimates by subtracting the costs of substituted medical services and the costs of medical services with decreased usage from the costs of medical services with increased usage.

Spreadsheet 4 – Cost implications to the MBS from substitutions and other increases and decreases

Calculate the above sets of estimates of number of deliveries and costs in Spreadsheet 4 (‘Cost implications to the MBS from substitutions and other increases and decreases’) of the standardised Excel workbook.

E4 Estimated financial implications for the MBS

INFORMATION REQUESTS

- Estimate the net financial implications for the MBS in each year over five years by subtracting the net cost offsets for both the aggregated estimates calculated in Sub-section E3 from the corresponding estimates calculated in Sub-section E2.
- Use Spreadsheet 5 of the standardised Excel workbook to calculate the results presented in this Section.

Spreadsheet 5 – Net cost of the proposed medical service to the MBS

Calculate the two sets of net financial implications in Spreadsheet 5 (‘Net cost of the proposed medical service to the MBS’) of the standardised Excel workbook.

E5 Identification, estimation and reduction of uncertainty

INFORMATION REQUESTS

- In each step of the calculations, assess the sources of uncertainty and distinguish the type and degree of uncertainty in utilisation and financial estimates.
- Where possible, explain the nature of each uncertainty and its impact on the overall estimates.
- Estimate the level of the uncertainty and propose ways to reduce it.
- Provide a separate workbook to generate the results of any calculations (e.g. sensitivity analyses and scenario analyses) to examine the impact of uncertainty. Summarise these in Spreadsheet 5 of the standardised Excel workbook.

Nature of uncertainty

When presenting the most likely utilisation and financial estimates, consider the degree of uncertainty of those estimates. Two types of uncertainty should be distinguished:

- 1) *usage that differs from expectations* — generally arises from uncertainty within and across particular variables in the analysis. Sensitivity analyses should be presented to examine the impact of this source of uncertainty; and
- 2) *usage that extends beyond the restriction* (sometimes called ‘leakage’) — generally arises from uncertainty as to whether the requested restriction would achieve its intended objective in limiting use. Usage beyond the requested restriction raises doubts about the overall cost-effectiveness of the proposed medical service where the intention of the restriction is to exclude its subsidised use in patients for whom that use would not be acceptably cost-effective. Scenario analyses might be relevant to examine the impact of this uncertainty.

Sources of uncertainty

The following lists summarise the factors that could be considered when assessing uncertainties in predicted utilisation patterns and financial implications resulting from the listing of a proposed medical service as requested. The lists are not intended to be prescriptive, but generally reflect factors that have been considered previously by MSAC and may arise from epidemiological data, treatment prevalence data, expert opinion and assumptions used in generating the quantified predictions. Any of these factors might provide information that will increase understanding of the uncertainties present in utilisation estimates. It might be useful to consider these factors explicitly, but not all the factors will apply to all assessment reports. Thus, it might not be necessary to address any or all of these questions for each assessment report, as the uncertainties outlined might be very small or of little importance to the overall cost to the MBS. Therefore, consideration should be given to how relevant each of the factors might be for a particular assessment report.

Factors that could affect the extent of usage within the requested restriction

Consideration of the following factors might provide relevant information on uncertainties within the requested restriction. Some factors might not be relevant in all assessment reports or might have a negligible impact on the overall estimates:

- Promotion might result in greater identification of the proposed medical service, resulting in more health care practitioners considering patients for treatment.
- Indirect media exposure to consumers might result in some consumers being more aware of the proposed medical service and seeking treatment with it. 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 medical service. This could be positive or negative, and could emerge at the time the assessment report is lodged or be expected to occur within five years of listing.
- More health care practitioners and patients might seek treatment if the proposed medical service treats a medical condition for which the alternatives are considered to be substantially inferior to the proposed medical service (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 a requested restriction 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.
- Patients might be treated more often than expected, particularly in the case of medical conditions with episodic manifestations.
- There might be a likelihood of treatment increasing over time.

Factors that could affect the likelihood of usage beyond the requested restriction

Some of the factors listed above might also affect the likelihood of usage beyond the requested restriction. More detailed guidance is given in Section A about ways of designing a restriction to minimise usage beyond its intention, however the following factors might be considered:

- The requested restriction is for a subset of the types of patients who are eligible according to the TGA-approved indication(s).
- The requested restriction is for a subset of the types of patients who were eligible for the randomised trial(s) published for the proposed medical service, or there are randomised trials demonstrating evidence in other medical conditions.
- The requested restriction is for 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 requested restriction is for a subset of the types of patients for whom the applicant plans to promote use of the proposed medical service before or after the listing for MBS funding is implemented.
- The requested restriction is for 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 medical service (e.g. a difficult differential diagnosis, ambiguity in the wording of the restriction, 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 .

Estimating and reducing uncertainty

The following three aspects should be addressed in any consideration of uncertainty:

- the direction of impact on the estimate (underestimate or overestimate);
- the impact on the magnitude of the estimate (small or large); and
- the likelihood that another estimate should replace the base-case estimate (probable or improbable).

Although quantitative estimates of uncertainty are preferred, semi-quantitative assessments may need to be given in many instances. Where the effects of some uncertainties are difficult to quantify, this should be noted. As a general principle, the more sensitive the overall financial implications are to a particular source of uncertainty, the more important it is to minimise that uncertainty.

One way to reduce uncertainty is to use data from multiple sources, where available. Where estimates derived from different sources are concordant, there might be more confidence and therefore less uncertainty in the resulting estimates. Where this is not the case, the disparity between the estimates might contribute to the estimate of uncertainty.

This can be 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).

Summary of calculations

Summarise the results of any calculations (e.g. sensitivity analyses and scenario analyses) to examine quantitatively the impact of uncertainty in Spreadsheet 5 (‘Net cost of medical service to the MBS’) of the standardised Excel workbook. Do not include the supporting calculations in that workbook. If additional calculations need to be explained, a separate workbook should be provided for any analysis other than the base-case analysis (most likely). Spreadsheet 1 (‘Background and assumptions’) of the separate workbook should highlight the differences from the base-case workbook.

Section F

Options to present additional relevant information

Introduction

Over time, a number of issues have arisen that are important for some assessment reports, but are not necessary for all assessment reports. These have included equity principles, ‘rule of rescue’ and other relevant factors that can affect MSAC’s assessment of proposed medical services.

This Section is intended to assist the consideration of such issues in relation to an assessment report. It does not cover all possible issues. Ultimately, an applicant may include in an assessment report any information that is relevant to MSAC’s decision.

F1 Other relevant factors

INFORMATION REQUESTS

- If the assessment report raises any issue relating to equity principles, discuss it in descriptive terms.
- If the assessment report raises any equity assumptions that particularly affects consideration of the cost-effectiveness of the proposed medical service, describe the implications, where appropriate, with reference to a sensitivity analysis.
- If the assessment report makes any claim that the ‘rule of rescue’ is applicable, set out the basis for that claim.
- If the assessment report identifies any other relevant factor not requested elsewhere, discuss it in response to this Section.

Equity principles

From a general policy viewpoint, the MBS promotes fairness in its subsidy arrangements by promoting affordable access to safe, effective and cost-effective medical services. Thus, any listing that is likely to promote particularly, or hinder, these or any other general equity principles should be discussed. For example, if the requested listing of the proposed medical service would raise particular patient affordability considerations, their implications should be discussed.

Equity assumptions

From a technical viewpoint, many elements of an economic evaluation contain embedded equity assumptions (e.g. see utility valuation in Appendix 4). In the rare cases in which such underlying assumptions might be important enough to influence a particular MSAC decision, a description of how the issue affects consideration of the cost-effectiveness of the medical service, and preferably an examination of its impact in a sensitivity analysis, should be sufficient.

Guidance on the 'rule of rescue'

Four factors, which apply in exceptional circumstances, are particularly influential in favour of listing. When all four factors apply concurrently, this is called the 'rule of rescue'. The four factors are as follows:

- No alternative exists in Australia to treat patients with the specific circumstances of the medical condition meeting the criteria of the restriction. This means that there are no suitable medical services 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, the younger the age at which a person with the condition might die or the closer a person with the condition is to death, the more influential the rule of rescue might be in the consideration by MSAC.
- 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 the consideration by MSAC. However, MSAC is also mindful that the MBS is a community-based scheme and cannot cater for individual circumstances.
- The proposed medical service 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 the consideration by MSAC.

As with other relevant factors, the rule of rescue supplements, rather than substitutes for, the evidence-based consideration of comparative cost-effectiveness. A decision on whether the rule of rescue is relevant is only necessary if MSAC would be inclined to reject an assessment report because of its consideration of comparative cost-effectiveness (and any other relevant factors). If MSAC concludes that the rule of rescue is relevant (such as for last-line treatment for terminally ill patients), it will consider whether there is a strong enough case for listing for MSAC to reverse a decision not to recommend listing in the absence of the rule of rescue.

This guidance on the rule of rescue is kept deliberately narrow. Although there are relevant arguments for broadening the guidance, MSAC is concerned that doing this would reduce the relative influence of the rule of rescue when it is applied to a broader set of eligible assessment reports. In other words, the greater the proportion of assessment reports that the rule of rescue is applied to, the smaller its average impact has in favour of listing across the identified assessment reports.

One issue that has arisen concerning the rule of rescue is that a second medical service to treat the medical condition considered to meet the requirements of the rule is not suitable for this consideration. This is because, by definition, the second therapeutic medical service does not meet the essential first factor of the four factors (i.e. that there is no currently alternative intervention). This causes a difficulty if listing of the second therapeutic medical service is sought on a cost-minimisation basis.

Another difficulty is that indiscriminate application of arguments such as the rule of rescue can lead to overall inefficiencies, unless MSAC compensates when considering medical services that clearly fall outside the rule.

Discuss any other relevant factor

If any other relevant factor is thought to be worth emphasising and is not already requested elsewhere for inclusion in the assessment report, discuss it in the response to this Section.

Part III

Alternative clinical evidence for proposed medical services to be considered by MSAC

Section D(i)

Economic evaluation for the main indication: presenting a cost-minimisation approach

Introduction

The purpose of this Section is to present an economic evaluation of substituting the proposed investigative medical service for the main comparator in the context of the listing requested. As already described in Sub-section B8 and D(i)1, the economic evaluation of the proposed medical service initially depends on whether the therapeutic conclusion shows:

- the proposed investigative medical service is superior to the main comparator; or
- the proposed investigative medical service is non-inferior (equivalent) to the main comparator.

If the proposed medical service has been shown to be superior to the main comparator, presentation of the economic evaluation according to Part II, Section D is appropriate. However, if the proposed medical service has been shown to be non-inferior (equivalent) to the main comparator, cost-minimisation analysis is appropriate (or cost analysis under limited circumstances where the proposed medical service is non-inferior to the main comparator, but has a superior safety profile that generates cost offsets from reduced use of health care resources to manage adverse reactions).

Cost-minimisation analysis

A cost-minimisation analysis applies when the proposed medical service is demonstrated to be no worse (non-inferior / equivalent) than other medical services at the same or a lower price. Assuming MSAC accepts the alternative therapies as providing acceptable outcomes in terms of both effectiveness and safety for their cost, a new treatment that offers these outcomes at a lower cost is preferable.

Cost analysis

A cost analysis compares costs only and so 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 medical service is demonstrated to be no worse in terms of effectiveness but to have a superior safety profile compared with the main comparator.

D(i)1 Presentation of a cost-minimisation analysis or a cost analysis

INFORMATION REQUESTS

- Present a cost-minimisation analysis OR a cost analysis.
- Provide copies of all sources of data in an attachment or a technical document (cross-referenced from the main body of the assessment report) and electronic copies of all computer-based analyses.

Cost-minimisation analysis

When the proposed medical service is regarded as non-inferior to its main comparator in terms of both effectiveness and safety, the appropriate type of economic evaluation is a cost-minimisation analysis. That is, the difference between the proposed medical service and the main comparator is reduced to a comparison of costs.

Such an assessment report need only present an abbreviated assessment report Section D, except where there are differences in the costs of delivering the two alternatives. Take particular care to justify any decision to model a difference due to a factor that is excluded in the trials. Only rarely has a model been accepted that contradicts a conclusion from the evidence of direct comparison randomised trials that fail to detect a statistically significant advantage when designed to do so.

If the conclusion of non-inferiority is not also supported by clinical data, the assessment report will be difficult to evaluate.

Cost consequences related to the provision of resources

Listing a non-inferior medical service might have cost consequences related to its differing mode of administration. These have sometimes arisen if the proposed medical service and its main comparator are available in different forms. If this applies in an assessment report, identify the types of other resources affected, estimate the extent to which the quantity of each type of resource provided would change (in its natural units of measurement) following a listing, and multiply by the relevant unit costs. Aggregate this with the medical service cost impact to estimate the net cost impact within the cost-minimisation analysis.

Cost analysis to reflect cost consequences related to management of adverse reactions

If the proposed medical service is demonstrated to be no worse in terms of effectiveness but to have a superior safety profile compared with the main comparator, the generally preferred approach would be to compare also the improved health outcomes due to this safety advantage with the associated incremental costs in a cost-consequence, cost-effectiveness or cost-utility analysis (see Part II, Sub-section D1). However, cost analyses have sometimes been presented and found to be acceptable in these circumstances. The cost analysis could be presented to quantify a claim that the costs offsets from the reduction in resources provided to treat the adverse events avoided are sufficient to reduce the incremental cost to zero or a negative value. In a cost analysis, the extent of the health impact would not be assessed other than to estimate the extent to which the provision of the identified types of other resources is reduced i.e. the economic claim could be that, at the MBS fee requested, the overall cost of treatment with the proposed medical service is the same or less than the overall cost of treatment with the main comparator.

Sources of information and electronic calculations

Separately provide copies of the original sources of all data (beyond those already presented in Section B of the assessment report) or expert opinion used in the model in an attachment or technical document. Cross-reference the extraction of data from each source to the level of the page, table or figure number of the source document.

Also, to enable independent verification of each analysis, provide an electronic copy of any computer-based calculations of the analysis (refer Section 5).

Appendices

Appendix 1 Relevant factors influencing provision of advice by MSAC

This Appendix provides lists of quantitative and qualitative factors that are relevant to the provision of advice by MSAC

Table A1.1 Examples of factors that are more readily quantified

Relevant factor	Description
Comparative safety	Presented as safety of the service under consideration compared with the appropriate comparator(s) when used in the target population and setting
Comparative health gain	Presented as effectiveness This is assessed in terms of both magnitude of effect and clinical importance of effect.
Comparative cost-effectiveness	Presented as cost-minimisation analysis or incremental cost-effectiveness ratios (including incremental cost-utility ratios). Includes a consideration of comparative costs, including the full spectrum of cost offsets
Patient affordability in the absence of MBS subsidy	Presented as cost/patient/course for acute or self-limited treatment, or cost/patient/year for chronic or continuing treatment Calculations for episodic treatment are more difficult.
Financial implications for the MBS	Presented as the projected annual net cost to the MBS
Financial implications for government health budgets	Presented as the projected annual net cost/year

Table A1.2 Examples of factors that are less readily quantified

Relevant factor	Description
Uncertainty	<p>The extent and nature of assumptions compared with the extent and nature of data-sourced evidence are important considerations.</p> <p>The presence of uncertainty increases the hesitation involved in making the decision, increasing the likelihood that a risk-averse decision will be made from the perspective of the MBS.</p> <p>Issues which may impact the decision of MSAC include (but are not limited to) uncertainty related to:</p> <ul style="list-style-type: none"> • the direct comparison randomised trial evidence; • an indirect comparison of two or more sets of randomised trials involving one or more common references; • the non-randomised study evidence; • translating the direct comparison randomised trials to the listing requested; • translating an indirect comparison of randomised trials or non-randomised studies to the listing requested; • the economic evaluation; • cost minimisation; • the utilisation and financial estimates; or • the plausibility of the valuation of health outcomes.
Equity	<p>Affordable access is a central policy principle of the MBS and is considered alongside the economic evaluation.</p> <p>There are many implicit equity and ethical assumptions in the use of quality-adjusted life-years gained; for example, age and socioeconomic and geographical status. This means that these assumptions might also need to be reconsidered alongside the economic evaluation on a case-by-case basis.</p>
Presence of effective alternatives	<p>This distinguishes between:</p> <ul style="list-style-type: none"> • an active comparator or placebo for add-on treatment; and • a placebo for no active intervention. <p>It also helps to define the clinical need for the proposed medical service.</p>
Severity of medical condition treated	<p>This depends on any restriction requested.</p> <p>The emphasis here is only on the nature and extent of disease as it is currently managed (see Part II, Sub-section A2).</p>
Ability to target treatment with the proposed medical service precisely and effectively to patients likely to benefit most	<p>If the proposed medical service appears not to be acceptably cost-effective across the broader population, it might become acceptably cost-effective in patients likely to benefit more than the average (assuming costs of the treatment do not increase proportionally).</p> <p>Claims of benefits greater than the average result from the ITT analysis should be supported by appropriate trial evidence.</p>

Appendix 2 Expert opinion

This Appendix outlines the situations in which expert opinion can be used, and explains how expert opinion should be collated and presented in an assessment report.

Expert opinion, where sought, will be considered in conjunction with advice provided to MSAC by the PASC and the ESC during the assessment report assessment and evaluation stages, respectively.

A2.1 Uses of expert opinion

Expert opinion is not a substitute for sound scientific evidence. Therefore, expert opinion is only considered where there are no observed data available, or where such data addressing the matter for which expert opinion has been sought are unlikely to become available in the near future. Observed data might come from randomised trials or non-randomised studies, including from cross-sectional studies or case studies. Expert opinion can also supplement observed data; for example, to review the likely representativeness to the national level of a cross-sectional study conducted in a single locality or in another country. Such supplementation will help the interpretation of observed data, and therefore reduce its uncertainty.

Expert opinion can be useful in several aspects of preparing assessment reports for MSAC; for example, to help:

- define the clinical need for the proposed investigative medical service and thus the context of its use by defining the proposed medical service's place in treatment in terms of the main indication(s) based on what should be recommended (see Part II, Section A), and the main comparator(s) and clinical management algorithms based on what is likely to change (see Part II, Section A);
- interpret the clinical importance and patient relevance of the outcome measures reported in the trials (see Part II, Section B);
- modify the patterns of resource use and, very rarely, the clinical outcomes measured in studies conducted in different settings, such as in other countries (see Part II, Section B);
- predict which resources would be used and how often each would be used to manage outcomes reported in the clinical studies that inform the evaluation in Section B, but not followed up (see Part II, Section C);
- identify the proportion of patients with the medical condition who would meet the eligibility criteria established by the requested restriction (see Part II, Section E);
- predict the proportion of patients within this eligible population who would take the proposed medical service (see Part II, Section E);
- predict the rates of uptake of the proposed medical service (see Part II, Section E); and
- predict the extents of substitutions, increases and decreases of other medical services that are MBS-listed (see Part II, Section E).

A2.2 Presenting expert opinion

INFORMATION REQUESTS

- Present expert opinion as a technical document or an attachment to the assessment report, with clear cross-references to the relevant Sections of the main body of the assessment report.
- Justify the need for expert opinion.

If expert opinion is included, its use should be justified in the introduction of the relevant Section of the application report. Include a clear rationale for, and the aims of, eliciting the expert opinion. Where expert opinion is used to fill in a gap in information, describe the nature of this gap clearly and indicate the steps that have been taken to address the gap, such as a literature search.

A2.3 Describing the collection and collation of expert opinion

INFORMATION REQUESTS

- Describe and justify the approach chosen to elicit expert opinion.
- Describe the methods used to obtain and collate the opinions, and summarise the opinions together with the extent of any variability in the opinions (see Table A2.1).
- Indicate how the opinions have been used in the main body of the assessment report and justify the approach used in the sensitivity analysis (see Part II, Sections D and E) to reflect any variability in the opinions obtained.

Using a well-designed methodology to elicit expert opinion helps to reduce uncertainty. The methods used might vary from large, published questionnaires and surveys with statistical analysis to a summary of interviews with a panel of clinical experts. Expert opinion might be presented as qualitative or quasi-quantitative information.

There are many approaches to addressing information gaps. The choice of the preferred approach might be influenced by the availability of existing surveys, small numbers of health care practitioners with appropriate expertise and resource limitations (e.g. time). Options for primary collection of opinions include interviews, focus groups, self-administered questionnaires and telephone surveys. If the survey is to determine what changes a health care practitioners might make to their prescribing behaviour, ensure that the hypothetical future scenario is clearly detailed.

When summarising the opinions and their variability, interpret the findings and discuss the limitations and biases of the method chosen. Indicate how the opinions have been used in the main body of 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. Where expert opinion is used to modify estimates from randomised trials or non-randomised studies, particularly estimates reported in Part II, Sections B or C or any other input into the economic evaluation in Part II, Section D, compare the results and justify the modification. Present a summary table that compares multiple sources or multiple variables. Table A2.1 in Appendix 2 provides guidance on the details that should be included.

Table A2.1 Methods to collect and collate expert opinion

Information to be provided	Notes
The criteria for selecting the experts	Prefer: <ul style="list-style-type: none"> • a random or comprehensive set of health care practitioners likely to deliver the proposed therapeutic medical service, OR • the appropriate medical specialty group.
The number of experts approached ^a	
The number of experts who participated ^a	Assess whether the extent and characteristics of the non-responders are likely to diminish the representativeness of the opinions provided, compared with the intended sample approached.
Declaration of potential conflict(s) 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 and the contractor, and the arrangements made between the contractor and those whose opinions were sought.
The 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, it might help to ask the experts to define the comparative clinical place of the proposed medical service and the main comparator based on this background information. Including the experts' definitions in the technical document or attachment would allow an assessment of the consistency of the background information with the evidence provided in the assessment report.
The method used to collect the opinions	For example, were the experts approached individually or was a meeting convened? Was any incentive used to maximise responses?
The medium used to collect the opinions	For example, was information gathered by direct interview, telephone interview or self-administered questionnaire?
The 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. <p>On a question-by-question basis, assess:</p> <ul style="list-style-type: none"> • 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 to be made, include in the technical document (or as an attached copy) the questionnaire or an outline of the interview questions.
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.

Information to be provided	Notes
The 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.
<p>Whether all experts agreed with each response, and, if not:</p> <p>(i) the approach used to finalise the estimate; and</p> <p>(ii) the approach used to present the variability in the opinions.</p>	<p>(i) For example, the majority opinion or a Delphi technique could be applied; for quantitative results, point estimates (such as the mean, median or the mode) could be presented.</p> <p>(ii) For example, present the range of opinions including common and outlying views expressed; for quantitative results, measures of variance (such as confidence intervals, range, centiles) could be presented.</p>

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 also ask the reason 'why' (e.g. explain why would you do this, explain why is this important). Conveying these reasons alongside expert opinion-based estimates might help improve their acceptability, particularly if a small group of experts has been approached. Including these explanations in the technical document or attachment would allow the opinions to be assessed on the basis of the underlying reasoning, rather than only depending on the authority of the experts.

Appendix 3 Assessment of non-inferiority

A3.1 Introduction

Non-inferiority means that, in terms of effectiveness, the proposed investigative medical service is no worse than its main comparator. It is used to support a claim of equivalence, because it is not adequate to demonstrate the *absence* of a statistically significant difference between the treatments to claim equivalence; such a lack of a significant difference might occur when the trials are too small to demonstrate a real difference in the effects of the compared services. The appropriate comparison to present is the point estimate of the difference with its 95% confidence interval. This allows MSAC to assess whether the confidence interval contains the minimal clinically important difference.

Thus, an assessment report should support any conclusion for non-inferiority with the information contained in its assessment report Sections as referred to below.

A3.2 Service delivery information

As part of the information provided in Section B of the assessment report, ensure that the service delivery relativity used in the clinical studies presented is appropriate.

A3.3 Non-inferiority threshold

As part of the information provided in Section B of the assessment report, explain and justify on clinical or other grounds the value of the non-inferiority threshold difference in treatment effect between the proposed medical service and its main comparator. Show how a difference greater than this nominated non-inferiority threshold difference would be clinically important. A specifically designed non-inferiority direct comparison randomised trial would have specified a non-inferiority threshold in its power calculation, and so might have provided one or more grounds to justify this threshold as a pre-specified minimal clinically important difference (MCID). Demonstrate that a systematic approach has been taken in the search for relevant and appropriate references to support the nominated threshold and provide the supporting citations, including any references to one or more regulatory agencies that might have provided guidance on any such thresholds in medical conditions similar to the proposed main indication.

If the basis of the clinical evaluation is an indirect comparison of randomised trials and the nominated non-inferiority threshold relates to an absolute comparison (e.g. absolute risk difference or weighted mean difference) rather than a relative comparison (e.g. relative risk or odds ratio), discuss the issues raised by relying on an indirect comparison of the difference between absolute treatment effects rather than on an indirect comparison of the ratio of relative treatment effects.

A3.4 Method of analysis

Also as part of the information provided in assessment report Section B, indicate whether the analysis of each key study was conducted on a per protocol basis (which is appropriate for an analysis in support of a conclusion of non-inferiority, because it helps examine any impact on the conclusions of losses to follow-up or poor compliance), as well as the standard ITT basis (which is the generally preferred basis for an analysis).

If one or more specifically designed non-inferiority clinical studies are available, also describe the primary analysis of non-inferiority in detail for each such study, including the pre-specified non-inferiority threshold (or MCID) used in the power calculation and whether the preferred per protocol basis rather than the ITT basis was used in the context of this non-inferiority analysis. Comment on any differences in the pre-specified non-inferiority thresholds across these studies and with the nominated non-inferiority threshold.

For any direct comparison randomised trial that was not designed as a non-inferiority trial, also describe its primary analysis in detail, including the pre-specified MCID used in the power calculation.

A3.5 Presenting an assessment of non-inferiority

Assessing non-inferiority based on an indirect comparison of randomised trials

As part of the information provided in response to Section B, present the results of each comparative analysis using, where possible, both the per protocol and the ITT basis of each trial with their 95% confidence intervals in a way that allows for direct comparison with the nominated non-inferiority threshold identified. Comment on any differences between the results for the per protocol and ITT populations. Where there is more than one trial reporting the same outcome, statistically combine these results using the random effects method and, where possible, both the per protocol and the ITT basis. Report each result with its 95% confidence interval in a way that similarly allows a comparison with the nominated non-inferiority threshold. Comment on any differences between the results for the per protocol and ITT populations. If the per protocol basis differs across trials, justify the approach to resolve this in the meta-analysis.

If one or more specifically designed non-inferiority direct comparison randomised trials are available, also report the results and stated conclusion of the primary analysis of non-inferiority for each such trial. Report whether the entire 95% confidence interval of the treatment effect between the two medical services is more favourable to the proposed medical service than the pre-specified non-inferiority threshold corresponding to the proposed medical service being less effective. If so, there is statistical support to the conclusion of non-inferiority based on an appropriate pre-specified trial design.

If the primary analysis of a specifically designed non-inferiority direct comparison randomised trial does not present the 95% confidence interval, and/or adopt a per protocol population basis for the analysis, and/or compare this interval with the non-inferiority threshold identified, then present the results, where possible, using the per protocol and the ITT basis of each trial with their 95% confidence intervals in a way that allows for direct comparison with this threshold for non-inferiority. Discuss whether these results might influence the conclusion of the primary analysis of the trial.

For any direct comparison randomised trial that was not designed as a non-inferiority trial, also report the results of the primary analysis as pre-specified. Report whether the entire 95% confidence interval of the treatment effect between the two medical services is more favourable to the proposed medical service than the pre-specified MCID corresponding to the proposed medical service being less effective. If so, there is post hoc statistical support to the conclusion of non-inferiority. Investigate whether the conclusion of non-inferiority is impacted by a comparison of an analysis conducted on a

per protocol basis and/or whether the 95% confidence intervals compared with the non-inferiority threshold identified would modify this conclusion. Report these investigations.

Supplementary analyses might be helpful to support conclusions of non-inferiority that have to rely on primary outcome analyses that were not adequately powered to assess non-inferiority. Base these supplementary treatment comparisons on the results for secondary outcomes that are known to be most responsive to change.

Assessing non-inferiority based on an indirect comparison of randomised trials

The general approach described above for direct comparison randomised trials needs to be adapted for an indirect comparison of randomised trials. Report the point estimates for the indirect relative treatment effect with their 95% confidence intervals in a way that allows for direct comparison with the nominated non-inferiority threshold for inferiority. Report whether the entire 95% confidence interval of the treatment effect between the two medical services is more favourable to the proposed medical service than this non-inferiority threshold corresponding to the proposed medical service being less effective. If so, there is indirect statistical support to the conclusion of non-inferiority.

Where possible (and appropriate noting that there is no basis for a pre-specified non-inferiority design for an indirect comparison of randomised trials), provide additional investigations and supplementary analyses as described above for direct comparison randomised trials.

A3.6 Assessing comparative harms in the context of non-inferiority

As part of the information provided in Section B of the assessment report, examine whether the extended assessment of comparative harms also supports a conclusion of non-inferiority.

A3.7 Interpretation of the clinical evidence

As part of the information provided in Section B of the assessment report, discuss any results to support a conclusion for non-inferiority in the context of the similarity or otherwise, of the mechanism of action(s) of the proposed medical service and the main comparator.

If providing a narrative conclusion in relation to whether the proposed service is non-inferior to the main comparator, provide justification why a statistical conclusion in this regard cannot be reached or is not feasible. For example, for Applications that predominantly contain lower level evidence, statistical proof on non-inferiority may not be feasible. However, where the quality of the evidence base is such that it lends itself to statistical methods to come to a conclusion of non-inferiority, these methods as opposed to narrative conclusions is preferred.

Appendix 4 Utility valuation of health outcomes

A4.1 Use of health-related QALYs gained and cost-utility analysis

The QALY is a measure of adjusted survival time where the adjustment is by means of health-related quality-of-life preference weights derived for specific health states. Expected survival time in each of these health states is adjusted using the preference weights and then summed across the duration of survival to generate expected QALYs gained. The use of preference weights distinguishes QALYs from other quality-of-life measures.

The QALY has become widespread as a measure of health outcome in the economic evaluation of medical services. The key characteristics of the QALY are as follows:

- it combines extension of life and quality of life in a single index that allows comparison across medical services;
- the utility weight index measures strength of preference on a cardinal index anchored on a 0 to 1 interval of death to full (perfect) health, with equal intervals measured in such a way as to have equal value and an allowance for the existence of health states perceived to be worse than death (i.e. <0); and/or
- the utility weights that underpin the QALY measure are based on a sample of individual preferences. These preferences are obtained in a way that involves a trade-off between quality and quantity of life. This provides some validity to the QALY as representing societal trade-offs and therefore social values.

The implication of using this scale is that one year of life in full health is counted as one QALY. Even though one year of life in normal health is less than one QALY, this does not necessarily mean that all incremental QALY gains are numerically smaller than incremental life-year gains. This is because incremental QALY gains can also encompass the possibility of improving quality of life, and such improvements can happen for a long period before any improvement in survival happens.

Theoretically, at least, the QALY provides a measure of health outcomes that is comparable across medical services. This form of analysis should therefore be considered whenever it is appropriate to the outcomes of the proposed medical service. However, many concerns over the estimation of QALYs have been documented.

Guidance on when a cost-utility analysis should be presented is provided in Part II, Sub-section D1.

Other relevant factors (see Part II, Section F and Appendix 1) should be considered alongside, not within, a cost-utility analysis. These include 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. Therefore, an assessment report should draw these issues to the attention of MSAC where this is thought important and relevant.

A4.2 Obtaining utility weights

Several approaches to obtaining utility weights are discussed in these Guidelines:

- using a MAUI in a direct comparison randomised trial;
- creating scenarios to indirectly elicit utility weights;
- directly eliciting utility weights in a randomised trial;
- obtaining a sample of patients matched to trial participants and eligible patients, and using a MAUI;
- mapping results of other quality-of-life instruments to the utility weight anchors of a 0 to 1 interval of death to full (perfect) health; and
- reporting utility weights from published sources.

The generally preferred method of measuring QALYs is by the repeated application of a valid, reliable and responsive MAUI questionnaire to participants in a direct comparison randomised double-blind trial, together with the application of an appropriate scoring algorithm.

However, it is recognised pragmatically that such instruments are not routinely included as an outcome measure in many trials, so it is anticipated that there will be a lag time before this preference can be met routinely. It is also recognised that in many cases it will be necessary to attach utility weights to health states that are not observed within a trial; for example, because they are the result of events that occur outside the trial time frame. Accordingly, guidance is also provided on alternative approaches (see Sub-sections A4.4 and A4.5 of this Appendix). In some circumstances, it is possible that an alternative approach would be preferred to the use of a trial-based MAUI (see Sub-section A4.4 of this Appendix).

Post-trial transformation to estimate preference weights ('utilities')

Preference weights are preferably generated directly from a trial using MAUIs or might subsequently be elicited with the aid of scenarios. Several other approaches have been presented in major assessments, and are discussed and assessed briefly below in Sub-section A4.5 of this Appendix. MAUIs and scenario-based elicitation of preference weights are further assessed in Sub-sections A4.3 and A4.4 of this Appendix, respectively.

MAUIs (multi-attribute utility instruments)

MAUIs have three defining elements:

- a generic health-related quality-of-life instrument. This element of a MAUI is a descriptive system (a questionnaire containing a set of items or statements with multiple response categories) that provides a description of the health-related quality of life of each respondent;
- a scaling technique, such as time trade-off (TTO) or standard gamble (SG). This is used to derive preference-based rankings for a sample of the health states covered by the descriptive system; and

- a model, which is used to extrapolate from this sample to generate cardinal weights for all health states covered by the descriptive system (i.e. to develop a preference-based scoring algorithm for the MAUI). Both mathematical and statistical models have been used to provide utility weights for any health state that can be described by the instrument in terms of its dimensions and levels. For these utility weights to be meaningful for an economic evaluation, the scaling technique must reflect the trade-offs that individuals are willing to make between health outcomes.

Together, these elements generate the unique advantage of trial-based measurement with a MAUI, which is that the direct observation of the actual health states experienced in the trial can be used to generate utility weights in an acceptable way using utility scores of the health states that have been generated in a separate population-based study.

Therefore, it is the combination of these three elements that enables acceptable post-trial transformations to estimate utility weights (see Sub-section A4.3 of this Appendix).

A4.3 Trial-based utility valuation of health outcomes

Measurement of QALYs using a trial-based MAUI

For MAUIs, the measurement of the health state happens in the trial itself, which enables more accurate and unbiased measurement of the health states as experienced by the patients receiving the relevant treatments. The valuation step is then inferred using an acceptable scoring algorithm, which means that the valuation is conceptually and practically separated from the assessment of the particular disease or treatment, and therefore not subject to bias.

To maximise comparability across assessment reports, it would be ideal to request that a single ‘off-the-shelf’ MAUI be used in randomised trials across all assessment reports presenting a cost-utility analysis. Criteria to guide the selection of such an instrument include that it is valid, reliable and responsive, and that it uses an acceptable scoring algorithm and an acceptable preference elicitation technique. However, in practice, no single MAUI has demonstrated unequivocal superiority against all the others and no single MAUI has been universally accepted. There is also debate about whether generic MAUIs are sufficient to capture all important disease-specific factors that might be relevant for particular disease pathways and treatments. The advantages and disadvantages of trial-based MAUIs are discussed further below.

Advantages of relying on trial-based MAUI data

Trial-based MAUI data has the following advantages:

- (a) it promotes comparability across cost-utility analyses;
- (b) it minimises bias by eliminating the need for an analyst intermediary;
- (c) it can appropriately minimise observer bias by assessing the subjective outcome of health-related quality of life under appropriate blinded conditions;
- (d) it minimises the information asymmetry of the health state being assessed because the trial participant is directly measuring the health-related quality of life of the health state as it is being experienced;
- (e) it applies the scoring algorithm of the general population (which can minimise a source of uncertainty if this was elicited in an Australian population or possibly from socioeconomically similar countries with similar life expectancy) to take responses

from the MAUI questionnaires to generate utility weights using an acceptable technique. In other words, the utility scores in the scoring algorithm have been elicited separately from the reporting of the responses in the trial context for each MAUI. The utility weights are calculated by a validated linkage between the response from the MAUI questionnaire in the trial and the utility score inferred for that response from respondents in the general population using the scoring algorithm;

- (f) as a direct translation, it minimises the number of steps between the direct trial-based measurement of health-related quality of life and its valuation;
- (g) it estimates some of the distribution and heterogeneity variation of health states in a population;
- (h) it maintains a fixed period of assessment to which the MAUI applies;
- (i) repeatedly applying the MAUI during the trial allows for direct conversion into the net present value of the future flow of realised QALYs gained and incremental QALYs gained and might provide a basis for extrapolation beyond the horizon of the trial;
- (j) it provides a benchmark against which to compare any more specific elicitation of preferences presented as supplementary evidence (e.g. using a scenario-based approach; Sub-section A4.4 of this Appendix);
- (k) it provides advantages for applicants and analysts in terms of time and cost to assess the appropriateness of using an acceptable 'off-the-shelf' MAUI in a trial;
- (l) it provides efficiency advantages for respondents and analysts, because no MAUI developed so far takes more than five to eight minutes to complete when self-administered (and less when using computer-based, interviewer-administered questionnaires) and because analysis of each of the main MAUIs is well developed;
- (m) the main MAUIs have been developed with the objective of having international applicability, so it is anticipated that this preference for trial-based MAUI utility weights will have increasing relevance over time to the multinational trial programs for new medical services.
- (n) it is possible to conduct an independent and peer-reviewed verification of any preferred MAUI — including its reliability, validity and responsiveness, the clinical importance of any differences detected by the instrument, and other desirable psychometric properties; and
- (o) the use of a consistent MAUI would allow replication (and potentially meta-analysis) of results across similar direct comparison randomised trials conducted between the proposed medical service and its main comparator.

Disadvantages of relying on trial-based MAUI data

Trial-based MAUI data has the following disadvantages:

- (a) the MAUI might be relatively insensitive to the patient-relevant outcomes affected by the proposed medical service, particularly if its main treatment effects or the impacts of the medical condition do not fall within the domains examined by the MAUI. This interpretation of the results needs to be assessed against the possibility of a true negative (i.e. that the proposed medical service has no overall perceptible incremental effect on utility; see also Sub-section A4.4 of this Appendix). The MAUI should therefore be demonstrated not to fit the context of the proposed

- medical service and the medical condition by comparing the results from the MAUI with an accepted nonutility quality-of-life instrument, such as the SF-36;
- (b) it is unlikely that, in the near future, a randomised trial would be designed to have the MAUI as its primary outcome. The trial might therefore be underpowered to detect a difference using the MAUI. As with all secondary outcomes, the results of the MAUI would need to be assessed with reference to the conclusion from the primary analysis of the trial; and
 - (c) trial participants might not be directly representative of the population for whom listing is requested, although an assessment of the distribution and heterogeneity of the results of this outcome might provide a basis for applying them to the targeted population.

Trial-based direct elicitation of utility weights

Conceivably, direct methods might be used within a trial to ask patients to value their current health state at baseline (or over a recent period of time at baseline), and at one or more time points during the trial follow-up (or over a recent period of time at each time point). Advantages (a)–(d), (f) and (h) listed above would also apply to trial-based direct elicitation of utility weights.

The main disadvantage for direct elicitation in the trial setting is the time horizon assumption for TTO or SG (i.e. the trial participant is required to answer a hypothetical question assuming that they remain in the current health state for the rest of their life expectancy). In a scenario-based setting, the entire framework is hypothetical, so there is less risk of any distortion arising from the respondent first having to conceptualise what it might mean to remain in the current health state for a prolonged period.

This approach might also raise potentially important issues to do with adjusting utility weights for groups of patients in certain disease groups (e.g. quadriplegics) and with different adaptations. The defined range of a utility scale is full health (1) to death (0), but people with cancer and other diseases adapt (or adjust up) their estimate of utility closer towards 1 — such people's 'normal health' might be considerably less than 1, but they adapt up to 1. This potentially biases against the allocation of further health resources (so-called double jeopardy). Some groups, when making the adjustment, could also eliminate their capacity to benefit.

Presenting trial-based direct elicitation and results

If utility weights have been directly elicited in a randomised trial, provide details of the method used and justify the selection of the approach taken (e.g. SG or TTO; interview-based and/or computer-based). The same considerations for the design of the preference elicitation task apply in this context as in a scenario-based approach (see Sub-section A4.4 of this Appendix). Report and assess the results as for MAUIs, above.

A4.4 Scenario-based utility valuation of health outcomes

Background

As discussed in Sub-sections A4.2 and A4.3 of this Appendix, obtaining utility weights using a MAUI within the context of a direct comparison randomised double-blinded trial is the preferred method. This Section of this Appendix presents a less preferred alternative, because there is an expected lag time before most major assessments would

be able to report utility weights on this basis. Furthermore, given that most randomised trials are designed overseas, few randomised trials would be conducted primarily to ensure that useful economic information is generated from this preferred source of evidence for MSAC and similar decision makers.

An assessment report might seek to justify the inclusion of a scenario-based approach to valuing health states in utility weights as supplementing trial-based utility weights. Alongside this justification for providing these supplementary estimates, present both sets of methods and results, and comment on the interpretation of the results compared to each other. As with the interpretation of the results of any measure of health outcomes, any claim for an improved sensitivity in quantifying the utility weight of smaller advantages needs to be assessed against the possibility of a true negative (i.e. that the proposed medical service has no overall perceptible incremental effect on utility; see also Sub-section A4.3 of this Appendix). Document the evidence that supports any claim that any difference in results between trial-based utility weights and scenario-based utility weights is attributable to the special characteristic of the health state and not some idiosyncrasy in the utility measurement procedures that have been adopted. This would help justify any apparent diminution in comparability across assessment reports that provided trial-based utility weights. Similarly, if using a scenario-based utility valuation to capture the impacts of health outcomes only occurring beyond the horizon of the trial, document the evidence that supports any claim that the scenario-based utility weights reflect the trial-based utility weights (e.g. by including one or more health states captured and valued within the trial as part of the scenario-based utility valuation study).

Other situations where a scenario-based approach might supplement trial-based utility weights include those in which:

- the health states are associated with quantitatively important ‘ex ante’ anticipated factors (in which one or more elements of the health state are anticipated rather than experienced, so that concepts such as anxiety, risk aversion, fear, hope or dread might be captured) or non-health outcome factors, such as convenience; and
- the health outcomes are significantly affected by prognosis.

If the introduction of the proposed medical service is expected to induce a succession of changing health states that have a significant interactive effect on utility and the composite utility is not equal to the sum (in which a profile of health states would need to be valued), this then suggests that the QALYs approach is unlikely to be suitable, and an alternative and technically more complex approach might be more appropriate, such as a healthy-year equivalents approach.

An assessment report might need to present a scenario-based approach to valuing health states as utility weights in the absence of any trial-based utility weights. In this situation, the main objective of achieving a comparable approach across assessment reports is diminished. Furthermore, many of the issues in interpreting scenario-based utility weights in the absence of trial-based utility weights are similar in nature to the issues in interpreting any results of non-randomised studies in the absence of a direct comparison randomised trial. In particular, it is difficult to minimise the many sources of analyst bias that are intrinsic to this approach (including in the unblinded nature of the construction and presentation of the scenarios, the design of the methods to elicit values and the analysis and interpretation of the results, which are all conducted after the trial results are known).

A particular source of potential biases can be identified with post-trial scenario-based approaches to valuing health outcomes. This is because there is a justifiable preference for eliciting these values from individual respondents drawn from the general population (because they might better reflect the perspective of society overall as representing the balance of taxpayers and patients) rather than of patients alone (who are likely to recognise that they would be the beneficiaries of any new subsidised intervention). However, this inevitably leads to an information asymmetry for the respondent in relation to each specific post-trial scenario in a scenario-based utility study. Seeking to address this information asymmetry by loading more information into the scenarios raises the problem that respondents might manage this burden by unknown filter mechanisms used subconsciously when assimilating the information provided about the scenarios.

On the other hand, giving insufficient descriptions of the scenarios raises the problem that respondents might manage the gaps by unknown extrapolations, also used subconsciously, when assimilating the information provided about the scenarios. It is likely that both assimilation processes are operating simultaneously whenever a respondent is interpreting the presentation of scenarios. It would therefore be expected that their responses would be sensitive to the construction and presentation of the background and scenarios by the analyst. However, any examination of the sensitivity of the results to these sources of bias would be limited by the number of scenario variations that can be examined for any one respondent or in any one study. In contrast, these sources of bias can be more successfully minimised by the trial-based MAUI approach outlined in Sub-section A4.2 of this Appendix, which separates the scoring of each health state by the fully informed but appropriately blinded patient who is actually experiencing it from the previous generation of the valuation of that health state by members of the general population (thereby avoiding the need for a further analyst to act as an intermediary after the trial).

The post-trial scenario construction process has a number of implications. The scenario-based approach runs the risk of presenting ‘extremes’ of health states for valuation rather than reflecting the distribution. Given the limited number of health states presented for valuation, there is rarely a basis to examine this source of uncertainty in sensitivity analyses. Furthermore, a key implication of analyst bias is the potential for the scenario-based approach to focus on particular symptoms and attributes, which would not necessarily be the way that a person experiencing the health state would perceive it. This leads to a distortion along the lines that ‘nothing seems as important as when you are asked to think about it’.

Presenting the methods of generating scenarios and of presenting them to respondents

If preference weights in utility units have been derived with the use of hypothetical health state scenarios, provide details of the methods used in the utility study as part of the information provided in Part II, Sub-section B1. Provide data and references that support the validity and reliability of these methods.

Describe the approach taken to construct the scenarios. The scenarios should be developed rigorously, including by demonstrating that consideration has been given to the following:

- Describe the basis of the derivation of the health state scenarios for the survey. Discuss the relationship between these scenarios and the quantified estimates supporting the conclusions presented in Section B of the assessment report or modified in Section C of the assessment report. Given the inherently subjective nature of this process, report any attempt to minimise selection bias in the process and its impact. A more convincing case would be based on a randomised trial that measured health-related quality of life frequently with one or more valid and reliable generic instruments, and the construction of the scenarios is justified and compared with the detailed quality-of-life information from the trial results using these instruments.
- Explain the derivation of the descriptions in each scenario. Discuss the approaches taken to reflect the experience of patients experiencing these health states in the text of the scenarios. For example, describe the derivation of the health state scenarios and weighting and whether they were derived directly using one or more facilitated focus groups (such a group should include Australians — users of the proposed medical service and people with some experience of the medical condition, as well as medical experts). In particular, explain how the five to nine attributes (see guidance in relation to text below) were selected for inclusion in each scenario from the range of patient experiences. Discuss the need for, and implications of, choosing a proxy (e.g. a carer; a family member; or a health care professional) in place of patients for this step.
- Examine whether the description of each scenario was understandable to Australian respondents. For example, report whether initial scenarios developed were piloted using in-depth interviews on all aspects of the respondents' thoughts and comments before undertaking the full survey. If a pilot study was conducted, advise whether it identified any issues and how these were addressed before the scenarios were used in the utility study.
- Report any assessment of the scenarios developed in terms of validity, reliability, responsiveness to change, and clinical importance. Report any assessment of the duration of the period covered in each scenario compared with the duration assumed in the choice-based preference elicitation task (see below).
- Clearly distinguish between elements in the scenarios relating to health and elements not relating to health (such as convenience of use, increased availability of options and any other externality). If non-health elements are included, ensure that elicited preferences can be presented separately as health elements alone or as health elements combined with other elements. The base case should be based on health elements alone. Use sensitivity analyses to examine the impact of including any other elements.

The text used to describe each health state scenario is crucial as the means to convey the basis of the utility weight elicited. Demonstrate that consideration has been given to the following:

- Respondents to scenarios are likely to be subject to cognitive overload when the number of attributes or aspects of the health state increases beyond five to nine.

- Each scenario should adopt the patient’s perspective, such that respondents are to imagine that they are in the health state described. The scenarios might be presented in the first or third person.
- Each scenario should be a single static health state rather than a profile of two or more different health states.
- The ‘ex post’ perspective (in which the health state is as experienced with a full diagnosis without considering the risk of a future event) is preferred in the description of scenarios to ensure that all relevant and important aspects are included explicitly and that all irrelevant aspects are excluded (e.g. the process of diagnosis and a range of possible prognoses). Provide a justification to support the use of an ‘ex ante’ perspective in any health state scenario. A possible example is the use of a medical service that is intended to prevent a future harmful event.
- As the scenarios are to be presented to individuals with limited technical knowledge, use simple language and a logical sequence of presentation of material to allow all respondents to understand the background and the scenarios. Avoid technical terms and unnecessary words.
- Minimise the possibility of framing and labelling effects in which apparently small changes in wording of the scenario can produce substantial shifts in response. A possible way of doing this is to provide more background context, but because each scenario is essentially a subjective matter, it is difficult to anticipate where problems could arise in any particular context. Report the results of any pilot testing for obvious framing and labelling effects (e.g. the use of emotive disease labels such as ‘cancer’ or ‘neurological disorder’ in the health state description) in the design and implementation of the scenario. An exception to the above example might be where an ‘ex ante’ perspective is justified.
- To minimise sponsor bias, the supplier should not be named during the survey. To focus on the health state, it would be preferable not to identify the medical service or the nature of the service. A justification should be provided if the service assesses some non-health outcome aspect of the treatment.
- Consider including questions to confirm the respondents’ comprehension of the background information and scenarios provided, and report the results of such a validation exercise.
- Justify the number of scenarios to be presented for valuation. The burden on respondents represents an upper limit, which is influenced by the complexity of the information presented and the number of attributes, as well as the number of scenarios. If the number of scenarios to be valued is less than this upper limit, consider including one or more extra scenarios that capture any important variation in the description of one or more health states to be valued. These extra scenarios would enable the presentation of sensitivity analyses of the impact of the description of the scenarios valued for the base case. An important limitation of the scenario-based approach to valuation is that sensitivity analysis of this important source of uncertainty is rarely presented.

Provide a copy of the information provided to the respondents as an attachment to the assessment report. Include in these materials any background information, the text of all health state scenarios, any questions used to confirm comprehension and the questions used to elicit preference weights (‘utilities’). Also provide a copy of any computer program used to facilitate the presentation of information and the elicitation of utility weights.

Outline the methodology adopted in implementing the survey instrument. Demonstrate that consideration has been given to the following:

- face-to-face interviews are preferred to facilitate comprehension of the background information provided, the description of the scenarios and the questions asked. Provide a justification to support the use of telephone interviews or posted self-administered questionnaires;
- the respondent should be asked questions throughout the background narrative to keep them involved and to ensure understanding;
- interviewers should be carefully trained to read material at an appropriate pace, and to use conversational inflection, pauses and eye contact in the appropriate manner; and
- material should be provided in a logical sequence and illustrated where appropriate with pictures, graphs or diagrams. Include display items to improve understanding and to increase interest.

Comment on how the study addressed the controversy of whose utility weights are elicited (e.g. a patient, a proxy for the patient, such as a care-giver or a member of the general population) discussed in the background above. The possibly unattainable ideal is that these utility weights are elicited from a representative cross-sectional sample of the Australian general population that is fully informed of all health implications of each health state scenario presented.

If respondents are not from the general population, this approach might also raise potentially important issues to do with adjusting utility weights for groups of patients in certain disease groups (so-called double jeopardy, see Sub-section A4.3 of this Appendix for further explanation). Therefore, for health states reflecting a chronic medical condition, also comment on whether the approach taken reflects adaptation of patients to the experience of the health state, and the implications this has for relating the valuation to the duration of the health state.

Elicitation, statistical analysis, reporting of results and interpretation of scenario-based utility valuation of health outcomes

Anchor the utility weights elicited on a 0–1 ratio scale of death to full (perfect) health. Elicit these weights using a choice-based preference elicitation task, which makes explicit that a choice or trade-off has to be made, and therefore allows for the strength of preference to be revealed. Justify the method chosen and provide details of the method used. The method chosen might be one of the following:

- SG: this method has the more direct theoretical foundation.
- TTO: this is a direct measurement tool designed specifically for use in health care evaluation. It is more appropriate for use by respondents who have difficulty in understanding probabilities. It is particularly useful in studies that compare alternatives in which TTO is the major clinical factor. The utility weight is based on how much quantity of life people are prepared to give up for additional quality of life.

- Each of these scaling techniques is confounded: TTO by time preference and SG by risk attitude. As both SG and TTO relative values are consistent in the direction of expected bias compared to each other, and comparison of the two techniques indicates that they provide similar results; therefore, either can be used as a scaling technique in an assessment report.
- The use of a MAUI to generate utility weights from a scenario is discouraged. This would not be a preference elicitation task, but rather a ‘mapping’ from one scenario to another MAUI-based scenario. If the scenario captures only a few domains covered by the MAUI, the respondent is forced to guess from the information provided what response should be given for the other domains covered by the MAUI. On the other hand, if the scenario is constructed to capture all domains, the analyst’s control of the scenario descriptions is so influential that the descriptive words chosen can tend to lead the respondent towards particular responses in each domain. In an extreme case, the analyst could effectively nominate the utility weight yielded by this approach based on their own expert opinion, and then align the text of the scenario descriptions to the text of the MAUI questions.
- Other methods for eliciting preferences, such as discrete choice experiments or other conjoint analysis methods, are still in development and thus any guidance here is preliminary. There are five main stages that characterise these types of study:
 - *Determine the attributes*: if based on one or more submitted randomised trials, the attributes should reflect the different components of the trial arms. If they are not defined on this basis, then literature reviews, patient group discussions and individual patient interviews will need to be used to solicit the attributes. These attributes should be important to the patients. If cost is used as an attribute, the technique can generate willingness-to-pay (WTP) under certain circumstances (see Appendix 5). To ensure that the analysis is being used to value health states rather than to value the treatments, it is important to exclude any other description or process aspect of the treatment.
 - *Define the characteristic levels*: justify the use of cardinal, ordinal or categorical scales. The levels should be realistic, be capable of being traded off and capture all relevant outcomes.
 - *Choose the scenarios to be presented in the stated preference experiment*: justify the presentation of the scenarios to ensure that they are realistic (e.g. ensure that the defined period of time for each scenario is consistent for both the proposed medical service and the main comparator) and that they make sense to the respondent (see guidance on constructing the scenarios in this Section of this Appendix, above). The number of scenarios will increase with the number of attributes and attribute levels, and it is generally not feasible to present all combinations of scenarios in a questionnaire. Use an appropriate experimental design, typically a fractional factorial design based on orthogonality, to choose the subset of scenarios to be presented in the experiment. Describe and justify the basis for generating the experimental design, including details of any software used. Provide the full experimental design in an attachment to the assessment report, including a list of all scenarios developed.
 - *Establish preferences using discrete choices*: present each respondent with a series of pairs or groups of options (choice sets) among the scenarios and request that a selection be made defining which is the most

preferred. Ranking and rating exercises have been used in conjoint analysis; however, the use of discrete choice experiments is preferred, because they are more consistent with the choice-based nature of SG and TTO, and have a more established basis in economic theory and statistical analysis.

- *Analyse data*: analyse the responses from the scenarios using regression techniques. Typically, a multinomial logit analysis is used because the dependent variable is a discrete random variable. Justify the modelling approach, including consideration of treatment of repeated observations and heterogeneity (e.g. use of mixed logit). Report on the extent to which the model explains the variation in preference selection. Explore the impact of possible confounding factors.

Claimed advantages of conjoint analysis include the ability to describe health state changes in terms of comparisons across the attributes, the duration of these changes and the probability of these changes occurring. Although the techniques of conjoint analysis are developing, they are still not yet sufficiently acceptable to have direct influence on MSAC decision making on their own. They are claimed to also explicitly consider non-health elements (in which case, results should be presented with and without including those elements). However, it is not clear that there is an acceptable framework outside the QALY framework in which to consider these claimed advantages in a comparable way across assessment reports.

Ensure that the sample size is large enough to measure population variance. The power of the study should be tested and between-group correlations should be demonstrated.

Present the results of the utility study as part of the information provided in response to Sub-section C2. Report the results as the point estimate of the mean utility of each health state scenario with its 95% confidence interval. In discussing these results, provide an overall assessment of the approach adopted to elicit preference weights from the hypothetical scenarios. Particularly, consider whether the methods by which the health state scenarios were:

- constructed allow all the critical changes in quality of life associated with the intervention to be captured and presented in such a way that they are accurately perceived by the respondents; and
- derived and constructed are likely to lead to bias in the valuation of health-related quality of life associated with the medical service; for example, by focusing on some aspects of health-related quality of life (such as example physical functioning) while excluding or minimising the impact of others (such as mental or social health).

From these results presented in Sub-section C2, identify and justify the estimates to be used as variables in the economic evaluation presented in Sub-section D5 for the base case and Sub-section D6 for the sensitivity analyses.

A4.5 Other methods for obtaining utilities

The following methods have all been presented in assessment reports for MSAC. Each raises a series of concerns, as detailed below.

Mapping of generic and disease-specific scales

In contrast with MAUIs, although other generic and disease-specific scales might be based on sophisticated psychometric techniques for instrument construction, none of those scales is capable of representing individual preferences on a scale of 0 = death and 1 = full (perfect) health, and so none can be used to calculate QALYs without some transformation. Despite this, a number of attempts have been made to ‘map’ from scores reported in randomised trials using generic or disease-specific quality-of-life measures into utility weights, which are then used to construct QALYs. Approaches vary from a simple intuitive mapping to the use of statistical techniques. For example, responses on a visual analogue scale of 0 to 100 to the question asking respondents to rate their health today have been divided by 100 and (wrongly) claimed to therefore measure utility weights on a 0 to 1 scale. Another example is the use of regression to ‘map’ an association between two sets of responses from a survey of respondents, each completing both the quality-of-life instrument and a MAUI, or other acceptable technique of eliciting preference weights. This regression ‘map’ is then used to transform into ‘utilities’ the responses to the quality-of-life instrument reported by respondents in another trial.

These are not well-established procedures. Where statistical techniques have been used, tests of reliability might include the predictive value of the technique across a range of quality-of-life values and changes in quality of life within, and differences between, respondents with the relevant medical condition. Where this approach is adopted, extensive sensitive analysis around the estimates generated should be undertaken to examine the sensitivity of results of the economic evaluation to this variable. Where such ‘mapping’ is presented, special attention needs to be given to establishing that the results generated are plausible and unbiased, particularly where the preference weight estimates generated have a substantial impact on the results of the economic evaluation.

It is difficult to illustrate the assessment of plausibility and bias in these circumstances. An approach that does not ‘map’ to an adequate utility instrument (i.e. that satisfies characteristics (b) and (c) of the QALYs shown in Section A4.2 in this Appendix) would not meet an essential prerequisite in estimating a preference weight index. An approach that is not based on a study that concomitantly measured the quality-of-life measure and such an index would also not meet an essential prerequisite to generate an association. Other issues to assess include the difficulties of ‘mapping’ ordinal (ranking) scales to the cardinal utility scale, the presence of floor and ceiling effects in most quality-of-life measures, and whether an acceptable range of important dimensions are adequately captured. A more structural approach might be taken to map specific dimensions of a generic quality-of-life instrument to corresponding dimensions of a MAUI (possibly best exemplified by the mapping of the SF-36 to the SF-6D), but this involves a much greater amount of developmental research work.

Population-matching studies

Another alternative occasionally used involves recruiting a separate sample of patients with characteristics similar to those in the randomised trials and for whom listing is requested. These matched patients then complete a MAUI reflecting their current health state (as a surrogate for a trial participant directly completing the MAUI), which is then used to estimate utility weights for the economic evaluation.

This population-matching approach is also subject to multiple sources of bias and thus uncertainty, particularly related to how similar the sampled patients are to those in the economic evaluation and the inability to blind the sampled patients from the objectives of the study. This can be context specific; for example, if there are important side effects, it might be particularly important to ensure that the sampled patients are exposed to the medical service and its side effects at the time the MAUI is completed.

This approach might be strengthened by getting the sampled patients to complete another quality-of-life instrument that was completed in the trials, and using the results of this concurrent instrument to more closely match a subset of sampled patients with trial participants and with the population for whom listing is requested. It can also be used to develop sample-based statistics of variance around the utility weights, which can be used in the sensitivity analysis of the economic evaluation.

Preference weights ('utilities') sourced from the literature

'Off-the-shelf' utility estimates might be available from the literature, and have been most often used when seeking to examine the impact of quality-adjusting a survival claim estimated in terms of life-years gained. As for any presentation of secondary (or even tertiary) data or analysis, the validity of the utility estimate depends on the methods used to elicit the estimate. Accordingly, present and assess the results against the preferred characteristics of a primary utility study, including:

- how the studies were identified (e.g. systematic search preferred to selective reporting);
- how representative the health state in each identified study is of the health state in the presented economic evaluation (including in dimensions of the type and severity of symptoms, and the duration of the health state);
- how the health state was captured (e.g. MAUI versus scenario based);
- how the preference was elicited (e.g. SG or TTO);
- what sample was chosen to respond to the MAUI questionnaire or scenario (e.g. members of the general public, patients, care givers, health care professionals);
- what assessment was made of the nature and direction of bias that might arise given the sample and methods; and
- how the sensitivity analyses examined variation in the identified utility options.

A particular difficulty in interpretation has occurred when a cost-utility analysis relies on combining utility weights across different sources for different health states within an economic evaluation, particularly across different sources that used different methods.

Appendix 5 Monetary valuation of health outcomes

A5.1 Preference for cost-utility analyses over cost-benefit analyses

Cost-benefit analyses are not preferred by MSAC because they are not likely to be helpful to most MSAC deliberations. The reasons for this are as follows:

- Cost-benefit analyses are typically applied in the context of a fixed-decision rule, which does not incorporate the breadth of equity and ethical considerations that are relevant to MSAC decision making (see also Appendix 1).
- The use of WTP to elicit monetary valuation for a cost-benefit analysis, which will be influenced by an individual's income and assets, is inconsistent with the principles of MSAC as a subsidy program to ensure equity of access.
- There remain considerable problems with interpreting WTP responses in the context of the Australian health care system where individuals do not typically face market prices. It could be argued further that the MBS, which uses fixed levels of co-payment and safety nets to achieve its objective in minimising low income as a barrier to accessing medical services that are MBS-listed in Australia, removes price signals even more than other elements in the Australian health care system.
- The methods for deriving monetary valuations of health gains presented to date have not satisfactorily minimised the hypothetical nature of the responses elicited or the incentives for the respondents to provide values that reflect a desire to have the MBS subsidy proceed in the full knowledge that the respondent will not directly incur this cost. Although it is theoretically possible to improve the realism of the scenarios and of the questions asked to elicit plausible monetary values (see Sub-section A5.2 of this Appendix), there remains a residual uncertainty in aligning the provision of resources valued in monetary units with welfare outcomes, which are apparently valued in the same monetary units.
- Cost-benefit analyses typically assign preference weights including to other welfare changes beyond the primary focus of MSAC on health outcomes (these include production changes and process changes), which have tended to reflect the construction of the scenario or attribute used to elicit the monetary valuation rather than to reflect the weights assigned by MSAC when considering a fuller range of other relevant factors, particularly equity.
- For the above reasons, there is unlikely to be a consistent exchange rate between monetary valuation and the utility weight that is the preferred basis for assessing strength of preference (see Sub-section A4.1 of Appendix 4). Therefore, considering these two approaches to valuing outcomes in parallel would predictably result in inconsistent decisions across assessment reports. This is undesirable.

- Although it is possible to use utility-based instruments in randomised trials to estimate the strength of preference for different health outcomes, this is not yet practical for monetary-based instruments. Therefore, the advantages outlined in Sub-section A4.3 of Appendix 4 for trial-based utility weights cannot be generated for monetary valuation. There are therefore disadvantages in common between scenario-based utility valuation (see Sub-section A4.4 of Appendix 4) and scenario-based monetary valuation (see Sub-section A5.2 of this Appendix).

Given the above reasoning, monetary valuation of health outcomes is allowed but is considered to be supplementary to utility valuation. Therefore, if both a cost-utility analysis and a cost-benefit analysis are presented in an assessment report, discuss the differences in the results and any differences in conclusions. In the absence of a cost-utility analysis, discuss why only a cost-benefit analysis is thought to be informative and why a cost-utility analysis is not possible. For example, consideration of such analyses might be justified in some situations to provide informative insights to the perception of the respondents to the clinical performance of a proposed medical service; however, such analyses should be interpreted cautiously in the absence of a worthwhile gain in health outcomes. Further guidance is provided in Sub-section A5.2 of this Appendix.

A5.2 Scenario-based monetary valuation of health outcomes

Background

Monetary valuation of health outcomes is typically scenario based. The issues raised in Sub-section A4.4 of Appendix 4 regarding the use of scenarios as a basis for eliciting the strength of preference in a utility metric largely overlap with their use as a basis for eliciting the strength of preference in a monetary metric. It is conceivable that monetary valuation could be elicited in the context of a randomised double-blind trial, but the practicalities of addressing the issues raised below suggest that this will not occur in the near future.

This Appendix seeks to identify those areas where monetary valuation might be informative in situations where utility valuation is problematic. Situations identified to date have tended to arise due to concerns over the lack of sensitivity of utility valuation to perceived increments in health outcomes. These have included short-term changes in health outcomes, differences in health outcomes that are too small to be detected with utility-based instruments, and differences in adverse outcomes for two medical services that are otherwise similar in terms of comparative effectiveness. An alternative metric might be justified in these circumstances, because underlying the QALY approach is the fact that survival duration is the metric, and there might be health gains that are valued, but are not sufficient for individuals to trade-off survival. However, this reduces comparability across assessment reports, because it introduces a new valuation system that is not necessarily interpreted the same way in the valuation step by the respondent as utility valuation. It also brings in other aspects, whether implicit or not, beyond valuing health outcomes.

An assessment report seeking to supplement a utility valuation of health outcomes with a monetary valuation of health outcomes should provide a justification for doing so. Alongside this justification for providing these supplementary estimates, present both sets of methods and results, and comment on the interpretation of the results compared with each other. As with the interpretation of the results of any measure of health outcomes, any claim for an improved sensitivity in quantifying the utility weight of smaller advantages needs to be assessed against the possibility of a true negative (i.e. that the

proposed medical service has no overall perceptible incremental effect on strength of preference; see also Sub-sections A4.3 and A4.4 of Appendix 4). Document the evidence that supports any claim that any difference in results between utility-based valuation and monetary-based valuation is attributable to the special characteristic of the health state, and not some idiosyncrasy in the utility measurement procedures that have been adopted. This would help justify any apparent diminution in comparability across assessment reports that provide utility weights.

An assessment report that provides monetary valuations of health outcomes without corresponding utility valuations would be more difficult to assess in terms of comparability across assessment reports.

Consistent with the request in Part II Section D and Sub-section A4.4 of Appendix 4, an assessment report that seeks to provide a monetary valuation of any attribute other than health outcomes (e.g. a production change; see Appendix 6) should do so separately from the valuation of health outcomes. This can be done by providing a supplementary economic evaluation that adds the additional information to the base-case economic evaluation. A request in an assessment report for MSAC to consider a non-health outcome or process attribute (such as convenience of use, increased availability of options and any other externality) would need to be judged on its merits, which would be informed by the direction and extent of the impact of its inclusion on the base-case economic evaluation. This distinction is therefore important both to promote consistency of decision making based primarily on health outcomes and to allow flexibility to consider other factors that MSAC might accept as relevant.

Presenting the methods of generating scenarios and presenting them to respondents

If preference weights in monetary units have been derived with the use of hypothetical health state scenarios, provide details of the methods used in the study as part of the information provided in Sub-section B. Provide data and references that support the validity and reliability of these methods. Refer to the text under the corresponding subheading of Sub-section A4.4 of Appendix 4 to identify the information to be provided, including a clear description of the attributes that are compared between the proposed medical service and its main comparator. Additional information specific to monetary valuations includes the following:

- describe the attributes in each scenario in a way that matches the policy question and the underlying theoretical construct to be addressed in the contingent market;
- whenever a probability of any type is included for an attribute in a scenario, examine more than one level of probability when eliciting monetary values in order to assess the degree of understanding (e.g. that a greater probability of benefit yields a greater monetary value of WTP); and
- where scenarios are developed as changes in health states rather than as the health states themselves, describe the likelihood, extent and duration of each change.

Elicitation, statistical analysis, reporting of results and interpretation of scenario-based monetary valuation of health outcomes

The most commonly used method is contingent valuation (CV) to elicit WTP. If a CV study is included in an assessment report, provide a justification for its inclusion, including why it would be informative for MSAC decision making.

The assessment report should outline the methodology adopted in designing and implementing the CV survey instrument. Demonstrate that consideration has been given to the following:

- The contingent (hypothetical) market should be a simple out-of-pocket payment to elicit the individual's strength of preference by considering the question of spending their private income to estimate the value of the change in health states being presented. Ensure that respondents understand the nature of the payment vehicle and that their responses are interpreted appropriately. The average WTP across respondents from this valuation might not necessarily be the WTP that society overall has for subsidising medical services to improve health outcomes for the population as a whole, but it is not clear that changing the hypothetical market to reflect a societal question of funding a public subsidy program would be meaningful to respondents. This market should also be described in simple language, eliminating unnecessary words and avoiding technical jargon.
- The initial WTP elicitation instrument describing the contingent market should be piloted alongside the piloting of the background information and the scenarios. Report any issues arising and how they were addressed before the full study began.
- Discuss the choice between a discrete choice format or an open-ended questionnaire format (with prompts or a payment card) to elicit responses. The closed-bid discrete choice format with randomly selected bids presented to each respondent — and only one bid per respondent — is more theoretically valid and less subject to bias than the other methods. Other issues to consider include the sample size required for the statistical analysis to infer the mean WTP from discrete choices, and the increased likelihood of nonresponse or protest response from open-ended questions. Justify the range of values used in the discrete choices or the prompts or payment cards. When conducting the survey, randomly allocate the selection of the order of discrete choices across respondents or the selection from the range of values in prompts and cards.
- To ensure some consistency within the time frames across different WTP studies, frame the questions in one of two ways:
 - as a one-off payment but constrained to within any one year, by invoking each respondent's annual (rather than lifetime) income; and
 - as a regular annual payment, with the value derived for 'this year' only, not for a 'hypothetical' year.
- Remind respondents of their budget constraints for their WTP throughout the survey.
- When conducting the survey, adopt a random ordering of questions across respondents.
- WTP studies should be conducted in a comparative sense and respondents should be made aware of any close substitutes. This would help to make clear the extent of incremental improvement in health across the alternatives.

- WTP is expected to be correlated to ability to pay. Indicate whether ability to pay has been assessed according to personal or household income (and, if the latter, whether this is adjusted for household size) and whether it has been assessed according to current income or also reflects assets that could be realised to make payments. Socio-demographic characteristics of respondents should be collected and included in the analysis.

From the above information, indicate the steps that have been taken to minimise the following sources of bias in the WTP survey:

- hypothetical bias: the respondent responds to a perception that the survey is hypothetical with hypothetical and therefore meaningless answers;
- strategic bias: the respondent varies the WTP from the ‘true’ WTP to increase the chances of getting a preferred decision by influencing the decision maker;
- interviewer bias: face-to-face or telephone interviews run the risk that valuation will be influenced (purposefully or accidentally) by the interviewer;
- starting-point bias: the initial prompt or bid in the bidding approach will anchor the respondent towards the starting bid, narrowing the distribution around the mean (portraying greater consensus than truly exists) and causing a loss in efficiency;
- ‘yea-saying’ bias: the respondent will agree with amounts as offered by interviewer;
- range bias: the elicitation procedure presents a range of potential WTP amounts that influences the WTP amount given by respondents; and
- sponsor bias: knowledge of the identity of the sponsor affects responses; minimised by not naming the sponsor of the survey or the manufacturer of the medical product.

The validity of the WTP depends on minimising sources of bias to reveal the true strength of preference in monetary terms.

Some preliminary guidance in relation to other stated preference methods, such as discrete choice experiments and conjoint analysis, is presented under the corresponding subheading in Sub-section A4.4 of Appendix 4. The methodological guidance on those methods should be considered in addition to the general guidance given above in this Section for valuing discrete health states. In addition, discrete choice experiments might also be used to calculate monetary measures of the composite of incremental health outcomes from the proposed medical service as a comparison of the alternative profiles of health outcomes over defined periods of time resulting from the proposed medical service and the main comparator. If so, justify the presentation of these profiles of health states to ensure that they realistically and accurately reflect the choice context (e.g. allowing for a ‘status quo’ or an ‘opt out’ option where appropriate for the presentation of the alternative profiles in each choice set) and that they make sense to the respondent (see general guidance on constructing the scenarios).

The statistical analysis, interpretation and reporting of data

Present the results of the scenario-based monetary valuation study as part of the information provided in response to Sub-section B1. Report mean WTP values on a net present value basis for each health state and then the overall aggregate with their 95% confidence intervals, interquartile range and full range.

Assess the results of the WTP survey as follows:

- present WTP values without adjustment for income. Also report WTP disaggregated across income group. Where the mean ability to pay in the survey differs from the national average, comment on the interpretation of the results;
- present the results both in an unadjusted fashion and with outliers removed. Discuss any difference in these results;
- report the response rate. Comment on the implications of the response rate and other potential sources of selection bias for the interpretability of the results of the survey;
- report the proportions of zero and very high bids. If either or both of these are greater than 10%, discuss the possible reasons for these proportions and their implications. Ask respondents to explain their reasons for responding with a zero bid;
- conduct regression analyses to assess the factors that might explain the WTP values given. Variables to examine include an ‘interviewer’ variable, a ‘question order’ variable, a ‘prompt’ variable (of the range of starting values in the prompt) and an ‘income’ variable; and
- assess whether the results make economic sense (i.e. that WTP increases with the size of both health gains increases and ability to pay increases).

WTP values are context specific, so values should only be used and applied to the specific circumstances for which they were obtained. WTP values are interpreted as an upper limit to true valuation. From these results presented in Sub-section B1, identify and justify the estimates to be used as variables in the economic evaluation presented in Sub-section D5 for the base case and Sub-section D6 for the sensitivity analyses.

Appendix 6 Including non-health care resources and non-health outcomes in a supplementary analysis

This Appendix provides additional guidance on the preparation of supplementary analyses of an economic evaluation to incorporate changes in non-health care resources and/or non-health outcomes that would be attributable to the listing of the proposed medical service (see Part II Sub-section D1).

A6.1 Identifying, measuring and valuing non-health care resources

Occasionally, because of the medical condition under treatment or the age of the patients, consideration of direct non-health care costs such as social services (home help, day care, meals on wheels, private travel to access health care, etc.) might be relevant.

If incorporation of non-health care resources is relevant for a supplementary analysis, adapt the general principles as detailed in Part II Sub-section D4 for health care resources to generate and present these variables. In brief, the resources should be identified and defined. An appropriate unit of measurement should be identified and the extent of change in the provision of the resources should be estimated. Present and justify an appropriate unit cost to estimate the value of the resources.

A6.2 Identifying, measuring and valuing non-health outcomes

Occasionally, listing a proposed medical service 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.

If incorporation of changes in non-health outcomes (including economic outcomes) is relevant for a supplementary analysis, adapt the general principles outlined in Part II Sub-section D4 for health outcomes, including by reference to Sub-section A5.2 of Appendix 5, as appropriate. In brief, the outcome should be identified and defined. An appropriate unit of measurement should be identified and the extent of change in the outcome should be estimated. Present and justify an appropriate valuation of the outcome.

Production changes

A production change is the value estimated in monetary units of the potential working time gained or lost measured in time units (days, weeks, years, etc.), which is realised as productive activity. It might also include realising the productive change of the potential impaired working time gained or lost by a sick patient who continues to work (measured in similar time units together with a measure of any associated change in the extent of impairment). Production changes have been called indirect economic outcomes in recognition of the fact that subsequent decisions had to be made to realise the time gained as productive activity to the advantage of the rest of society rather than as any other activity.

Provide a strong justification if production changes are combined with surrogate outcome indicators in an economic evaluation because this combination is generally inappropriate.

If production changes are to be included in a cost-utility analysis, adopt a method that 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 to the individual receiving the proposed medical service. These health-related changes are therefore already appropriately included in the denominator of the cost-utility ratio.

Unlike direct health benefits, the economic benefit to society through patients' return to, or maintenance of, productive capacity is both difficult and controversial to estimate accurately. This is because the available methods and their application remain unresolved. Therefore, although changes in production as an outcome of treatment might be included in supplementary analyses in assessment reports for MSAC, they should not be included in the base-case analysis.

There are several difficulties in estimating the net present value of production changes. These estimates are underpinned by three assumptions:

- for short-term absence, production will be made up on the return to work;
- employers usually have excess capacity in the labour force to cover absenteeism; and
- for long-term absence, production will be made up by a replacement worker otherwise unemployed.

Where estimation of production changes can be justified in the assessment report, address each of the three underlying assumptions listed above when estimating production changes from the potential working time gained or lost (reported in time units). For example, the claim that there has been a recovery of production lost due to returning to health from an episode of illness depends on demonstrating that:

- the worker returns to work;
- 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 (note: if the worker is highly productive, the incentives to replace that worker are stronger); and
- no temporary replacement from outside has been employed (namely, that there is full employment).

As in this example, the **marginal** increase in society's production due to the return of healthy workers to the workplace is overestimated if the human capital method is used; that is, the workers' time regained is simply multiplied by the labour market value of the average worker (usually estimated by the average wage). It is not always likely to be zero either, but some proportion in between. Provide and justify the best estimate of the true proportion based on firm evidence.

Addressing the four questions in the example above would therefore help to convert the potential working time gained or lost reported in time units into production gains or losses reported in monetary units. The friction method has been advocated as a method that provides a basis to help make this type of conversion. Although there is no evidence that it has yet been applied in Australia, it is theoretically preferable to the human capital

method for this reason. However, in the example provided above, it only offers a basis for addressing the last two of the four questions, and only does so by proposing an indirect estimate at the national level rather than a direct estimate at the patient level. The friction method therefore still generates an upper estimate compared with an approach that could address all four of the questions above. Other evidence needs to be provided to address the first two questions, because not all healthy workers would choose to deploy the time gain to return to contributing to societal production. In the example above, recognising that this choice exists is important because deploying the time gain for some other purpose, such as a leisure activity, is an intrinsic part of valuing the improved health as a gain in utility weights rather than valuing it as a production gain to society in monetary terms.

Any evidence to support an estimate of the proportion of people who choose to return to contributing to societal production would also need to account for the influence of incentives provided through various types of sickness benefit payments provided by social security systems and employers, which vary across countries. This might hinder the translation of overseas evidence to Australia.

Answering all four questions satisfactorily in the example above would therefore help minimise double-counting across the denominator and the numerator of an incremental cost-utility ratio, because it would more accurately estimate the extent of production gains to society beyond the gains valued by the population benefiting with improved health. Valued in monetary terms, these production gains would represent a more suitable estimate for inclusion in the numerator of this ratio.

The above example is intended to illustrate the application of the three more general reasons. A similar approach would be needed in other contexts, such as a medical service that prevents future episodes of illness, or a medical service that might improve production capacity in individuals who, without the proposed medical service, would otherwise stay at work, although unwell, and therefore perform at less than full production capacity.

Present the results of the economic evaluation excluding the production changes in the base case. Assess the impact of including these changes in a supplementary analysis. This separation allows MSAC to consider the impact of their inclusion on the direction and extent of change on the base case.

At the same time, MSAC can weigh up, as another relevant factor, the inevitable equity implications of varying the base case to include an element that explicitly favours those who make a greater contribution to production. Inclusion of production gains favours those medical services that improve the health of people who are able and choose to return to contributing to societal production.

The present value of production changes should be calculated. This means that where production gains are anticipated over a number of time periods (beyond one year) these should also be discounted. Discounting future costs and benefits is a standard feature of economic evaluation. Costs or benefits are discounted at an annual rate of 5%.

A6.3 Resources and outcomes to be excluded

Costs should be limited to those associated with the medical condition under treatment. In other words, do not include as consequences in the economic evaluation other unrelated medical conditions that, in the fullness of time, are likely to afflict patients who live longer as a result of effective treatment that they receive now.

Appendix 7 – Codependent applications across MSAC and PBAC

Introduction

This Appendix provides guidance on the preparation of an application to MSAC that involves a codependent medicine.

What are codependent technologies?

Health technologies are codependent where the patient health outcomes related to the use of one health technology (eg a medicine) are improved by the use of another health technology (eg a pathology test or an imaging technology). The use of the technologies needs to be combined (either sequentially or simultaneously) to achieve or enhance the intended clinical effect of either technology. Therefore, the net clinical benefits of the joint use of the technologies, as distinct from the net clinical benefit of each technology in isolation, needs to be determined for a health technology assessment. The cost-effectiveness and financial implications of the joint use of the technologies are also considered as part of the reimbursement decision.

The most common example of a codependent technology is a medicine-test combination where a new medicine seeking listing on the PBS has a related pathology test that may help to determine the population group eligible for that medicine. MSAC classifies such tests as ‘investigative medical services’.

An investigative medical service can have several purposes,⁶ of which the following are most likely to be relevant to codependent technologies:

- establishing a predisposition or estimating a prognosis
- identifying a patient as suitable for a therapeutic medical service by predicting a variation in the effect of the therapeutic medical service
- measuring an early treatment effect on a surrogate outcome as the basis for predicting the extent of a later treatment effect on more patient-relevant outcomes
- monitoring a patient over time after an initial investigation to guide subsequent treatment decisions if the service needs to be repeated.

To achieve an improvement in health outcomes, the investigative information from the test must result in a change in the management of a subsequent therapeutic service. In this sense, the test can only indirectly improve health outcomes and any improvement also needs to be balanced against any harm that the service might cause. This purpose defines the need for a codependent technology to be assessed.

⁶ Medical Services Advisory Committee. Technical guidelines for preparing assessment reports for the Medical Services Advisory Committee – service type: investigative (version 2.0). Canberra: MSAC, March 2016,

When is a codependent submission required?

A material codependency requiring a codependent submission arises when the Minister for Health requires advice from two different expert advisory committees because listing of the codependent technologies would involve two separate reimbursement schemes. For example, codependent technologies that require new listings or amendments to both the PBS and the Medicare Benefits Schedule (MBS) would need advice from both the PBAC and MSAC.

There are two different processes by which advice relating to the two reimbursement schemes can be formulated for the minister:

- **integrated codependent submission** – a combined submission for the two technologies is prepared and considered *jointly* by MSAC and the PBAC
- **streamlined codependent submissions** – individual submissions for each of the technologies (one for the test and one for the medicine) are lodged at the same time and are considered by MSAC and the PBAC, respectively, in parallel.

Further details of each process are provided below. Flowchart A7.1 shows the scenarios in which technologies with a material codependency have been considered for reimbursement by MSAC and the PBAC using an integrated or streamlined approach.

Integrated codependent submissions

Integrated submissions involve a submission of a medicine to the PBAC which also involves a codependent test or other investigative service that either:

- is not listed in the MBS; or
- requires a substantial amendment to the MBS to list it as intended, and thus entails joint consideration by both the PBAC and MSAC.

The format outlined in Subsection A7.2 is sufficient to meet the expectations of both the PBAC and MSAC with regard to a submission for PBS listing of the medicine and also a submission-based assessment for the MBS listing of the codependent pathology test or other investigative service.

In particular, lodge an integrated codependent submission when:

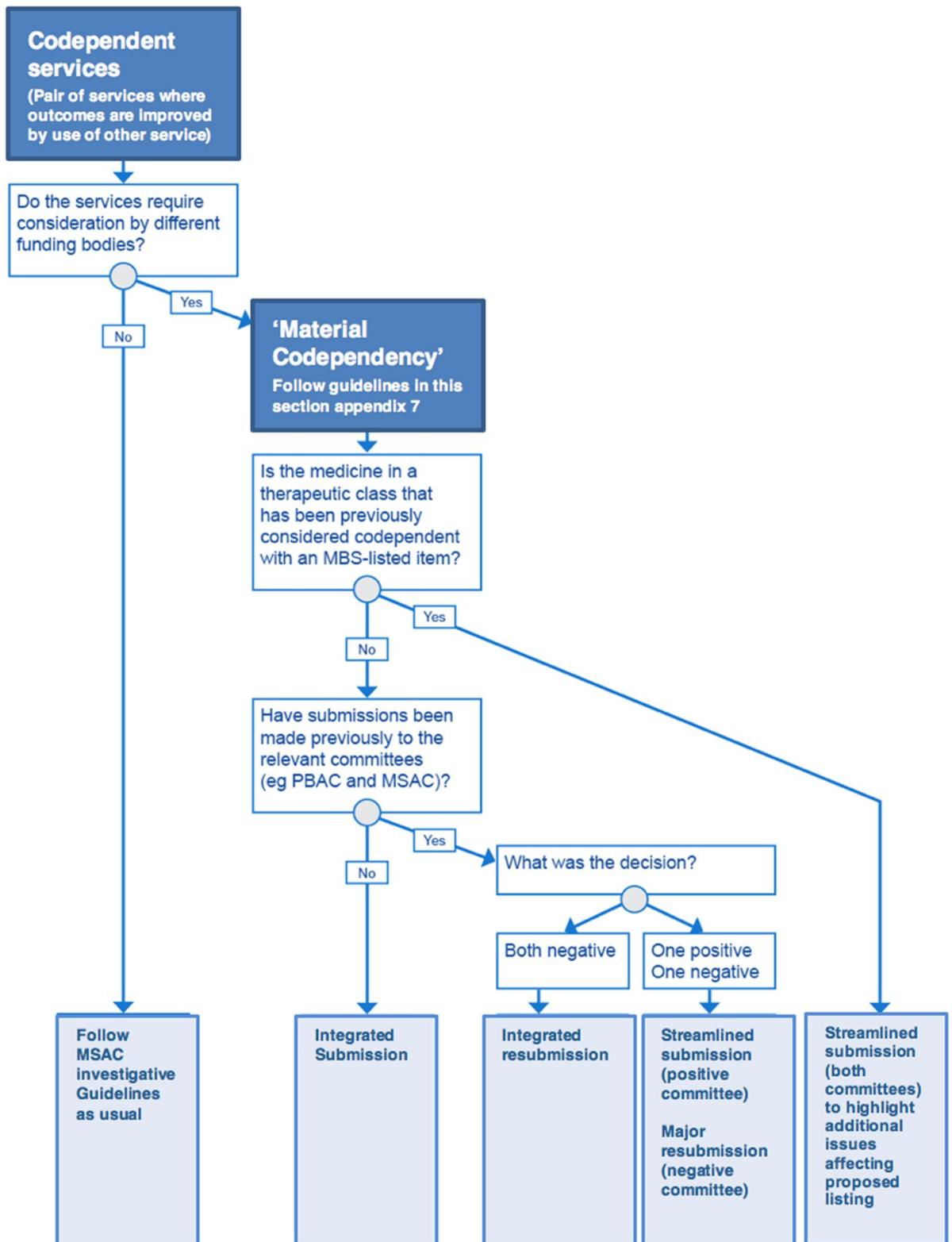
- the test and the medicine require a listing on the MBS and the PBS, respectively, and neither technology has been considered previously by either committee (MSAC or the PBAC)
- an integrated codependent resubmission is needed (ie both committees have indicated they are not satisfied with the information in the previous submission)
- the medicine is of a different therapeutic class to one that has been previously considered to be codependent with the MBS-listed companion test.

Streamlined codependent submissions

Codependent technologies can be efficiently *reconsidered* when, after previous consideration, only one committee has foreshadowed support for a technology in the pairing. For example, if MSAC has foreshadowed support for a codependent test or other investigative service, the lodgment of a resubmission to the PBAC may occur separately but in parallel to the lodgment of a streamlined resubmission to MSAC to ensure that MSAC's advice is expeditiously aligned with the circumstances of any PBAC recommendation for the codependent medicine. Similarly, if an MBS item descriptor for a

test or other investigative service needs minor amendment to accommodate access to a codependent medicine in the same therapeutic class as one that has been previously PBS listed, then a streamlined codependent submission to amend this item descriptor may be lodged with MSAC alongside the submission to the PBAC for the codependent medicine.

Flowchart A7.1 Classification of integrated and streamlined codependent submissions



MBS = Medicare Benefits Schedule; MSAC = Medical Services Advisory Committee; PBAC = Pharmaceutical Benefits Advisory Committee

Note: In the situation where the medicine is listed but the test is not, a material codependency does not exist because the decision to list the test falls to MSAC alone.

A7.1 Overview of information requested in codependent submissions

As already indicated, the amount of information to include in a codependent submission is contingent on any previous MSAC and PBAC consideration of public funding of the test and/or the medicine for the specific clinical condition under review.

Types of evidence

The approach to presenting evidence in an integrated codependent submission will differ according to whether direct evidence or linked evidence is available (see Figure A7.2 and Subsection A7.2 in Section 2 – Clinical evaluation).

Direct evidence

‘Direct evidence’ describes studies that compare groups of people receiving either the currently used diagnostic test/test strategy or the proposed diagnostic test/test strategy and measures the differential impact of the diagnostic method on patient health outcomes. If patients are randomised to receive the test, then biomarker status would be known and, on that basis, subsequent targeted therapy or usual care could be decided. If patients are randomised to not having the test, then a treatment would be received that is not targeted by the biomarker result.

Linked evidence

The ‘linked-evidence approach’ was proposed by MSAC whereby evidence of test accuracy comparing the proposed and current test/test strategy could be linked (if considered to be appropriately transferable) to separately sourced evidence of treatment effectiveness to approximate the likely clinical effectiveness of the proposed test/test strategy.

For example, this might involve linking evidence of the test’s performance (eg diagnostic accuracy) with evidence demonstrating that the test result changes the medicines or treatment prescribed, and with evidence that the alternative medicines have different effectiveness and safety profiles.

Integrated codependent submissions

If lodging an initial integrated codependent submission, consider addressing all the items outlined in Subsection A7.2. If lodging an integrated codependent *resubmission*, pay particular attention to the issues raised by both committees in deciding not to support the proposed codependent technologies.

The key to the evaluation of codependent technologies is to establish the basis of the codependency claim. Make the relationship between the test for the biomarker (investigative medical service) and the medicine explicit, particularly whether it is based on treatment effect modification and/or a prognostic effect. An integrated codependent submission must demonstrate that the medicine interacts with the biomarker that is identified by the test to improve patient health outcomes. That is, there must be an improved treatment effect because the medicine targets either an intrinsic property of the biomarker, or a physiological process for which the biomarker is a proxy.⁷

⁷ Merlin T, Farah C, Schubert C, Mitchell A, Hiller J, Ryan P. Assessing personalized medicines in Australia: a national framework for reviewing codependent technologies. *Med Decis Making* 2013;33(3):333–42.

If the codependent technologies have evidence of both a background prognostic effect (tied to a specific biomarker) and treatment effect modification, present the net treatment effect (ie treatment effect modification controlling or adjusting for the background prognostic effect) in the submission.

Where a new codependent test and medicine targeting an as yet unproven biomarker are submitted for reimbursement, complete all information items in Subsection A7.2.

When a new biomarker(s) is proposed as part of a group of biomarkers in a submission, the aim is to gauge whether the addition of this new biomarker(s), when targeted by the medicine, results in further improvements in patient health outcomes. For a submission of this type, focus on the multiple biomarkers identified by a single test (rather than sequential testing). It is probable that the committees will have already addressed a codependent relationship between at least one of the biomarkers and the medicine and, in the interim, knowledge has advanced on how biomarkers work together to interact with the medicine. This scenario could encompass the possibility of a new or currently listed medicine, as well as a new or currently listed test. In this situation, all the items in Subsection A7.2 need to be addressed. Also seek advice from the Pharmaceutical Evaluation Branch on how to address these types of codependent technologies.

The preferred structure for an integrated codependent submission to MSAC and the PBAC is given in Figures A7.2–A7.4. This is an adaptation of the preferred structure for submissions to the PBAC for medicines outlined in the [PBAC Guidelines](#).

Figure A7.2 Preferred structure of an integrated codependent submission for Section 1

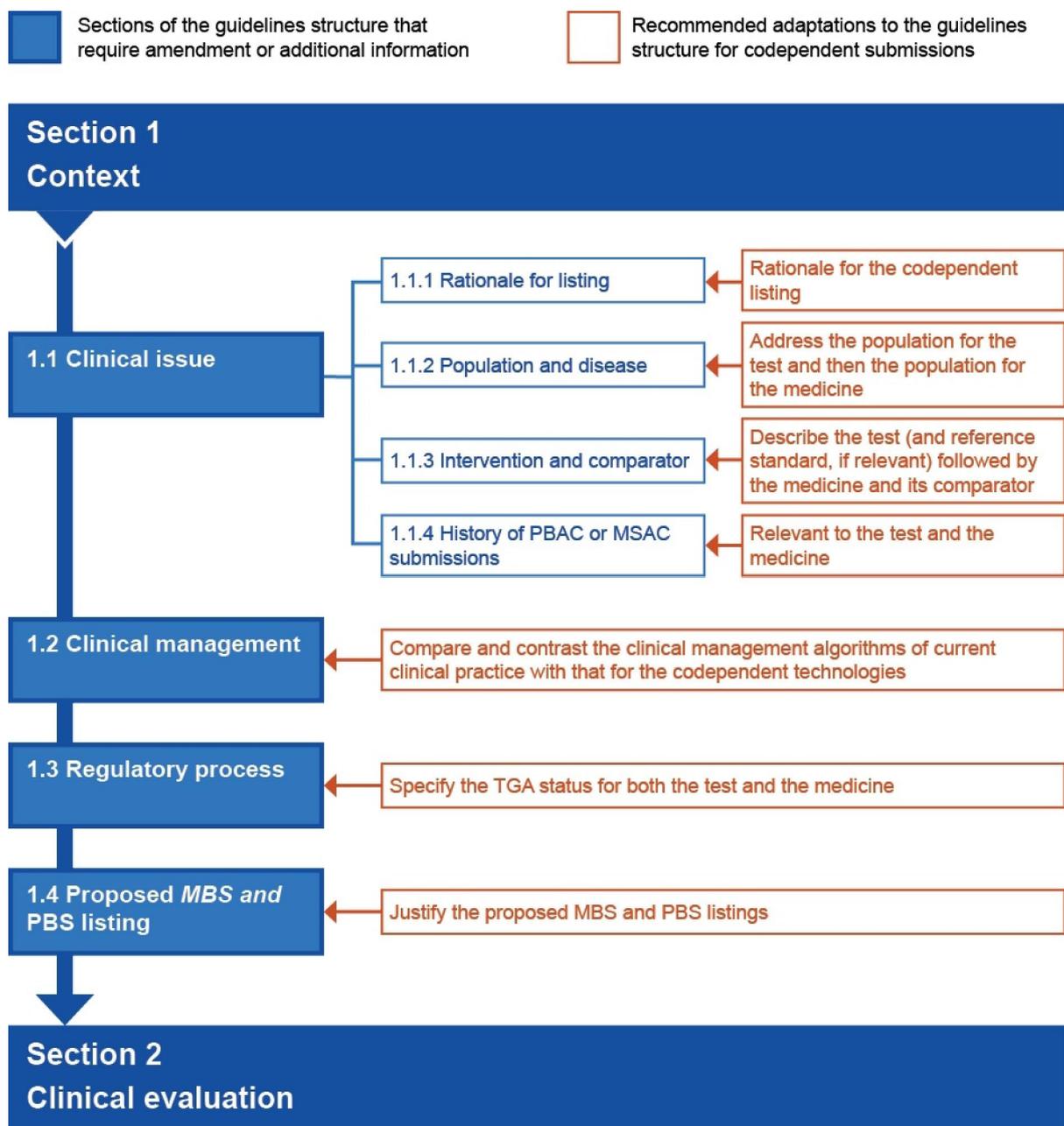
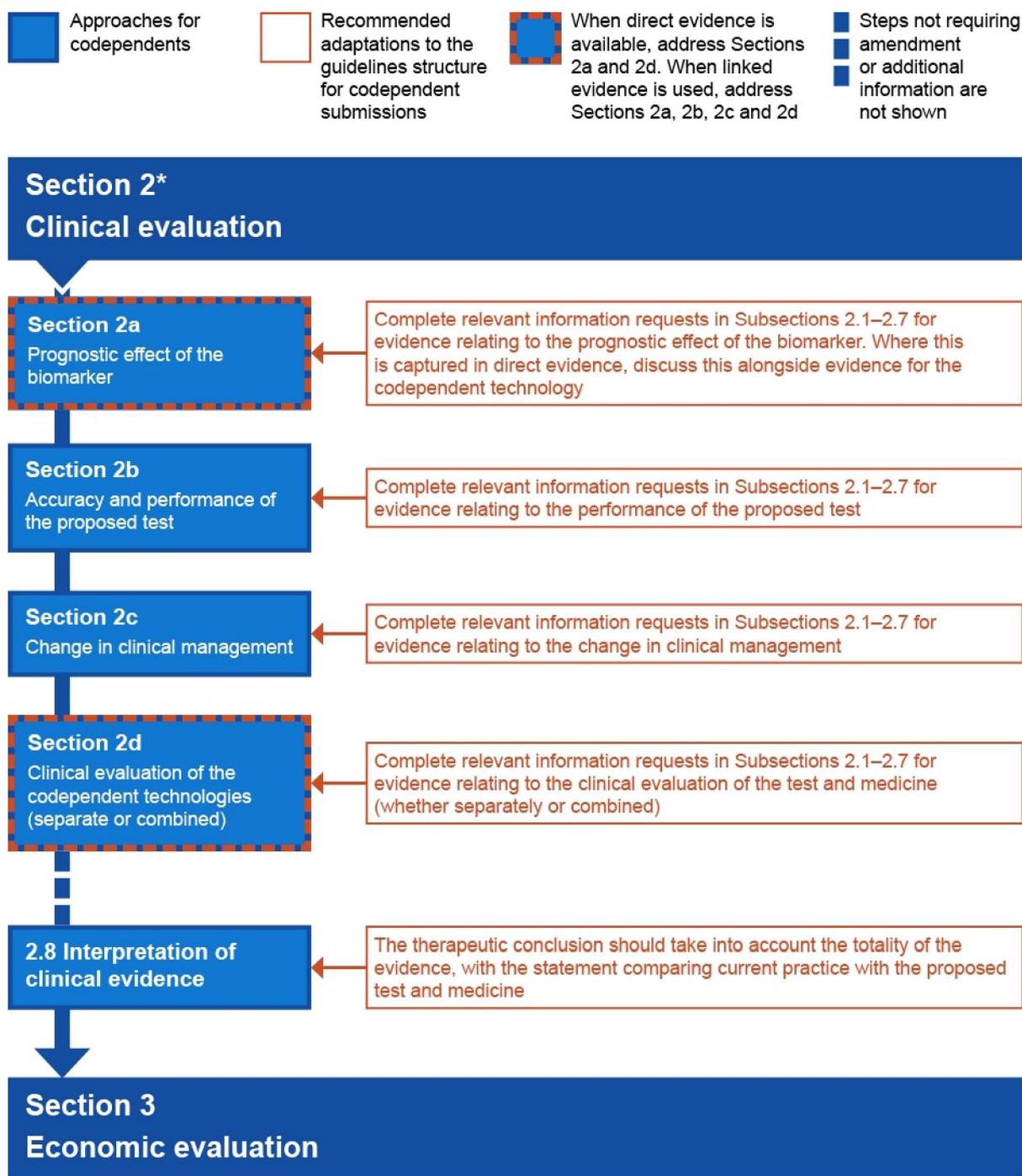
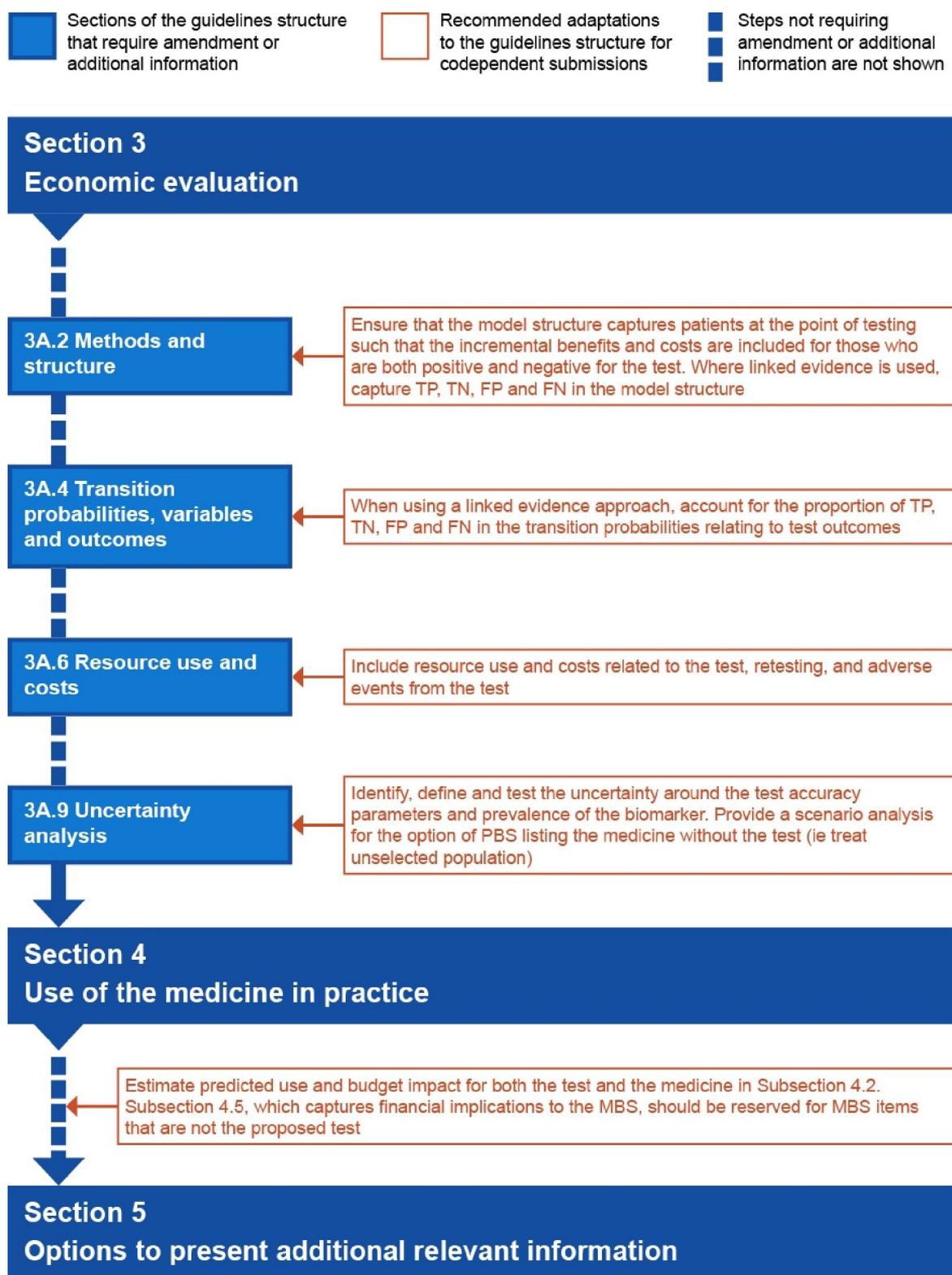


Figure A7.3 Preferred structure of an integrated codependent submission for Section 2



* The approach taken in Section 2 will depend on the available evidence. Perform a literature search to find direct evidence (of current practice vs the proposed test/medicine). Where this is not available, perform a literature review for each of Sections 2a, 2b, 2c and 2d.

Figure A7.4 Preferred structure of an integrated codependent submission for Sections 3 and 4



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

Streamlined codependent submissions

The principle of streamlined codependent submissions is to ensure that the substantive submission to seek the advice of one committee is coordinated with a less substantive submission to the other committee which has already signalled some support for the other codependent technology. In practice so far, the substantive submissions have had to be lodged with the PBAC, and streamlined submissions have been needed to ensure timely alignment of MSAC advice in the event of a PBAC recommendation.

Most experience has occurred where the PBAC has decided to defer or not to recommend a codependent medicine. In this circumstance, the purpose of the resubmission to the PBAC is to address the issues raised by the PBAC in reaching this outcome. There is also experience in the situation where the proposed medicine is in the same therapeutic class as a PBS-listed medicine, and listing is sought for essentially the same population, including with reference to patient eligibility being informed by biomarker test results. In this circumstance, the submission to the PBAC should address the items in Subsection A7.2 to the extent that is relevant, especially in relation to determining the individuals for inclusion in the proposed PBS restriction, in the submitted clinical evidence, and in the estimates of cost-effectiveness and financial implications.

Where a resubmission is being made to the PBAC, address the following matters in the streamlined resubmission to MSAC:

- a request to create or amend the MBS item for the proposed investigational medical service corresponding to the proposed medicine
- proposed wording for the MBS item descriptor (which should reflect the requested PBS restriction and the existing MBS item and/or the previous MSAC advice relevant to the proposed investigative medical service)
- a proposed MBS fee
- the costs to the MBS of the proposed listing (which should reflect the corresponding costs in the submission to the PBAC)
- a summary of the previous MSAC advice relevant to the proposed investigational medical service, and either
 - confirming that the applicant agrees with each aspect of the advice (including, where appropriate, indicating how it has followed it in the submission to the PBAC and/or MSAC); or
 - indicating where the applicant disagrees with any particular aspect of the advice, providing reasons (and an assessment of the consequences of adopting the applicant's alternative approach rather than MSAC's advice).

Where a submission is being made for a proposed medicine that is in the same therapeutic class as a PBS-listed medicine, and listing is sought for essentially the same population, including with reference to patient eligibility being informed by biomarker test results, address the first four of the matters listed above in the streamlined resubmission to MSAC. Also provide a detailed description of the testing strategy used in the trials presented in the submission to the PBAC, and the testing strategy for the corresponding trial of the comparator medicine – so that MSAC can assess any differences between these 'evidentiary standards'.

A7.2 Specific codependent technology information requests

Subsection A7P4.2 contains 62 item numbers (information requests) intended to meet the evidence requirements of the PBAC and MSAC when assessing codependent technologies for aligned reimbursement decisions. These items are accompanied by additional clarification on what is meant by the information request, as well as where it would be appropriate to present the information throughout the integrated codependent submission.

Item numbers are tagged with (T), (M) or (O), which indicate whether the item number is relevant to the test, the medicine or overlaps both. This flags the information that is primarily relevant to MSAC (T), the PBAC (M) and both committees (O).

Since listing circumstances will vary between the codependent technologies, the information needed to reduce decision-maker uncertainty will also vary. Follow the guidance given in Part A, Sections 1–5, about the medicine for all integrated codependent submissions. In addition, new items and/or expanded information are requested to address specific codependency issues. These are organised here by the main sections of a submission (ie Sections 1–5). Refer to Subsection A7.1 for an overview of which codependency information items apply to an integrated or streamlined submission, and to Figure A7.2 for how the submission should be structured.

Section 1 – Context

The following information requests are relevant to Part A, Section 1, of a submission to the PBAC.

Details about the biomarker, the test and the medicine

ADDITIONAL INFORMATION REQUESTS

- 1 (O) Describe current reimbursement arrangements for the test and the medicine
- 2 (T) Identify the sponsor of the test
- 3 (M) Identify the sponsor of the medicine
- 4 (O) Describe the biomarker
- 5 (T) Describe the proposed test
- 6 (O) Describe the medical condition or problem being managed
- 7 (O) Describe the relevant clinical management pathways

1 (O) Current reimbursement arrangements

Include in Subsection 1.4

Indicate whether the proposed biomarker(s) has been previously accepted as valid by MSAC and the PBAC for the proposed clinical indication (eg validated using another test). Describe current reimbursement arrangements for the test and the medicine.

The response to this item defines whether the submission is integrated or streamlined (see Subsection A7.1).

2 (T) Test sponsor

Include in Subsection 1.1.1 or 1.1.3

Identify the source(s) of the test options (eg commercial sponsor, research laboratory, National Association of Testing Authorities [NATA]-accredited pathology provider, pathology practice). This includes clinical sponsors of tests, given that tests guide both the initiation and cessation of therapy. If a specific test (eg the evidentiary standard; see Item 5) is not specified, this item is not needed.

3 (M) Medicine sponsor

Include in Subsection 1.1.1 or 1.1.3

Identify the sponsor of the medicine. This enables a different sponsor to be identified, if necessary, for each component of the codependent technology.

4 (O) Biomarker

Include in Subsection 1.1.2

Describe the biomarker in a way that is consistent with the proposed MBS item descriptor and to enable differentiation from other possible biomarkers. Additional detail can be provided at Item 8.

The most common type of integrated codependent submission has involved pharmacogenetic technologies assessing genetic DNA biomarkers whereby one genetic locus at a time is evaluated. However, codependent technologies that include genetic panel testing or genomic testing (ie assessment across the genome, testing hundreds or thousands of loci simultaneously) can also be submitted.

5 (T) Proposed test(s)

Include in Subsections 1.1 and 1.4

First describe the evidentiary standard test method (ie the test used in the key evidence supporting the requested listing). Include sufficient detail for a laboratory technician to be able to perform it. If more than one test is proposed or available, then specify the range of techniques used to measure the biomarker (eg polymerase chain reaction, high-resolution melting), and indicate which method, if any, is regarded as the reference or ‘gold standard’ test.

List the other available test options that fall within the scope of the proposed MBS item descriptor. If other test options are available in Australia, or the evidentiary standard test is not available in Australia, then provide a comparison of all available tests for the biomarker that fall within the scope of the requested MBS item descriptor.

Include the proposed MBS item descriptor by modifying Subsection 1.4 to ‘Proposed MBS and PBS listing’.

6 (O) Medical condition or problem being managed

Include in Subsection 1.1.2

Describe the population proposed for testing for the biomarker in terms of what previous tests have been undertaken or what clinical signs are present. Describe whether the proposed population has been enriched in terms of biomarker prevalence.

Issues to consider when judging the value of adopting any enrichment or triaging strategy include:

- the quantified effect on the Australian prevalence of being test positive (and hence on the number of patients who would need to be tested to target treatment)
- the confidence in the clinical diagnosis being able to identify likely patients with the biomarker and to minimise erroneous inclusions and exclusions from the patient pool selected as eligible for the test
- the consequences of misallocation of treatment due to false positive or false negative test results brought about by these erroneous inclusions and exclusions, which can vary across clinical settings – for example, between first-line therapy (where there are effective alternative treatments) and last-line therapy (where there are not)
- the amount of tissue needed to make multiple types of diagnosis when tumour tissue is limited (eg via fine needle aspirate biopsy) and so the need for larger tumour samples or re-sampling has implications for harm to patients and costs to the health care system
- whether the clinical diagnosis itself might also modify the treatment effect, independent of the testing strategy (eg the effect of the proposed medicine might vary according to histology type, in addition to biomarker status).

If different test result thresholds are likely to determine eligibility for the medicine, or if eligibility for the medicine is determined subjectively, consider providing alternative requested PBS listings in Subsection 1.4.

7 (O) Clinical management pathways

Include in Subsection 1.2

Describe and compare the proposed clinical management of a typical patient up to the point of being offered the proposed test and subsequent therapy with the proposed medicine, as compared with the currently existing clinical pathway(s) where the proposed test and/or medicine is not available.

Ensure that the clinical management pathways outline all alternative tests/test strategies (whether the tests occur in series or concurrently) and all alternative treatments (including nonmedicine treatments) for the target clinical indication, both with and without knowledge of the patient's biomarker status.

Identify tests and treatments that are commonly used and likely to be supplemented or replaced by the codependent technologies (see Item 14).

If it is important for patients with a rapidly progressive disease or condition to ensure that a timely test result is available to determine eligibility for the medicine, indicate whether the test is likely to be performed earlier in disease or condition progression than currently (also see Item 12).

The nomination of when to test compared with when to treat can be influenced by many factors, including:

- the urgency of knowing the test result to inform the start of medicine therapy
- the costs of block retrieval and costs (and patient harms) of obtaining new samples
- the confidence that the sample or previously obtained test result represents the status of the patient at the time of deciding which treatment to start (eg the stability of a mutation over

time or in response to previous treatment, or between the primary tumour and metastatic disease)

- the clinical and cost-effectiveness consequences of misallocation of treatment due to false positive or false negative conclusions based on changes in mutation status.

A ‘no testing’ pathway and dealing with data scarcity

To demonstrate the test’s impact on patient health outcomes, indicate a pathway where testing for the biomarker is not undertaken. Then estimate the effectiveness (and cost-effectiveness) of the medicine using the economic model both with and without use of the test (see Item 37). This approach is requested because it may be more cost-effective to provide the medicine without the test if the test has poor accuracy and/or the medical condition is prevalent.

Because data are often scarce, the aim of codependent technology evaluation is to maximise the use of the available evidence on the two technologies. If the effect of the medicine in the total population is being estimated and data are not available on the biomarker negative population, it may be sufficient to use data/transition probabilities associated with the total population (ie biomarker positive and negative) if the prevalence of the biomarker is low in the total population and if sensitivity analyses are conducted to vary the estimates/inputs within a plausible range.

If the prevalence of the biomarker is high in the total population, it will be important to test whether it would be more cost-effective to deliver the medicine without use of the test. If the ‘biomarker negative’ arm is receiving usual care, then an effect that was consistent with treatment effects before the introduction of the new medicine would be expected. When the new medicine is replacing usual care and there are no data on the biomarker negative population, or in the event that there is a fairly even distribution of the biomarker negative and positive in the total population, then collect and extract data on false positive patients (ie true negatives with incorrect test result) to determine response to therapy in the alternative condition (biomarker negative group).

Rationale for the codependency

ADDITIONAL INFORMATION REQUESTS

- 8 (O) Define the biomarker
- 9 (O) Provide a biological rationale for targeting the proposed biomarker with the proposed medicine
- 10 (O) Define any other biomarker(s) that modify the comparative treatment effect of the medicine
- 11 (O) Define the prevalence of the condition being targeted in the population that is likely to receive the test

8 (O) Definition of the biomarker(s)

Include in Subsections 1.1.2 or 1.1.3, and 5.3

Describe the nature of the biomarker (eg single nucleotide polymorphisms, mutation, copy number variation).

Where relevant, include the following elements describing the context for the biomarker:

- the disease or condition
- the specific function of the biomarker
- the critical parameters which define when and how the biomarker should be identified.

If the biomarker is a specific genetic mutation, describe exactly what the test is identifying (eg an expression microarray of tumour tissue that identifies a cancer that can be inhibited by activating a particular pathway). Categorise any mutation biomarker as germline or somatic, or both. If a mutation biomarker is classified as germline, then consider issues related to heritability in Subsection 5.2 (eg testing of relatives and genetic counselling, ethical and medico-legal implications of testing).

Issues to be considered when judging the optimal definition of the biomarker include the following:

- the patient and cost consequences of different sampling needs to support different test options when it is difficult to obtain sufficient material to test for the biomarker (eg tumour samples)
- the prevalence of the different types of mutations in the disease or condition identified, noting that the evidence is likely to be greater for common mutations compared with rare mutations
- the frequency and predictive consequences of multiple mutations in a single sample (eg tumour heterogeneity and mosaicism), or indeed the impact of mutations in genes other than those nominated that may influence the effectiveness of the proposed medicine
- the evidence of impact on health outcomes for each type of mutation, either directly (eg if it is included in the evidentiary standard definition and ideally shows treatment effect modification), or from in vitro studies, or by inference (eg if there is a biologically plausible basis to differentiate among different types of mutations, such as activating or inactivating mutations, or mutations that predict resistance, sensitivity or neutrality to the medicine effect)
- the clinical and cost-effectiveness consequences of misallocation of treatment because of false positive or false negative results based on these conclusions.

9 (O) Biological rationale for targeting that biomarker(s)

Include in Subsection 1.1

Present the initial evidence that was used to select the biomarker for targeting with the proposed medicine. Describe and explain the overall approach to the selection of the biomarker, including methods and relevant aspects of study design and statistical analysis. Describe the rationale for selection of the population sample studied in the biomarker qualification.

Where the biomarker is genetic, present the criteria used for selecting candidate genes (eg candidate by position or by function, based on expression profiling data). Justify, using molecular biological or pharmacological principles, the plausibility of treatment effect modification (ie interaction) between the biomarker itself and the medicine, or, alternatively, between the medicine and another factor for which the biomarker is a proxy. Advise whether this biological rationale preceded the data collection underpinning the key evidence.

10 (O) Other biomarker(s) that modify the comparative treatment effect of the medicine

Include in Subsections 1.1 and 5.3

If testing for any other biomarkers is already reimbursed for targeted treatment with the medicine for the same condition, consider these codependent technologies in the choice of comparator.

If another biomarker is a genetic mutation, then:

- provide details on the specific mutation and the nature of the mutation
- explain whether the treatment effect in patients with this other mutation is consistent with the effect under consideration.

Note: This item may be relevant even if these other biomarker(s) are claimed but a test for the biomarker is not yet reimbursed.

11 (O) Prevalence of the condition being targeted in the population that is likely to receive the test

Include in Subsections 1.1 and 3A.4

Estimate the prevalence of the condition being targeted as measured by the true positive biomarker; this is relevant to calculate the performance of a test in terms of its negative and positive predictive value.

Indicate in Subsection 1.1 whether there is a ‘gold standard’ or reference standard test to determine whether a patient is true positive for the biomarker. Provide evidence to estimate the prevalence of the biomarker in an Australian population.

In the absence of an accepted reference standard test to correctly identify biomarker status, use an alternative appropriate methodology to estimate the prevalence (eg adjudication by a third test or sensitivity analyses of the prevalence of the biomarker given different assumptions).

The denominator for the prevalence calculation (the source population) is the number of patients considered eligible for the test according to the proposed MBS item descriptor.

The source population consists of patients in the clinical pathway up to the point of being offered the test or the medicine in the absence of the test.

Proposed impact of codependent technologies on current clinical practice

ADDITIONAL INFORMATION REQUESTS

- 12 (T) State whether the proposed test results are expected to be consistent over time, including over the course of the disease or condition
- 13 (T) Indicate whether the proposed test could be used with other treatments and/or for other purposes
- 14 (T) State whether the proposed test is additional to another test(s) currently defining the condition, or a replacement test, or both (ie depending on the test result, replace some tests or be additional to other tests)
- 15 (T) Describe how the proposed test will be offered in Australia
- 16 (T) Identify the biospecimen or sample needed for the test, and whether this specimen needs to be collected specifically for the test or has already been collected for another purpose
- 17 (T) Describe the need for subsequent testing to monitor the development of new somatic mutations and/or to guide dosage or cessation of therapy with the codependent medicine (if relevant)
- 18 (O) Indicate whether the proposed medicine can be used with other specific tests for that biomarker, other than the test proposed. Describe the available methods for testing for the biomarker

12 (T) Consistency of the test results over time

Include in Subsections 1.1 and 1.2

Where test results for a patient may change over time (eg between a primary tumour and subsequent metastases in cancer), provide sufficient detail to clarify the relationship and timeframes between test results and the appropriateness of treatment.

For example, rat sarcoma viral oncogene homolog testing of the primary colorectal cancer tumour is usually representative of the findings in metastases, regardless of therapy. In another example, epidermal growth factor receptor results change with, for example, exposure to radiotherapy, so the results of testing the primary tumour may not be representative of what is happening in non-small cell lung cancer metastases.

13 (T) Use of the proposed test with other treatments and/or for other purposes

Discuss in Subsections 1.2 and 4.2

If other treatments or purposes are relevant, consider whether their use is currently reimbursed or whether there is the possibility of leakage. Refer to the clinical management pathways provided in response to Item 7.

14 (T) Use of the test in the clinical management pathway

Include in Subsections 1.2, 3A.6, 4.2 and 4.5

Refer to the clinical management pathways provided in response to Item 7. The test is most likely to be an additional test, although occasionally, if the biomarker is a strong predictor, it could replace another test in the pathway.

15 (T) Provision of the test in Australia

Include in Subsections 1.2 and 1.3

Indicate whether the test is likely to be widely accessible or available in a few selected laboratories across the country. Explain how the test would be undertaken in practice, and what impact it would have on the patient and health professionals (Subsection 1.2).

Specify the TGA status of the proposed test options (if relevant). Assess access and quality assurance issues. Identify how many laboratories offering the test have NATA accreditation for that test (Subsection 1.3).

16 (T) Specimen or sample collection

Include in Subsection 1.1 (plus Subsections 2.5, 2.6, 2.7, 3A.6, 4.2 and 5.1 if a new specimen needs to be collected)

Identify the biospecimen or sample needed for the test – for example, blood, tumour material (formalin-fixed paraffin embedded [FFPE] or fresh), bone marrow, cytology specimen or mouth swab.

Identify whether this specimen needs to be collected specifically for the test or has already been collected for another purpose. For example, tumour already removed can be tested if archival FFPE is available and the test can identify the biomarker from this tissue.

If a new specimen needs to be collected, specify the costs (Subsections 3A.6 and 4.2), risks (Subsections 2.5–2.7) and feasibility of collecting the sample (Subsection 5.1). In some instances, such as a blood sample, the costs and risks would be trivial. In other instances, such as when a new biopsy is required, there may be significant costs as well as safety risks for the patient.

17 (T) Use of the test for monitoring purposes (if relevant)

Include in Sections 1.1–1.4, 3A.4 and 4.2

If relevant, describe the need for subsequent testing to monitor the development of new somatic mutations and/or to guide dosage or cessation of therapy with the codependent medicine.

This will impact on the clinical need for the proposed test (discuss in Subsections 1.1–1.4), as well as associated transition probabilities (Subsection 3A.4) and costs (Subsections 3A.6 and 4.2). If a new biopsy is required, cross-reference to Item 16.

18 (O) Availability of other tests for the biomarker

Include in Subsection 1.1 (if other tests are publicly funded) or Subsection 5.1 (if other tests are not publicly funded)

Indicate whether the proposed medicine can be used with other specific tests for that biomarker, other than the test proposed. Describe the available methods for testing for the biomarker.

If other tests are publicly funded to identify the biomarker, amalgamate this item with Item 10. If other tests are available or are emerging but are not yet publicly funded, address this item in Subsection 5.1.

Section 2 – Clinical evaluation

The following section contains information requests for establishing the clinical benefit of the codependent technologies in terms of patient health outcomes.

An integrated codependent submission may need to present more than one Section 2 to support the proposed listing of the medicine and the test. The extent of information requested is discussed in Subsection A7.1, and will be further contingent upon the availability of direct evidence or the need to use linked evidence. An overview is shown in Figure A7.2.

The following general approach to presenting a submission may be appropriate:

Approach based on direct evidence

- Section 2a – prognostic effect of the biomarker
- Section 2d – clinical evaluation of the codependent technologies (evidence of combined use)

and/or

Approach based on linked evidence

- Section 2a – prognostic effect of the biomarker
- Section 2b – performance and accuracy of the proposed test
- Section 2c – change in clinical management
- Section 2d – clinical evaluation of the codependent technologies (separate)

Each Section 2 should follow the steps presented in Part A of these guidelines.

Direct evidence approach

ADDITIONAL INFORMATION REQUESTS: DIRECT EVIDENCE

- 19 (O) Determine whether the biomarker test can predict differences in patient health outcomes irrespective of the clinical management provided
- 20 (O) Indicate whether the search for direct evidence was comprehensive and whether the selection process was unbiased
- 21 (O) Assess bias, confounding and the impact of chance on the findings presented in the direct evidence

Section 2a Evidence of prognostic effect of the biomarker

19 (O) Prognostic effect of the biomarker

Include in Section 2

Determine whether the biomarker test can predict differences in patient health outcomes irrespective of the clinical management provided.

It is important to discriminate the background prognostic effect of biomarker status from the impact of any treatment effect modification associated with the biomarker. This requires a comparison of outcomes in patients receiving usual care conditioned on the presence or absence of the biomarker.

Use the approach described in Section 2 to systematically review the evidence of the presence or absence of a prognostic effect of the biomarker, as identified by the proposed test. Searching the literature for prognostic information is typically more complex than searching for intervention (treatment) studies. For example, literature searches would not be limited to randomised controlled trials. Advice from an information specialist is recommended.

Section 2d Clinical evaluation of the codependent technologies (combined)

Most of the information needed for this section is already covered by the information requests in Part A of the [PBAC Guidelines](#). Additional requests are given below.

When ‘direct evidence’ is available this should be presented in the submission. Direct evidence can include the following trial designs (illustrations of the different trial designs are provided in Merlin et al,⁸ supplemental data 1 file):

- **Double-randomised controlled trial:** A trial that randomises patients to use of the test or not, then randomises to use of the medicine or its main comparator, and then follows patients to measure the effect of the treatment on clinical (health) outcomes.
- **Single-randomised controlled trial of test:** A trial that randomises patients to use of the test or not, and then follows patients to measure the effect of targeted treatment with the new medicine on clinical (health) outcomes.
- **Prospective biomarker-stratified design:** A trial that prospectively tests eligible patients, then randomises those that are test positive or negative to use of the medicine or its main comparator, and then follows participants to measure the effect of treatment on clinical

⁸ Merlin T, Farah C, Schubert C, Mitchell A, Hiller J, Ryan P. Assessing personalized medicines in Australia: a national framework for reviewing codependent technologies. *Med Decis Making* 2013;33(3):333–42.

(health) outcomes. The ‘no test’ or ‘alternative test’ arm is not included in this biomarker-stratified design.

- **Retrospective biomarker-stratified design:** A trial that randomises eligible patients to use of the medicine or its main comparator, then follows participants to measure the effect of treatment on clinical (health) outcomes, and then analyses results across subgroups of patients defined by whether they are positive for the test (or biomarker) or whether they are negative to the test (or biomarker).

The design of a double-randomised controlled trial can be used as a template within which the available direct clinical evidence can be hypothetically mapped (see Merlin et al, supplemental data 2 file). Identify areas where information is missing in the economic modelling in Section 3.

For example, given that a single-randomised controlled trial of a test does not provide information on the test (biomarker)-medicine relationship (ie evidence that the biomarker is a treatment effect modifier and/or has a prognostic effect), consider supplementing this evidence with information from prospective and/or retrospective biomarker-stratified study designs.

As prospective and retrospective biomarker-stratified study designs are without a ‘no testing’ trial arm (ie to determine biomarker status), the impact of false positive and false negative test findings cannot be determined from the reported patient health outcomes. Consider providing supplementary information from the linked-evidence approach described below, so that a comparison of the proposed test/test strategy and existing test/test strategy can be made with respect to their relative diagnostic accuracy or test performance.

Retrospective biomarker-stratified study designs may use archival tissue/sampling to determine biomarker status. Exercise caution when interpreting results from these studies, because biomarker status might change over time, particularly if there is evidence that an intervening treatment may modify the biomarker result.

20 (O) Selection of the direct evidence

Include in Subsections 2.1 and 2.2

Indicate whether the search for direct evidence was comprehensive and whether the selection process was unbiased. Present a systematic review of direct evidence (study designs given above) concerning the proposed biomarker test and the proposed medicine, with prespecified inclusion/exclusion criteria and study selection outlined in a PRISMA flowchart⁹ (ie indicating how trials were selected and the reasons why any potentially relevant trials were excluded).

21 (O) Quality of the direct evidence

Include in Subsections 2.3 and 2.6

Assess bias, confounding and the impact of chance on the findings presented in the direct evidence. Give particular attention to the impact of selection bias and confounding with respect to any subgroup analyses. For example, were the subgroup analyses prespecified (involving stratified randomisation) and was blinding maintained? Was the subgroup

⁹ Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62(10):1006–12.

analysis exploratory (eg determined on the basis of retrospectively obtained samples)?
Were the results adjusted for potential confounders?

Linked-evidence approach

ADDITIONAL INFORMATION REQUESTS: LINKED EVIDENCE

- 22 (T) Describe the analytical performance of the proposed test
- 23 (T) Define the reference standard or a gold standard against which the performance of the proposed test will be measured
- 24 (T) Indicate whether the search for evidence on the diagnostic accuracy or predictive accuracy of the proposed test was comprehensive, and whether the evidence selection process was unbiased
- 25 (T) Indicate whether the evidence reporting on the diagnostic accuracy or predictive accuracy of the proposed test is (i) of good quality and (ii) applicable to the requested MBS target population
- 26 (T) Report on the performance of the proposed test in terms of its diagnostic accuracy or predictive accuracy. If several tests are proposed or no specific test is specified, indicate which test has the best performance. If test accuracy cannot be determined, calculate agreement or concordance between tests
- 27 (T) Indicate which test is the most accessible/available/used. (Only relevant if several tests are proposed or no specific test is specified)

A full linked-evidence approach is only meaningful when the evidence for the proposed test and the evidence for the proposed medicine have been generated in similar patient populations, and so it is clinically sensible to link the two datasets. If the test identifies patients earlier or with a different spectrum of disease than the patients in whom the medicine has been trialled, then it is not clinically sensible to link this evidence. In this circumstance, present direct evidence of the impact of biomarker testing on patient health outcomes.

Section 2b Test performance and accuracy

22 (T) Analytical test performance

Include alongside Subsection 2.5

Analytical test performance assesses how accurately and how consistently the test identifies biomarker status (eg the coefficient of variation and other appropriate statistics). Present any differences across laboratories in how they characterise test results (eg a kappa statistic or other concordance statistic). Identify whether there is an external quality assurance program by which laboratories can benchmark their assays, and whether the test is performed and interpreted accurately and reliably. An assessment of the analytic validity of the evidentiary standard test, relative to other existing test options, would be helpful for decision making.

23 (T) Reference standard or a gold standard for test performance

Include in Subsection 1.1

Define the reference standard or a gold standard against which the performance of the proposed test will be measured. Provide evidence that the reference standard is considered to be accurate and is an appropriate benchmark. (This is not needed if the reference standard has already been identified and ratified by the Protocol Advisory Subcommittee [PASC].)

Note: The reference standard is not necessarily the same as the relevant comparator for the codependent test. The comparator is the current test/test strategy being used in the absence of the proposed test; this may be different to the benchmark (reference standard) test for determining test accuracy. For example, a reference standard for a new genetic

test might be Sanger sequencing, but the comparator for the new genetic test might be a high-resolution melting method.

Also note that the comparator for the test is different to the comparator for the medicine.

Test accuracy

In the instance where a reference standard is available

If a reference standard is available, test performance is determined using diagnostic accuracy measures (eg using a cross-sectional study design). Compare the proposed test to the designated reference standard by cross-classifying the test results of patients who are representative of the intended population receiving the test. The proposed test will be referred to as the ‘evidentiary standard’ if it is the test used in the key evidence presented in the submission.

Use the reference standard designated by the PASC, or select and justify the choice of a reference standard if this has not been previously specified by the PASC.

In the instance where no reference standard is available

If no reference standard is available, test performance can be determined using predictive accuracy (eg using a longitudinal study design, with the clinical outcome providing the benchmark for identifying whether the patient does or does not have the condition).

If a reference standard is not available or is unacceptable for the requested use and/or the requested population, consider the various options for dealing with imperfect or missing reference standards in the guidance provided by Reitsma et al.¹⁰ If the guidance by Reitsma et al is not followed, justify the approach used.

Note that if sensitivity and specificity of the proposed test are to be estimated using a composite/constructed standard, the new reference standard should be developed independently from the analysis of results of the proposed test (ideally, in advance of collecting any specimens). Consult with statisticians and health professionals before constructing the reference standard.

If measures of concordance or agreement (positive per cent agreement and negative per cent agreement) are calculated instead of measures of test performance, ensure that the terms ‘sensitivity’ and ‘specificity’ are not used, as these estimates are not of test accuracy but of agreement between the proposed test with the nonreference standard.¹¹

24 (T) Selection of the evidence on test accuracy

Include in Subsections 2.1 and 2.2

Indicate whether the search for evidence on the diagnostic accuracy or predictive accuracy of the proposed test was comprehensive and whether the evidence selection process was unbiased.

¹⁰ Reitsma J, Rutjes A, Khan K, Coomarasamy A, Bossuyt P. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009;62(8):797–806.

¹¹ United States Food and Drug Administration. Guidance for industry and FDA staff: statistical guidance on reporting results from studies evaluating diagnostic tests. Rockville, MD: Food and Drug Administration, Center for Devices and Radiological Health, US Department of Health and Human Services, 2007.

For example, systematically review test performance studies for the proposed test (evidentiary standard) with prespecified inclusion/exclusion criteria and a PRISMA flowchart. Indicate how test performance studies were selected and the reasons why any potentially relevant studies were excluded.

Note that literature searching for test performance studies will need to be more exhaustive than for treatment trials, because indexing and filtering of these studies is less reliable in bibliographic databases. Suggestions for identifying test accuracy studies in literature searches is given in Chapter 7 of the *Cochrane handbook for systematic reviews of diagnostic test accuracy*.¹²

25 (T) Quality of the test accuracy studies

Include in Subsection 2.3

Indicate whether the evidence reporting on the diagnostic accuracy or predictive accuracy of the proposed test is of good quality and applicable to the requested MBS target population.

This can be done using a [QUADAS-2 assessment](#)¹³ for each test accuracy study in terms of risk of bias and applicability for use in Australia on the domains of patient selection, index test, reference standard, and flow and timing.¹⁴ Display the results as a table or graph. Note that QUADAS-2 is a critical appraisal tool, whereas tools like STARD and the ACCE framework are used for reporting test accuracy studies and genetic test interventions, respectively.

26 (T) Performance of the proposed test

Include in modified version of Subsection 2.5

Report on the diagnostic accuracy or predictive accuracy of the proposed test. If several tests are proposed or no specific test is specified, indicate which of the tests has the best performance. If test accuracy cannot be determined, calculate agreement or concordance between tests.

Diagnostic accuracy or predictive accuracy

Provide test performance measures such as sensitivity, specificity, likelihood ratios, positive and negative predictive values, or area under the receiver-operator characteristic curve. Ensure that test failure (invalid results) for either test is documented (proportion of failures), but do not include these results in the test accuracy estimates.

Summarise (if a meta-analysis is performed) test accuracy measures and approaches, as appropriate to the available evidence base. Consider the presence of heterogeneity and/or test threshold effects. Various methods are described by Takwoingi et al.¹⁵

¹² De Vet H, Eisinga A, Riphagen I, Aertgeerts B, Pewsner D. Chapter 7: searching for studies. *Cochrane handbook for systematic reviews of diagnostic test accuracy* version 04 [updated September 2008], The Cochrane Collaboration 2008.

¹³ www.bristol.ac.uk/social-community-medicine/projects/quadas/quadas-2

¹⁴ Whiting P, Rutjes A, Westwood M, Mallet S, Deeks J, Reitsma J et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155(8):529–36.

¹⁵ Takwoingi Y, Guo B, Riley R, Deeks J. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2015;June 26 [online first](doi: 10.1177/0962280215592269).

When interpreting the results of the studies, prioritise assessing the trade-offs in false positive and false negative test findings. For example, consider whether there is a clinically accepted test performance level below which a new test should not be used (ie either false positives are too great or false negatives are too great) for the intended purpose.

The main issues to consider are that:

- false negatives are of greater concern when the clinical setting of the proposed medicine is as last line with best supportive care as its comparator
- false positives are of greater concern when the proposed medicine is being compared with effective alternatives.

If the reference standard being used to determine test accuracy is imperfect, and it is therefore unclear whether the false positives or false negatives ascertained using the codependent test are actually true positives and true negatives, provide evidence of the clinical (health) outcomes of those patients found to be false positive or false negative and report these under the ‘Direct evidence’ section, if possible.

The positive predictive value and negative predictive value should also be calculated, since these data are key to the calculation of transition probabilities in Subsection 3A.4.

Calculate estimates of sensitivity and specificity, adjusted to correct for any (verification or partial verification) bias that may have been introduced by not using the reference standard to its fullest extent (ie to verify all the results obtained with the new test).¹⁶

Agreement or concordance

If agreement data are provided, rather than test accuracy data, measures such as positive predictive value and negative predictive value (used in Section 3) cannot be calculated since the subjects’ condition (as determined by a reference standard) is unknown. In this situation, report the 2×2 table of results, comparing the candidate test with the nonreference standard test, and report the agreement measures along with their confidence intervals or kappa statistics. Alternatively, odds ratios could be reported indicating the likelihood of an outcome, given that particular test result.

27 (T) Test availability

Include in Subsection 5.1

Consider which test is the most accessible/available/used. (Only relevant if several tests are proposed or no specific test is specified.)

Where testing is both complex and uncommon, there are important quality and pathology laboratory performance considerations that need to be addressed – for example, biospecimens may need to be shipped to a small number of high-throughput pathology laboratories.

Where biospecimens are relatively transportable, it may not always be an access advantage to bring the test closer to the patient.

¹⁶ Reitsma J, Rutjes A, Khan K, Coomarasamy A, Bossuyt P. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009;62(8):797–806.

Section 2c Change in clinical management

ADDITIONAL INFORMATION REQUESTS: LINKED EVIDENCE

- 28 (O) Substantiate whether knowledge of the test result will cause a change in the management of the patient by the treating clinician. Identify instances where management would not change, despite the test indicating that the biomarker is present

28 (O) Change in management of the patient because of knowledge of test result

Include in Subsections 2.1–2.5

Substantiate whether knowledge of the test result will cause a change in the management of the patient by the treating clinician. Identify instances where management would not change, despite the test indicating that the biomarker is present.

There may be ‘leakage’ issues identified through an assessment of the ‘change in management’ part of the linked evidence. Often a test is done to rule out use of a medicine (eg to avoid potential medicine-related adverse events or the development of resistance), but the medicine is given anyway, or, alternatively, the test is used to select a specific medicine, but the medicine is not provided. Since codependent tests are used to guide therapeutic decisions, explicitly address this by searching for literature that reports on the management of patients identified with and without the biomarker.

Section 2d Clinical evaluation of the codependent technologies (separate)

ADDITIONAL INFORMATION REQUESTS: LINKED EVIDENCE

- 29 (T) Identify any safety considerations that will impact on the entire process of testing
- 30 (M) Indicate whether the search for evidence on the therapeutic effectiveness of the proposed medicine was comprehensive and whether the evidence selection process was unbiased
- 31 (M) Indicate whether the evidence reporting on the therapeutic effectiveness of the proposed medicine is of good quality
- 32 (O) Provide evidence (if relevant) of treatment effect modification (ie interaction) as a consequence of biomarker status
- 33 (O) Provide evidence (if relevant) that using the test results in better targeting of patients that are likely to respond most to the medicine (ie by using the prognostic effect of the biomarker to determine the baseline risk of disease or condition progression)
- 34 (O) Indicate whether the effect of the medicine, as conditioned by the test or biomarker result, has a clinically important and statistically significant effect on patient-relevant health outcomes (both safety and effectiveness)

29 (T) Safety concerns regarding the proposed test

Include in Subsection 2.7

Identify any safety considerations that will impact on the entire process of testing. For example, patient contraindications to the testing procedure, required biospecimen size, additional risk of harm (with reference to Item 16), or processing time impacting on treatment initiation.

30 (M) Selection of the evidence on the therapeutic effectiveness of the medicine

Include in Subsections 2.1 and 2.2

Indicate whether the search for evidence on the therapeutic effectiveness of the proposed medicine was comprehensive and whether the evidence selection process was unbiased.

This evidence should include:

- the therapeutic effectiveness of the medicine when conditioned by the test or biomarker result
- the therapeutic effectiveness of the medicine in unselected patients (where biomarker status has not been determined).

For example, present a systematic review of the available comparative clinical evidence of the proposed medicine versus its comparator in patients with and without the biomarker, as well as the available comparative clinical evidence of the proposed medicine versus its comparator when patient biomarker status is not known.

Ensure that the systematic review has study inclusion/exclusion criteria delineated, and include a PRISMA flowchart indicating how trials were selected and the reasons why any potentially relevant trials were excluded.

31 (M) Quality of therapeutic effectiveness evidence

Include in Subsection 2.3

Indicate whether the evidence reporting on the therapeutic effectiveness of the proposed medicine is of good quality.

Assess bias, confounding and the impact of chance on the results. Particular attention should be given to the impact of selection bias and confounding on any subgroup analyses. For example, were the subgroup analyses prespecified (stratified randomisation) and was blinding maintained? Or was the subgroup analysis exploratory (determined on the basis of retrospectively obtained samples)? Were the results adjusted for potential confounders? Where the study design involves biomarker positive patients only, assess study quality according to the usual guidance in Subsection 2.3.

Depending on the study design, confounding may occur where biomarker status is a prognostic factor and when there are imbalances in biomarker status in the proposed medicine and comparator medicine trial arms.

32 (O) Evidence of treatment effect modification

Include in Subsection 2.6

Provide evidence (where available) of treatment effect modification (ie interaction) as a consequence of biomarker status.

For example, is there evidence of substantial variation in a measure of relative treatment effect between the proposed medicine and comparator/usual care trial arms after stratifying on biomarker status?

Treatment effect modification in this setting identifies a relationship between the biomarker and the medicine, which is likely to be unique or limited to companion tests assessing a particular biomarker and medicines with a particular mechanism of action (cross-reference to Item 9). This means that both technologies are needed to produce or optimise a clinical benefit.

33 (O) Evidence of prognostic effect

Include in Subsection 2.6

Provide evidence (if relevant) that using the test results in better targeting of patients that are likely to respond most to the medicine (ie by using the prognostic effect of the biomarker to determine the baseline risk of disease or condition progression).

For example, is there evidence of minimal variation in a measure of relative treatment effect between the proposed medicine and comparator/usual care trial arms, but determining biomarker status helps identify patients at greatest risk of an event, which, in turn, helps maximise the absolute treatment effect?

Amalgamate with Item 19 if this issue has been addressed there.

If an improvement in treatment effect is a result of better targeting of those patients that are likely to respond most, this identifies a relationship between the biomarker and a potentially broader range of existing and future treatment options (potentially including nonmedicine treatment options) than is likely to apply for treatment effect modification. This may allow reimbursement of either the test or the medicine of both technologies.

This apparent improvement in treatment effect is simply because a certain patient subgroup (flagged by a specific biomarker) will always do better, so the biomarker is considered prognostic.

It is possible for both treatment effect modification and prognostic effect to coexist. In this case, to assess the unique contribution of the medicine, an assessment of its effect must be made relative to usual care and an adjustment made for the background prognostic effect of the biomarker.

34 (O) Size of the treatment effect on patient-relevant health outcomes

Include in Subsections 2.6 and 2.8

Indicate whether the effect of the medicine, as conditioned by the test or biomarker result, has a clinically important and statistically significant effect on patient-relevant health outcomes (both safety and effectiveness). Relate this to the following factors:

- factors intrinsic to the proposed medicine
 - treatment effect modification when prognostic effect is not present in the medicine/biomarker relationship (see Item 32)
 - absolute treatment effect when prognostic effect is present in the medicine/biomarker relationship (see Item 33)
- the factor intrinsic to the proposed test
 - accuracy of identification of biomarker status given the test result (ie positive predictive value and negative predictive value), and the impact of inappropriately treating or not treating patients who received an inaccurate biomarker test result.

When the proposed MBS listing either cannot include the test used in the evidence base or also encompasses other test options, delineate the consequences of using the other test options in place of the evidentiary standard test for health outcomes and the provision of subsequent health care resources in Subsection 2.7.

Applicability of the effectiveness of the codependent technology

ADDITIONAL INFORMATION REQUESTS

- 35 (O) Indicate whether the evidence supporting the clinical effectiveness of the codependent technology is applicable to the Australian population and to the circumstances of using each of the technologies

35 (O) Applicability of the evidence

Include in Subsection 2.7, with any economic implications included in Subsection 3A.3

Indicate whether the evidence supporting the clinical effectiveness of the codependent technology is applicable to the Australian population and to the circumstances of using each of the technologies. For example, is the biomarker prevalence in the trial similar to that in the target MBS population? Is the medicine, dosage and frequency of use in the trial similar to that proposed for the target PBS population? How are any inconsistencies identified in the submission addressed?

Section 3 – Economic evaluation

The following section contains information requests for establishing the cost-effectiveness of the codependent technologies in terms of patient health outcomes.

Structure of the model

ADDITIONAL INFORMATION REQUESTS

- 36 (O) Indicate whether the model structure is consistent with other published economic evaluations in the same broad clinical management setting, initiating before the decision to test or treat
- 37 (O) Indicate whether the model structure is consistent with the clinical pathways provided in response to Item 7
- 38 (O) If relevant, provide a supplementary analysis of the nonhealth-related impacts associated with using the proposed test

36 (O) Consistency with other published economic evaluations

Include in Subsection 3A.2

Indicate whether the model structure is consistent with other published economic evaluations in the same broad clinical management setting, initiating before the decision to test or treat.

Indicate whether and why there are differences in model structure compared with the identified economic evaluations.

37 (O) Consistency with the clinical management pathways

Include in Subsection 3A.2

Indicate whether the model structure is consistent with the clinical pathways provided in response to Item 7. Indicate whether and why there are differences between the model structure and the clinical pathways, considering the following factors:

- The start point is testing of the eligible population (ie only a subset of the tested population goes forward to receive the proposed medicine). The less-preferred alternative is to start with the treatment and back-calculate the number (and costs) of testing the larger population.
- Where the model is constructed using a linked-evidence approach, include model arms to account for both accurate and inaccurate test results (see Items 39–42). This is not necessary if a single-randomised trial of the test is available (ie randomised to test versus no test trial arms) and only the evidentiary standard test is to be listed in the MBS – then the impact of inaccurate testing is incorporated in the health outcomes of the patients (this is analogous to a trial-based economic evaluation of the test and medicine pair). Where true positive, false positive, true negative and false negative test results are accounted for in the model, present a table specifying what source of estimates is used for each of the health outcomes and the health care resource provision in each of these four situations.
- A scenario analysis is provided where the proposed medicine is used without testing to show the extent of improvement in the incremental cost-effectiveness ratio (ICER) associated with using the test (see Items 7 and 58).

38 (O) Nonhealth-related impacts

Include in Subsection 3A.2

If relevant, provide a supplementary analysis of the nonhealth-related impacts associated with using the proposed test.

The same considerations for caregiver impact apply to codependent technologies as for other technologies, so the guidance provided in Part A of these guidelines will apply. The base-case economic model should be from a health system perspective. If other significant nonhealth impacts are expected, provide a supplementary analysis from a societal perspective. Discuss this in a supplementary analysis section. This could include the value to patients of being informed of their biomarker status.

Transition probabilities relating to test outcomes

ADDITIONAL INFORMATION REQUESTS

- If a linked-evidence approach was used in Section 2, calculate and include in the model:
 - 39 (O) the positive predictive value (PPV) of the proposed test
 - 40 (O) the complement of the PPV of the proposed test
 - 41 (O) the negative predictive value (NPV) of the proposed test
 - 42 (O) the complement of the NPV of the proposed test
- 43 (O) In the model, provide the incidence of adverse events associated with (i) the proposed medicine in patients with correct (true positive) and incorrect (false positive) positive test results, and (ii) the comparator medicine in patients with correct (true negative) and incorrect (false negative) negative test results; or (iii) reported from the direct evidence (ie in the circumstance that a double or single-randomised controlled trial of the test is available – analogous to a trial-based economic evaluation of the test/medicine pair)
- 44 (O) In the model, include the incidence of test-related adverse events for *all those tested*
- 45 (O) Where prognostic effect is operating in addition to treatment effect modification, ensure that the model adjusts for this factor when presenting absolute treatment effects

Calculate the following values for inclusion in the model using prevalence of the biomarker in the ‘tested’ population, and the sensitivity and specificity of the proposed test reported in Section 2:

- positive predictive value (PPV)
- negative predictive value (NPV)
- complement of PPV (1-PPV)
- complement of NPV (1-NPV).

39 (O) Positive predictive value of the proposed test

Include in Subsections 3A.4 and 3A.8

Calculating the PPV requires information on the sensitivity and specificity of the proposed test – as reported in the clinical evaluation section of the submission – and the prevalence (probability) of the biomarker in the target MBS population. It is the probability that a test positive result for the biomarker is correct. The PPV is used in a Bayesian manner to condition the model and calculate the transition probability associated with a true positive (use in Subsection 3A.4).

$$PPV = \frac{SN \times P}{SN \times P + (1 - SP) \times (1 - P)}$$

where SN = sensitivity, P = prevalence of the biomarker, SP = specificity

If agreement or concordance data are provided, rather than test accuracy data, measures such as the PPV cannot be accurately calculated since the subjects' condition (as determined by a valid reference standard) is unknown. In this situation, a range of indicative PPVs (using a test nominated as the reference standard) might be used as transition probabilities and tested in sensitivity analyses. These analyses would explore the impact on the ICER of discrepancies in the agreement between the evidentiary standard test and other nominated reference standard tests that will be used in Australia to identify the biomarker.

40 (O) Complement of positive predictive value of the proposed test

Include in Subsections 3A.4 and 3A.8

One minus positive predictive value (1 – PPV) is the probability that a test positive result for the biomarker is incorrect (false positive). It predicts the consequence that patients will be treated unnecessarily, with a consequent decrement in expected treatment effectiveness and increment in harms. It is used in a Bayesian manner to condition the model and calculate transition probabilities.

If agreement or concordance data are provided rather than test accuracy data, present the complement of the range of indicative PPVs used to address Item 39.

41 (O) Negative predictive value of the proposed test

Include in Subsection 3A.4 and 3A.8

To calculate the NPV also requires information on the sensitivity and specificity of the proposed test – as reported in the clinical evaluation section of the submission – and the prevalence (probability) of the biomarker (eg phenotypic expression of mutation) in the target MBS population.

The NPV is the probability that a test negative result for the biomarker is correct. It is used in a Bayesian manner to condition the model and calculate transition probabilities.

$$NPV = \frac{SP \times (1 - P)}{(1 - SN) \times P + SP \times (1 - P)}$$

where SN = sensitivity, P = prevalence of the biomarker, SP = specificity

If agreement or concordance data are provided rather than test accuracy data, refer to guidance provided at Item 39.

42 (O) Complement of negative predictive value of the proposed test

Include in Subsection 3A.4 and 3A.8

One minus negative predictive value (1 – NPV) is the probability that a test negative is incorrect (false negative) and predicts the scenario where patients receive usual care instead of the proposed medicine with a consequent decrement in expected treatment

effectiveness. It is used in a Bayesian manner to condition the model and calculate transition probabilities.

If agreement or concordance data are provided rather than test accuracy data, present the complement of the range of indicative NPVs used to address Item 41.

43 (O) Medicine-related adverse events in patients according to test result

Include in Subsection 2.5 or 2.6, and Section 3

In the model, provide the incidence of adverse events associated with (i) the proposed medicine in patients with correct (true positive) and incorrect (false positive) positive test results, and (ii) the comparator medicine in patients with correct (true negative) and incorrect (false negative) negative test results; or (iii) reported from the direct evidence (ie in the circumstance that a direct randomised trial of the test is available – analogous to a trial-based economic evaluation of the test–medicine pair).

Determine whether biomarker test status predicts or does not predict any comparative treatment effect variation in terms of adverse events (Subsection 2.5 or 2.6) and incorporate in the model (eg Subsections 3A.2 and 3A.4). Include the impact of medicine-related adverse events on patients with a positive test result.

44 (O) Incidence of test-related adverse events

Include in Subsections 3A.4 and 3A.6

In the model, include the incidence of test-related adverse events for **all those tested**. Refer to Items 16 and 29 (Subsection 3A.4). This includes adverse events from resampling to perform or reperform the test. Sometimes the original sample is not available or not of sufficient size to allow retesting, and a new sample is needed to reperform the test. Account for the costs associated with resampling in the model (Subsection 3A.6).

45 (O) Incorporation of net treatment effects (if relevant)

Include in Subsections 3A.2–3A.5

Where prognostic effect is operating in addition to treatment effect modification, ensure that the model adjusts for this factor when presenting absolute treatment effects.

Resource items and costs included in the model

ADDITIONAL INFORMATION REQUESTS

- Include the following costs in the model:
- 46 (O) unit test costs
 - 47 (O) cost of sampling (if relevant)
 - 48 (O) test administration costs
 - 49 (O) costs of patient consultations with medical personnel regarding the test results and treatment planning
 - 50 (O) costs of retesting and nonassessable results
 - 51 (O) costs for adverse events associated with testing
 - 52 (O) costs of additional and further testing as a result of the proposed test
 - 53 (O) costs of medicine-related adverse events, including those where the test result was false positive
 - 54 (O) costs of other relevant health care resources (eg diagnostic, medical, hospital, allied health)

46 (O) Unit test costs

Include in Subsection 3A.6

In estimating the cost of testing, include the cost of tests undertaken on all patients for whom the medicine is being considered, not just the cost of the test for those who were found to be suitable for the medicine. Include all relevant sources of costs (eg infrastructure, training, quality assurance) that need to be captured in, and associated with, rendering an MBS-funded test (eg a pathology test).

47 (O) Cost of sampling (if relevant)

Include in Subsection 3A.6

For example, taking, storing, retrieving and transporting biopsy samples.

48 (O) Other relevant costs of test administration

Include in Subsection 3A.6

49 (O) Costs for patient consultations with medical personnel

Include in Subsection 3A.6

Include costs of patient consultations with medical personnel regarding the test results and treatment planning. Include an explanation as to the extent that these costs overlap with the already-occurring consultations for medical management.

50 (O) Costs of retesting and nonassessable results

Include in Subsection 3A.6

This could be covered at Item 46. In some cases, the test result is invalid or not assessable, and retesting of the sample is required. Ensure that any costs associated with retesting are in the model.

51 (O) Costs for adverse events associated with testing

Include in Subsection 3A.6

Provide costs for the items mentioned at Item 44.

52 (O) Costs of additional and further testing as a result of the proposed test

Include in Subsection 3A.6

This includes costs associated with any changes in subsequent types of testing for other purposes brought about by the use of the proposed test.

53 (O) Cost of medicine-related adverse events

Include in Subsection 3A.6

Provide costs for the items mentioned at Item 43. Include these costs in all arms of the model, including false positive test result arms.

54 (O) Costs of other relevant health care resources

Include in Subsection 3A.6.

For example, costs for diagnostic, medical, hospital and allied health resources.

Uncertainties in the model

ADDITIONAL INFORMATION REQUESTS

- 55 (O) Assess the uncertainty around the medicine's therapeutic effectiveness
- 56 (O) If a linked-evidence approach was used in Section 2, assess the uncertainty around test accuracy
- 57 (O) If a linked-evidence approach was used in Section 2, assess the uncertainty around the prevalence of the biomarker
- 58 (O) If relevant, provide a scenario analysis for the option of PBS listing the medicine without the proposed test as a prerequisite

55 (O) Uncertainty around therapeutic effectiveness

Include in Subsection 3A.9

In instances where both treatment effect modification and prognostic effect are operating in the medicine-biomarker relationship, assess the uncertainty of the estimated incremental treatment effect and model this uncertainty.

56 (O) Uncertainty around test accuracy (if relevant)

Include in Subsection 3A.9

If a linked-evidence approach was used in Section 2, assess the uncertainty around test accuracy. In instances where there is heterogeneity in plausible test accuracy measures (sensitivity and specificity) in the collated evidence base, particularly for different eligible test options, vary these measures when calculating the PPV and NPV transition probabilities and assess the impact of this uncertainty on the estimated absolute treatment effect.

57 (O) Uncertainty around biomarker prevalence (if relevant)

Include in Subsection 3A.9

In instances where there is limited or heterogeneous information on the prevalence of the biomarker in the target MBS population, vary the plausible prevalence rate when calculating the PPV and NPV transition probabilities, and assess the impact of this uncertainty on the estimated absolute treatment effect.

58 (O) PBS listing the medicine without the biomarker test as a prerequisite

Include in Subsection 3A.9

Depending on the prevalence of the biomarker, in some cases there may be a net clinical benefit – which may be more cost-effective – to provide the medicine to patients without the use of biomarker testing. A scenario analysis should be used to make this explicit (see Item 7).

Section 4 – Use of the medicine in practice

ADDITIONAL INFORMATION REQUESTS

- 59 (O) Present a budget impact analysis incorporating both MBS and PBS components, with results split by sector (public, private, patient, other)
- 60 (O) Calculate an epidemiologic estimate for disease burden that is based on the prevalence of the biomarker as determined by the proposed test
- 61 (O) Estimate the cost of testing all patients eligible for the test and the cost of retesting when indeterminate or nonassessable results are produced. If testing is also required after therapy is initiated (ie to monitor therapy or to determine when therapy should cease), these costs should also be included. If relevant, provide a scenario analysis for the option of PBS listing the medicine without the proposed test as a prerequisite
- 62 (O) Estimate any other MBS costs that would be incurred if the test and medicine were listed

This section contains information requests for establishing the predicted use of codependent technologies and the financial implications to the Australian Government budget.

59 (O) Budget impact analysis incorporating both MBS and PBS components

Include in Subsections 4.1–4.6

Present a budget impact analysis incorporating both MBS and PBS components, with results split by sector (public, private, patient, other).

Present the cost of the proposed test alongside the proposed medicine in Subsection 4.2, if appropriate, and use Subsection 4.5 to present utilisation of, and costs associated with, other MBS items.

60 (O) Epidemiology estimate for disease burden

Include in Subsection 4.2

Calculate an epidemiologic estimate for disease burden that is based on the prevalence of the biomarker as determined by the proposed test.

A market-share estimate for a new biomarker scenario is likely to be inappropriate, because previous medicine utilisation will not have been targeted to this biomarker. Seek expert epidemiological advice on whether prevalence is expected to remain constant after listing.

First, estimate the number(s) of patients likely to be considered for the test (eg with the medical condition as defined). Second, based on the prevalence of the biomarker, estimate the proportion of patients likely to receive a positive test result with the proposed test (and be eligible for use of the medicine).

Where the biomarker has been validated using another test and is targeted by other reimbursed medicines, a market-share approach may be reasonable.

61 (T) Likely use and overall financial cost of the test

Include in Subsections 4.2 and 4.6

Estimate the cost of testing all patients eligible for the test (ie biomarker positive, biomarker negative and indeterminate biomarker status) and the cost of retesting when indeterminate or nonassessable results are produced. Include these costs if testing is also required after therapy is initiated (ie to monitor therapy or to determine when therapy should cease).

Any uncertainty about use of the test (ie biomarker prevalence) or changing availability of the test should be explored in Subsection 4.6. There may be 'leakage' issues identified through an assessment of the 'change in management' part of the linked evidence. A codependent technology is meant to target the use of a medicine to a patient who is biomarker positive, but, in some cases, the medicine is given even if the patient is negative for the biomarker. Similarly, a test may be done to rule out use of a medicine (eg to avoid potential medicine-related adverse events or the development of resistance), but the medicine is given anyway.

62 (O) Other MBS costs

Include in Subsection 4.5

Estimate any other MBS costs that would be incurred if the test and medicine were listed. Consider procedures for administration of the medicine and consultations for adverse events, consultations for resampling, genetic counselling and so on.

Appendix 8 Screening investigative medical services

Those investigative medical services (in isolation or in combination) that attempt to initiate a process leading to early detection of a target condition with a recognisable latent (asymptomatic) stage in the expectation of benefit which is supposedly offered by earlier detection are known as screening investigative medical services. There is no difference between the concept of screening, periodic health exams, case finding and ‘triage’ testing as they are all attempts at initiating early detection in the expectation of benefit that is supposedly offered by early detection. There are different subtypes of screening, each with their specific aims:

- mass screening aims to screen a whole population (or subset);
- multiple or multiphasic screening uses several screening tests at the same time.;
- case finding or opportunistic screening is aimed at patients who consult a health care practitioner for some other purpose; and
- targeted screening of groups with specific exposures (e.g. workers in lead battery factories) is often used in environmental and occupational health. For example, when targeted screening is done in groups with occupational exposures, the criteria for screening are not necessarily as strict as those for general population screening. The health effect that is prevented may be minor (nausea or vomiting), but screening may be a high priority if the effect reduces the patients’ ability to work. Many health effects from exposure to environmental hazards are graded and preventing a minor effect may also prevent more serious effects. Targeted screening can be legally required (e.g. in miners who work with lead or chromium) and used in follow-up to environmental health incidents.

Screening investigative medical services should not be confused with diagnostic medical services, which are those investigative medical services (in isolation or in combination) that tend to be applied to symptomatic individuals to elucidate information that explains and/or assists in managing their current clinical presentation. Diagnostic investigative medical services can be further classified into two broad subgroups:

- 1) those rendered in individuals in whom the underlying reason of their clinical presentation is not yet clear and the purpose of testing is to confirm a diagnosis that might explain and/or add value in managing that individual’s clinical presentation; and
- 2) those that generate information about an individual in whom it supposedly adds value to managing an existing diagnosis (i.e. staging, disease progress).

Deciding whether an investigative medical service should be incorporated as part of a population-based screening program is not simply a decision based on epidemiological evidence. The effectiveness of population-based screening depends on both the accuracy of the screening investigative medical service and the clinical effectiveness of early detection and intervention. A good screening investigative medical service must detect the target condition earlier than without screening, and with sufficient accuracy to avoid producing large numbers of false positives and false negative results. Screening and treating those who test positive should also improve the likelihood of favourable health

outcomes. The World Health Organisation has developed a set of principles¹⁷ that should be used to guide the decision-making process involved with the planning, operation and evaluation, and prioritising of screening programs are centred on these principles.

The principles for developing a screening program are as follows:

The Disease

- Must be an important public health problem.
- Natural history of the disease must be known.
- Must be a long latent or early asymptomatic phase.
- Clear and agreed population on who to target screening (and clear communication strategies to reach this population).

The Test

- An appropriate test must be available – acceptable, adequate sensitivity and specificity (this includes appropriate accreditation and quality assurance program, appropriate training/facilities/logistics to deliver test and clear support & recall services to facilitate access to subsequent services for patients who test positive).
- Clearly defined screening interval between tests that is backed up by evidence.

The Treatment

- An effective treatment must be available and benefits should outweigh harms.
- Treatment started early must be better than treatment started late.
- There must be an agreed policy on who to treat.
- Must have accessible follow up diagnostic services and treatment (adequate infrastructure in terms of clear referral pathways, timely access to treatment, appropriate training of staff delivering the service and adequate workforce).

The Program

- The program should be ongoing and sustainable.
- must be cost effective (benefits outweigh costs) - The costs should be appropriate within the costs of the wider health care program.
- Appropriate model of care to deliver program as well as supporting health system and governance requirements (funding/health information systems such as data collection/workforce/facilities etc).
- Case finding, not a one off event.

¹⁷ Wilson JMG and Jungner G (1968). *Principles and Practice of Screening for Disease*, Public Health Paper 34, World Health Organization, Geneva. Further information on the WHO screening principles can be found at: [http://www.health.gov.au/internet/screening/publishing.nsf/Content/pop-based-screening-framework/\\$File/screening-framework.pdf](http://www.health.gov.au/internet/screening/publishing.nsf/Content/pop-based-screening-framework/$File/screening-framework.pdf)

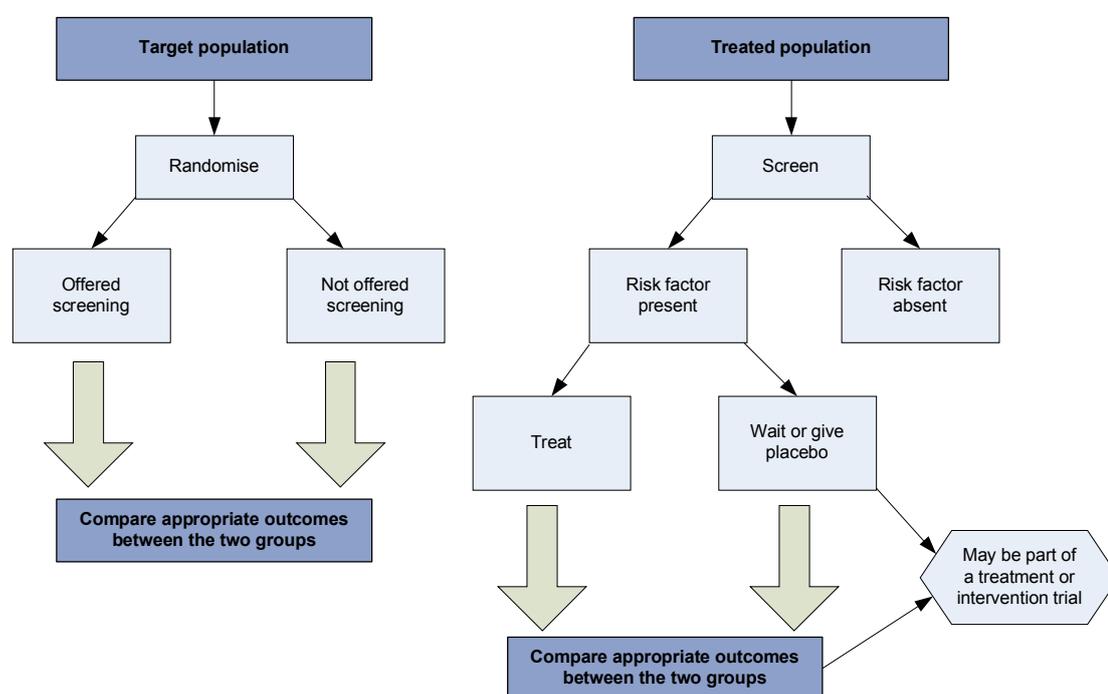
From the above discussion, balancing benefits and harms of a population-based screening investigative medical service depends on:

- how much inconsequential disease is detected;
- how large is the benefit for the true positives;
- how accurate is the investigative medical service (how many false positives and false negatives);
- baselines risk of disease;
- the screening interval; and
- whether there are standards that apply to the facilities and systems for follow-up and treatment.

The value of using an investigative medical service as part of a screening program becomes questionable if the use of a test in a screening program results in the detection of high numbers of insignificant lesions, the overtreatment of insignificant lesions, false positive tests leading to the unnecessary treatment to exclude cancer, a serious burden of diagnosis or if it requires a significant number of people to be involved to benefit only a few.

A key question that needs to be asked is whether there is direct evidence from randomised trials to show that that early detection and treatment works through an investigative medical service that is being used as part of a screening program. In general, there are two types of randomised trials that can be used to evaluate screening investigative medical services, as shown in Figure A8.1.

Figure A8.1 The two types of randomised trials used to evaluate screening investigative medical services



At times, early detection can offer no value, despite appearing beneficial for the following three reasons:

- **Lead time bias** — the interval between a diagnosis of a condition at screening, compared to when the condition would have been diagnosed due to the development of symptoms. In other words, lead time bias is the time gained in treatment and controlling the condition when detected earlier than usual. Studies assessing screening are vulnerable to lead time bias, because survival as measured from the time of diagnosis may be increased — not because patients live longer, but because screening lengthens the time they know they have the condition. To avoid lead time bias, the outcome should be the mortality rate (or other relevant time rate) rather than survival time after diagnosis. Lead time bias can also be avoided by using randomised trials of effects of screening on mortality or morbidity using an intention-to-screen analysis;
- **Length time bias** — the tendency of screening to preferentially detect a slowly progressing condition. That is, it is a systematic error that occurs when disproportionate numbers of long-duration cases are found in one group and not the other. Length time bias occurs because those with a long preclinical phase are more readily detected by screening than those with more rapidly progressing conditions. To avoid length time bias, use randomised trials with an intention-to-screen analysis; and
- **Selection bias** — occurs when people who volunteer for screening (or accept invitations to screening) are healthier and have lower death rates than people who decline. Selection bias is sometimes called health worker effect bias, because health workers are commonly selected for such studies (although not all health worker samples would necessarily be healthier than people selected from the general population). To avoid selection bias, use randomised trials with an intention-to-screen analysis.