

Technical Guidelines for preparing assessment reports for the Medical Services Advisory Committee – Service Type: Investigative

(Version 1.3)

Draft for public consultation 2015

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).

Like the Technical Guidelines for Therapeutic services, it is intended that this current draft be used by applicants until an agreed version is made available following the receipt of feedback.

Chair  
Medical Services Advisory Committee

Contents

Foreword ii

Contacts xii

Medical Services Advisory Committee Secretariat xii

Record of updates xiii

Abbreviations xiv

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

2.5 Writing and style conventions used in these Guidelines 6

2.6 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 13

4.3.1 Standard assessment report 13

4.3.2 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

5.2.1 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

Introduction 18

A.2 Proposed medical service 18

A.3 Proposed MBS listing or other public funding sought 19

A.4 Comparator details 19

A.5 Clinical management algorithm(s) 19

A.5.1 Algorithms for intended and current contexts 19

A.6 Differences between the proposed medical service and the main comparator 20

A.7 Clinical claim 20

A.8 Summarise the primary elements of the decision analysis (PPICO) 21

Section B Clinical evaluation for the proposed investigative medical service (analytical validity) 22

Introduction 22

Analytical validity 22

Evidence of health outcomes 22

B.1 Identification of a reference standard 23

Reference standard 23

B.2 Description of search strategies (accuracy studies) 25

Search strategies to identify accuracy studies 26

B.3 Listing of all accuracy studies 27

B.3a Listing accuracy studies of direct comparison 27

B.3b Listing accuracy studies for inclusion in an indirect comparison (where relevant) 31

B.4 Assessment of the measures taken by investigators to minimise bias in accuracy studies 34

Assessment of measures to minimise bias 34

Notes for quality checklist 35

Tabulate responses 37

Source documents 38

B.5 Characteristics of accuracy studies 38

Details of accuracy studies 38

Characteristics of the accuracy studies included in an indirect comparison (where relevant) 39

Source documents 40

B.6 Systematic overview of the results of the accuracy studies 40

Primary analysis 40

Analysis (including meta-analysis) 41

Identifying and interpreting heterogeneity between accuracy studies 42

Pooling data 43

Pooling sensitivity and specificity 44

Pooling likelihood ratios 44

Producing a summary DOR and ROC curve 44

Identifying a test threshold effect 44

Presenting the results of an indirect comparison 45

Adverse event data 45

Clinical importance 46

B.7(a) Extended assessment of reliability evidence 47

Primary analysis of studies reporting reliability/reproducibility 48

Consideration of effective analytical validity 48

B.7(b) Extended assessment of clinical validity 49

B.7(c) Extended assessment of comparative harms 49

B.8 Interpretation of the investigative evidence 50

Section C Translation Issues 53

Introduction 53

C.1Description of search strategies 55

Search strategies 55

C.2 Presentation of clinical trial evidence of therapeutic health outcomes 56

Clinical trial evidence of therapeutic health outcomes 56

Prognostic information 57

Co-dependency 57

Section C (CoD) *(Co-dependents)* Translation Issues 58

Introduction 58

C.1(CoD) ‘Overlap’ information 59

C.2 (CoD) Time course of the test results 63

Section D Economic evaluation for the main indication 64

Introduction 64

D.1 Overview of the economic evaluation 66

Generation of the base-case economic evaluation 66

Type of economic evaluation 67

Sources of information 72

D.2 Population and circumstances of use reflected in the economic evaluation 73

Population (demographic and patient characteristics) 73

Circumstances of use 73

Consistency across characteristics 74

D.3 Structure and rationale of the economic evaluation 78

Literature review 78

Software package 79

Fully accessible electronic copy of the economic evaluation 80

Structure of the economic evaluation 80

Justification of the structure 84

Time horizon and outcomes used in the evaluation 85

Methods used to generate the results 85

Sources of information 86

D.4 Variables in the economic evaluation 87

Variables used in the evaluation 87

Direct health care resources 88

Present value of direct health care resource costs 89

Present value of health outcomes 91

Additional considerations relating to necessary diagnostic criteria 92

D.5 Results of the economic evaluation 94

Health care resource costs 94

Stepped economic evaluation (requested if the evaluation includes variables derived from Section C) 99

D.6 Sensitivity analyses 102

Univariate sensitivity analyses 102

Multivariate sensitivity analyses 103

Sensitivity of the results to changes in the modelled economic evaluation 103

Section E Estimated extent of use and financial implications 104

Introduction 104

E.1 Justification of the selection of sources of data 107

Published data sources 107

Spreadsheet 1 (‘Background and assumptions’) 108

Copies of data 109

E.2 Estimation of use and costs of the proposed medical service 109

Numbers of patients 109

Spreadsheet 2 – Epidemiology of the disease and patient numbers’ 110

Number of times the proposed medical service is delivered 111

Aggregated cost calculations 111

Spreadsheet 3 – Cost of the proposed medical service to the MBS 111

E.3 Estimation of changes in use and cost of other medical services 112

Medical services likely to be affected by the listing of the proposed medical service 112

Number of times the proposed medical service is delivered 113

Costs over five years 113

Aggregated cost calculations 113

Spreadsheet 4 – Cost implications to the MBS from substitutions and other increases and decreases 113

E.4 Estimated financial implications for the MBS 114

Spreadsheet 5 – Net cost of the proposed medical service to the MBS 114

E.5 Identification, estimation and reduction of uncertainty 114

Nature of uncertainty 114

Sources of uncertainty 115

Estimating and reducing uncertainty 116

Summary of calculations 117

Section F Options to present additional relevant information 118

Introduction 118

F.1 Other relevant factors 118

Equity principles 118

Equity assumptions 118

Guidance on the ‘rule of rescue’ 119

Discuss any other relevant factor 120

Part III Alternative clinical evidence for proposed medical services to be considered by MSAC 121

Section C(i) Linking an investigative conclusion of superiority to the economic evaluation when there are no clinical trials of therapeutic health outcomes 123

Introduction 123

C(i).1Basis for linked evidence 125

C(i).2 Evidence for a change in clinical management 126

C(i).3 Evidence for change in health outcomes 127

Section D(i) Economic evaluation for the main indication: presenting a cost-minimisation approach 128

Cost-minimisation analysis 128

Cost analysis 128

D(i).1 Presentation of a cost-minimisation analysis or a cost analysis 129

Cost consequences related to the provision of resources 130

Appendices 131

Appendix 1 Relevant factors influencing provision of advice by MSAC 132

Appendix 2 Expert opinion 134

Appendix 3 Assessment of non-inferiority 138

Appendix 4 Utility valuation of health outcomes 141

Appendix 5 Monetary valuation of health outcomes 156

Appendix 6 Including non-health care resources and non-health outcomes in a supplementary analysis 162

Appendix 7 Measures taken by the investigators to minimise bias in non-randomised studies 166

Appendix 8 Screening investigative medical services 172

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

Figure B.1 Key information requests for assessment report Section B of a standard assessment for MSAC 23

Table B.3.1 Summary of identification of accuracy studies of direct comparison from the search of the published literature (see Section B.3) 28

Table B.3.2 Accuracy studies of direct comparison (and associated reports) presented in the assessment report 29

Table B.3.3 Reasons to exclude each accuracy study of direct comparison from further detailed assessment 30

Table B.3.4 Comparative summary of characteristics of accuracy studies of direct comparison 30

Table B.3.5 Comparative summary of results of accuracy studies of direct comparison 30

Table B.3.5 Accuracy studies (and associated reports) presented in the assessment report 32

Table B.3.6 Reasons to exclude each accuracy study from the indirect comparison 33

Box B.4.1 QUADAS tool checklist for assessing the quality of specific accuracy studies 35

Table B.4.1 Summary of the measures undertaken to minimise bias 37

Table B.4.2 Flow of participants in each accuracy study of direct comparison under consideration 38

Table B.5.1 Eligibility criteria in the accuracy studies under consideration 39

Table B.5.2 Characteristics of participants in the accuracy studies 39

Table B.6.1 Results of accuracy studies directly comparing proposed and comparator investigative medical services 45

Table B.6.2 Summary of results of the indirect comparison 45

Table B.7(a).1 Characteristics and summary results of accuracy studies reporting reliability/reproducibility 48

Table B.8.1 Classification of the investigative relativity of the proposed investigative medical service over its main comparator 51

Figure C.1 Pathways for assessment report Section C of a standard assessment for MSAC 54

Table C.1.1(CoD) ‘Overlap’ questions to answer when submitting a co-dependent assessment report to MSAC 60

Figure D.1 Key information requests for assessment report Section D of a standard assessment for MSAC 65

Table D.1.1 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 68

Table D.1.2 Type of economic evaluation that should be presented for various classifications of a service under MSAC consideration 70

Table D.2.1 Definitions for populations and circumstance of use that should be taken into account in the evaluation 73

Table D.2.2 Comparison of characteristics of trial and requested populations and circumstances of use 75

Table D.2.3 Example of a comparison of the characteristics of target, study and wider populations and settings 77

Figure D.3.1 Example of the structure of a basic economic evaluation 81

Table D.3.1 Key elements of the base-case economic evaluation 82

Table D.4.1 List of health care resource items and unit costs included in the economic evaluation 90

Table D.5.1 List of health care resource items and summary of cost impacts in the economic evaluation 95

Table D.5.2 List of health care resource items and summary of cost impacts for each health state in a state transition model 96

Table D.5.3 List of health states and summary of cost impacts included in the economic evaluation 96

Table D.5.4 List of health states and summary of health outcomes included in the economic evaluation 97

Table D.5.5 Assessment of the implications for the economic evaluation of applying the clinical evaluation (Step 1 then Step 2) 100

Table D.5.6 Assessment of the implications for the economic evaluation of extrapolating and transforming the clinical evaluation (Step 3) 101

Table D.6.1 Analyses of the implications for the economic evaluation of usage beyond the requested populations and circumstances of use 103

Figure E.1 Key information requests for assessment report Section E of a standard assessment for MSAC 105

Table E.1.1 Categories of data sources 108

Figure C.1(i) Key information requests for assessment report Section C of a standard assessment report for MSAC where there is an investigative conclusion of ‘superior’ from assessment report Section B but no clinical trials of therapeutic effectiveness 124

Figure D(i).1 Key information requests for assessment report Section D of a standard assessment for MSAC in which the therapeutic conclusion from assessment report Section B is non-inferior 129

Table A1.1 Factors that are more readily quantified 132

Table A1.2 Examples of factors that are less readily quantified 133

Table A2.1 Methods to collect and collate expert opinion 136

Figure A8.1 The two types of randomised trials used to evaluate screening investigative medical services 174

# Contacts

## Medical Services Advisory Committee Secretariat

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 proposed assessment reports or related matters. Any correspondence or assessment reports should also be lodged at via the address below. They 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](http://www.msac.gov.au/)

Email: [hta@heath.gov.au](mailto: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 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# 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 |
| AV | Analytical validity |
| CV | Contingent valuation |
| DOR | diagnostic odds ratio |
| DRG | diagnostic-related group |
| EAV | Effective analytical validity |
| ESC | Evaluation Sub-committee |
| Guidelines | *Guidelines for Preparing Investigative Assessment Reports for the Medical Services Advisory Committee* |
| Health | Australian Government Department of Health |
| HTA | health technology assessment |
| HTA Review | *Review of Health Technology Assessment in Australia* |
| ID | Identification |
| ITT | Intention-to-treat |
| MAUI | Multi- attribute utility instrument |
| MBS | Medicare Benefits Schedule |
| MESP | MSAC Expert Standing Panel |
| MCID | Minimal clinical important difference |
| MSAC | Medical Services Advisory Committee |
| NHMRC | National Health and Medical Research Council |
| PASC | Protocol Advisory Sub-committee |
| PBAC | Pharmaceutical Benefits Advisory Committee |
| PBAC Guidelines | *Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee* |
| 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 | Standard Reporting of Diagnostic Accuracy |
| TGA | Therapeutic Goods Administration |
| Therapeutic Guidelines | *Guidelines for Preparing Therapeutic Assessment Reports for the Medical Services Advisory Committee* |
| TTO | Time trade-off |
| QALYs | Quality adjusted life years |
| Q-Twist | Quality- adjusted time without symptoms of the disease or toxicity |
| QUADAS | Quality Assessment of Diagnostic Accuracy Studies |
| WTP | 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. 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](http://www.msac.gov.au/).

## 1.3 MSAC sub-committees

MSAC currently has two sub-committees: the Protocol 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) 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 MSAC’s assessment of an application are 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

Unless commercial-in-confidence information has been identified, the submitted evidence will be made public. Where agreed confidentiality has been reached, applicants will have access to and confirm a modified version of the assessment report (submitted or contracted), and will be able to view the MSAC Public Summary Document (PSD) for agreement to modifications before uploading on the MSAC website. Reports will be published as submitted to ESC. Any agreed errors of fact will be provided separately as an errata.

### 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.

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 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.

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.

Please note that 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

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 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);
* 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; 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) andwhere 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 4.3.2 of these Guidelines.

## 2.5 Writing and style conventions used in these Guidelines

Several conventions have informed the revision of these Guidelinesto 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**’ in boxes. 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 box 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.

## 2.6 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 (see page xii).

# 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 flexibility 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

The information should address the Final Protocol agreed by PASC, however, an applicant may choose to take the risk of not to adhere to the Final Protocol noting 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 Protocol.

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.

Figure 4.1 Sections of a standard assessment report

**Section A**

**CONTEXT**

**Section C**

**TRANSLATION ISSUES**

**Section B**

**CLINICAL EVALUATION**

**Section D**

**FINANCIAL EVALUATION**

**Section E**

**BUDGETARY IMPLICATIONS**

**Section F**

**OTHER**

(optional)

Assessment report Section

**Executive summary**

(required for all reports)

Content

Clearly set out the key aspects and issues presented in the main body of the assessment report.

Identify the type of funding being sought. If the application seeks MBS funding, include the proposed item(s) descriptors and fee.

Establish the context for the assessment report. Describe the proposed medical service, its intended use on the MBS or elsewhere, and the medical services that would be co-delivered or substituted (the medical service likely to be most replaced by health care providers in practice is the ‘main comparator’). Describe the agreed Final Protocol using the PICO approach.

Provide the best available evidence comparing the clinical performance (**therapeutic medical service**) or analytical validity (**investigative medical service**) of the proposed medical service with the main comparator. Provide details about the trials or studies and other sources of evidence, including the scientific rigour of the methods, the size, statistical precision, clinical importance and patient relevance of the results. Conclude with a comparative assessment of the proposed medical service.

**Therapeutic medical service** - describe the methods used in the pre-modelling studies to translate (apply, extrapolate and transform) the results of the evaluation of the clinical studies to the context of the requested listing. Include a description of the analytical plan and research questions, the data used (with reasons for exclusions) and analyses. Provide a table with the results of the analyses (i.e. the variables for use in any modelled economic evaluation).

**Investigative medical services** - include evidence for indirect therapeutic health outcomes and pre-modelling studies as for therapeutic medical services.

Provide an economic evaluation that focuses on changes in health outcomes and in the provision of health care resources due to the proposed medical service.

Present the structure and variables of any modelled economic evaluation, with the results in a disaggregated form, before aggregating them and applying extensive sensitivity analyses.

Include financial analyses for the MBS and government health budgets.

Present any additional information of relevance to the standard assessment report.

Main body of report

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

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 request boxes 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[[1]](#footnote-1)* 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, as 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.

### 4.3.2 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 as ‘overlap’ (refer to Part II, Section C.1).

# 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;
* 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 contact the Health Technology Assessment Team ([hta@health.gov.au](file:///D:\Users\mooran\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.Outlook\40QJ059F\hta@health.gov.au) or (02) 6289 7550) to receive up-to-date detail 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? |
| --- | --- |
| * The original, signed covering letter for the assessment report (with an attachment containing the complete index to the assessment report). | Yes/No |
| * 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: | Yes/No |
| * the numbers of copies of the main body of the assessment report and details of its contents | Yes/No |
| * the numbers of copies of other parts of the assessment reports and details of their contents. | Yes/No |
| * 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 |
| * 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 |
| * Any additional technical documents, attachments and references provided separately to the main body of the assessment report (where relevant), which should: | Yes/No |
| * 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) | Yes/No |
| * have the contents identified on the cover | Yes/No |
| * be legible and in English (or accompanied by a reputable translation). | Yes/No |
| * Bound copies of the main body of the assessment report (using the agreed template), which **must**: | Yes/No |
| * 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) | Yes/No |
| * have the contents identified on the cover | Yes/No |
| * 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) | Yes/No |
| * have consistent pagination throughout | Yes/No |
| * include dividers between each Section, attachments and references, with an appropriately labelled tab extending beyond the page width | Yes/No |
| * have all cost calculations in Australian dollars (A$) | Yes/No |
| * incorporate attachments containing reports of each of the relevant randomised trials (or each of the relevant non-randomised studies, if necessary), which **must** be:   (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 |  |
| * Supply the whole assessment report and any accompanying calculations and models in electronic format (with any spreadsheet compatible with Microsoft Excel 2003, any word-processing document compatible with Word 2003 and any other software package consistent with Sub-section 6.2). 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 |
| * 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 MSAC to ensure the acceptability for evaluation of any software that is not on the maintained list of software packages.

## 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](mailto: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 Final Protocol that has been considered by PASC in the pre-assessment phase of the application to MSAC.

**A.1 Address all items in the Final Protocol**

* All items in the Final Protocol 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 Final Protocol.
* Indicate if any additional information provided in the assessment report has been compared to the Final Protocol.

## A.2 Proposed medical service

Describe the proposed medical service as set out in the Final Protocol, including the health issue to be addressed, the purpose of the investigative medical service, mode of delivery and other details.

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);
* 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; or
* form the basis of a screening program in asymptomatic populations (see Appendix 8).

## A.3 Proposed MBS listing or other public funding sought

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

## A.4 Comparator details

* Identify the main comparator(s) described in the Final Protocol. 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.

## A.5 Clinical management algorithm(s)

* Present the clinical management algorithm(s) described in the Final Protocol.
* 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 Final Protocol.
* Present the corresponding algorithm depicting the current context as listed in the Final Protocol.
* 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.

### A.5.1 Algorithms for intended and current contexts

Clinical management algorithms are most relevant to an assessment report presenting a modelled economic evaluation (see Sub-section D.1). They are also helpful for estimating changes in use and cost of other medical services (see Sub-section E.3). 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 D.2 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.

## A.6 Differences between the proposed medical service and the main comparator

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

## A.7 Clinical claim

Provide information about the clinical claim with respect to the proposed medical service, as set out in the Final Protocol.

Based on the clinical claim, state what type of economic evaluation will be used) as specified in the Final Protocol.

## A.8 Summarise the primary elements of the decision analysis (PPICO)

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 Final Protocol.

Primary elements for an investigative medical service include:

* Population and medical condition (e.g. patients with non-small cell lung cancer);
* Prior tests (i.e. any tests that would be done before the proposed investigative medical service is used);
* Proposed investigative service (e.g. genetic testing);
* Comparator investigative service (e.g. no genetic testing);
* Reference standard (see Section B.1); and
* Outcome claims:

# Section B Clinical evaluation for the proposed investigative medical service (analytical validity)

## Introduction

Investigative medical services do not *directly* improve health outcomes and might impose risks that can result in harm to patients (e.g. via radiation or biopsy). Rather, a diagnosis generated by the investigative medical service has the potential to change clinical management and thereby alter health outcomes *indirectly*through the use, avoidance or adjustment of a therapeutic medical service, or to otherwise inform other management options for the patient. This indirect impact of an investigative service on a patient’s health outcome means that the presentation of evidence for an investigative service is different to that of a therapeutic intervention.

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

* Step 1: evidence of analytical validity; and
* Step 2: evidence of health outcomes.

### Analytical validity

The analytical validity of an investigative medical service shows how well it distinguishes patients who have the target condition or clinical information of interest from those who do not.

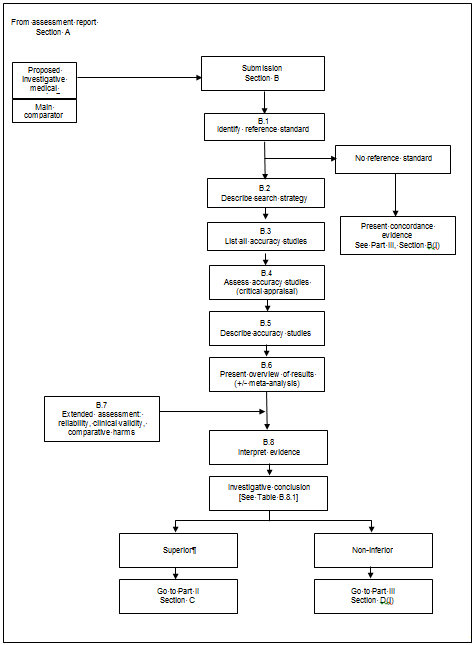
The purpose of Section B of the assessment report is to identify and present the best available clinical evidence to substantiate the analytical validity of the investigative medical service in relation to the main comparator.

Figure B.1 shows the flow of information requests for presentation of this evidence in Section B of a standard assessment report.

### Evidence of health outcomes

After the evaluation of analytical validity, the next stage is to assess the evidence for health outcomes from using the investigative service. There are a number of pathways for presenting this evidence, all of which ultimately include gathering and presenting information to provide the data required for the economic evaluation (in Section D of the assessment report). Details of the information requests for presenting the evidence of health outcomes are provided in Section C of these Guidelines.

Figure B.1 Key information requests for assessment report Section B of a standard assessment for MSAC

****

## B.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). |
| --- |

### Reference standard

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 and changes the nature of the presentation of evidence for an investigative claim.

The presence of a reference standard allows calculation of sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios and so on. This allows quantitative assessment of *analytical validity*[[2]](#footnote-2) (*accuracy* and *reliability*), *clinical validity*[[3]](#footnote-3) and *comparative harms*.[[4]](#footnote-4) This assessment informs an ‘investigative’ conclusion of whether the proposed service is superior or non-inferior to the main comparator.

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 and comparative harm. The absence of a reference standard creates more uncertainty around the investigative claim, 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.

## B.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. |
| --- |

Accuracy studies are used to document how well an investigative medical service, or a series of investigative medical services, is able to correctly identify patients with the target condition or clinical information of interest. To assess the accuracy of an investigative medical service (be it the proposed investigative medical service or the comparator investigative medical service), its results are compared with the results of the reference standard (see ‘Introduction, Step 1’, above).

The reference standard[[5]](#footnote-5) 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.

In the presence of a reference standard, the accuracy of the proposed investigative medical service and the main comparator 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 reference standard does not exist, Section III Section B(i) explains that evidence of concordance needs to be presented. If a reference standard does not exist, usual measures of accuracy cannot be generated.

### 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 and 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.

## B.3Listing 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. * 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.

### B.3a Listing accuracy studies of direct comparison

#### Search results

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

1. the study directly compared the proposed (or index) investigative medical service and the main comparator;
2. the study contained the nominated reference standard; and
3. 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 B.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 Sub-section B.3.

Table B.3.1Summary of identification of accuracy studies of direct comparison from the search of the published literature (see Section B.3)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | MEDLINE | EMBASE | Trial or accuracy registries | Other databases |
| Number of citations retrieved by search |  |  |  |  |
| Number of citations excluded after title/abstract review:   * 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:   * 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. Each citation without an annotation should thus be a report of an accuracy study included in the assessment report.

#### Master list of trials

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

Table B.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. *Journal* Year; Vol(No):pages  Author(s). Title. *Journal* 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 B.3.3from 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 study 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 B.3.3). If there is more than one type of reason for exclusion, arrange the excluded trials in Table B.3.3by the reason for exclusion.

Table B.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 B.3.4 and B.3.5).

Table B.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 B.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 B.8 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.

### B.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 Section A:

* 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 B.3a.

#### Master list of trials

From the two tables reporting the results of the expanded searches for the proposed investigative medical service, list all identified relevant citations for the proposed investigative medical service. Similarly, list all identified relevant citations of accuracy studies for the main comparator. Table B.3.5 provides a suggested format for presenting a master list of all the relevant accuracy studies identified in the search for the indirect comparison.

Table B.3.5 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. *Journal* Year; Vol(No):pages  Author(s). Title. *Journal* 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. *Journal* Year; Vol(No):pages Author(s). Title. *Journal* Year; Vol(No):pages | Yes/No |

#### Presentation of non-inferiority (equivalence) trials

If an indirect comparison is provided to support an investigative conclusion of non-inferiority or equivalence in Sub-section B.8, 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 B.3.6). 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 B.3.6 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.

## B.4Assessment 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 Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool, the Standards for Reporting of Diagnostic Accuracy (STARD) initiative and the ACCE[[6]](#footnote-6) Model Project (for genetic tests).

For simplification, the remainder of this Section is based on the QUADAS quality appraisal tool. The checklist in Box B.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.

Box B.4.1 QUADAS tool 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?   **Source:** Whiting et al (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 3:25. |
| --- |

### 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 groupwas recruited to the study to address the review question; and second, whether the method of samplingparticipants 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 accuracyof an investigative medical service is determined by comparison with an imperfectreference standard, some target condition misclassificationwill be introduced. Misclassification of the target condition by the reference standardwill tend to result in underestimation of the accuracy of theinvestigative medical service under consideration. Underestimation of the sensitivity is most likely when the prevalence of the target conditionis low, and the estimated sensitivity will be closer to thetrue sensitivity with increasing prevalence. Underestimationof the specificity will occur most when the prevalence of thetarget condition is high, and the estimated specificity willbe closer to the true specificity when the prevalence of thetarget condition is low. Good reference standards are independent of the investigative medical service under consideration, and are applied blindly or objectively to 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 B.4.1 and B.4.2provide a suggested format for presenting the summary in the main body of the assessment report.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trial ID | Sampling method/ concealment of allocation | Blinding | | |
| Participants | Investigators | Results assessors |
| Study 1 | A/B/C/Other | Yes/No | Yes/No | Yes/No |
| Study 2 | A/B/C/Other | Yes/No | Yes/No | Yes/No |
| Etc. |  |  |  |  |

ID = identification

A = central telephone randomisation service; B = third-party randomisation service (e.g. professional body, device company); C = sequentially labelled, fully opaque, sealed envelopes

Table B.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) |
| * Proposed investigative medical service | *n* (%) | *n* (%) | *n* (%) |
| * Main comparator | *n* (%) | *n* (%) | *n* (%) |

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 B.4.1 and B.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 B.4.2). Alternatively, if it is clearer for some tables, identify the source of information cell by cell, using footnotes.

## B.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 B.5.1 and B.5.2 provide a suggested format.

Table B.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 B.5.2provides a suggested format for this information.

Table B.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.

## B.6Systematic 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 B.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.
* 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 B.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.

If positive and negative predictive values are reported, refer to Sub-section B.7(b).

### 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 whenever there is more than one accuracy study reporting a common measure of accuracy.

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 B.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:

1. define the criteria used to select studies for the meta-analysis;
2. assemble the dataset from eligible studies;
3. justify assumptions made about included studies; for example, random/non-random sample of population and test procedures;
4. test for the presence/absence of heterogeneity (see above);
5. test for the presence/absence of threshold effect (see below); and
6. 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 B.6.1provides a suggested format for presenting and comparing accuracy data generated from several studies.

Table B.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*= *I2* 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 B.6.2suggests how topresent 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 B.6.2 Summary of results of the indirect comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study ID | Studies of proposed investigative medical service | | Studies of main comparator | |
| Relevant measures of accuracy  (95% CI) | Common  reference  standard    –  – | | 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 B.7(c) outlines how to present an extended assessment around comparative harms between the proposed and comparator investigative medical service.

### Clinical importance

Discuss the clinical importance of above results in terms of the clinical importance of false positives/false negatives and true positives/true negatives. Assigning the correct test is essential, if knowledge of the result generated by the investigative medical service can potentially change clinical management. 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.

Inaccurate test results can compound clinical problems created by the complexity of disease. Incorrect assignment of a test result potentially exposes the patient to an inferior treatment or denies consideration of an alternative therapeutic medical service. In some cases, misassignment 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. Where there is no effective alternative therapeutic medical service, a false negative is likely to outweigh a false positive in terms of clinical importance. Where there are effective alternative therapeutic medical services, a false positive is likely to outweigh a false negative in terms of clinical importance. This theme is further explored in Part II, Section C*.*

In addition, assess whether there is a **threshold** below which the test should not be used (e.g. either 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 (e.g. an investigative medical service that detects florid disease might not be much use, while an investigative medical service that correctly identifies subtle disease might potentially be more useful).

#### 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), as for Sub-section B.3.For a complex systematic overview, consider re-presenting the tables from the main body of the assessment report in a technical document or attachment, as described in Sub-section B.3, including additional columns or notes for each table to indicate the source of the data in each row or cell, as appropriate.

## B.7(a) 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 Bhas 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 B.7(a).1.*.*

Table B.7(a).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

### Consideration of effective analytical validity

‘Effective’ analytical validity should be considered following an assessment of accuracy and reliability. While analytical validity (AV) refers to the ideal performance of the investigative medical service, the effective analytical validity (EAV) refers to analytical validity in the real world. Applicants should identify additional factors that result in a difference between AV and EAV, and how these factors can be measured and ameliorated. These additional aspects of an investigative medical service are not traditionally considered in the context of comparative analytical validity, and would need to be presented, assessed and related to the initial investigative conclusion. For example, in pathology, the analytical validity will be affected by the nature or type of specimen required, the nature of specimen preparation, the stability of an analyte and so on. If the biopsy of a tumour taken for a mutation test does not have enough tumour cells, or the sample has been mishandled between extraction from the patient and insertion into the apparatus central to the test, the result might be inconclusive — an outcome that is relevant to the interpretation of the results in practice. In addition to potentially changing the qualitative conclusion (inferior, non-inferior, superior), this additional consideration might modify the quantitative assessment (i.e. the extent or degree of any superiority or inferiority).

## B.7(b)Extended assessment of clinical validity

| INFORMATION REQUESTS   * For each accuracy study listed, provide information on whether clinical validity of the investigative medical service was also measured. * 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. |
| --- |

*Clinical validity* relates to whether an investigative medical service answers the clinical question being asked, and refers to how an investigative medical service predicts the target condition under consideration — or, in the case of genetic testing, the degree of association between the genotype and clinical phenotype. 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. 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.

### B.7(c) 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). * 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. |
| --- |

Adverse events reported in accuracy studies can provide additional information about the direct harms of a particular test. The requirement to report adverse events applies equally to research on investigative medical services and research on therapeutic medical services. It can also be important to learn about the invasiveness and risks of the reference standard used.

Accuracy studies are often not a complete source of data on comparative harms. Thus, a wider basis of assessment of comparative harms from other sources (i.e. beyond the results of accuracy studies) is encouraged to complement rather than replicate the assessment of comparative harms presented in response to Sub-section B.5*.* 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. Extend the scope of this strategy beyond that presented in Sub-section B.2.

### B.8 Interpretation of the investigative evidence

| INFORMATION REQUESTS   * Provide a summary assessment of the overall evidence presented for an investigative claim. * Use this assessment to state the category from Table B.8.1 that best reflects the investigative conclusion of the proposed (index) investigative medical service over its main comparator, supported by the evidence presented. |
| --- |

In parallel to the key clinical evaluation of a therapeutic medical service, it is proposed that a conclusion be drawn about whether the proposed investigative medical service is non-inferior (no worse than) or superior compared to its alternatives. Include in this assessment of the evidence the main findings of:

* comparative analytical validity:
  + the level and quality of the evidence (Sub-section B.2–B.4);
  + the measures of accuracy generated (Sub-section B.5);
  + the consistency of the results throughout the trials presented (Sub-sections B.5);
  + the statistical precision of the evidence (Sub-section B.6);
  + the clinical importance and patient relevance of study results (Sub-section B.6);
  + the measures of reliability generated (Sub-section B.7(a)); and
  + the effective analytical validity generated (Sub-section B.7(b))’ and
* comparative harms (Sub-section B.7(c)).

The interpretation of the clinical data presented in assessment report Section B is crucial in determining the success of the assessment report. It is important to classify the investigative profile 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 the investigative claim). Table B.8.1 sets out a framework for this classification.

Table B.8.1 Classification of the investigative relativity of the proposed investigative medical service over its main comparator

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Comparative safety | Comparative analytical validity a | | | |
| Inferior | Uncertain b | Non-inferior c | 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 a | 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   
**a** In this table, ‘linked evidence approach’ refers to assessment of indirect health outcomes via the *Therapeutic Guidelines* (Sections B and C) or a linked approach via Section C. See the introduction of Section C(for further explanation of these pathways. **b** ‘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 between specificity, sensitivity and/or the comparative safety considerations.

**c** An adequate assessment of ‘non-inferiority’ is the preferred basis for demonstrating equivalence.

The essential difference between assessing whether the proposed investigative medical service is superior 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.

The advantage of distinguishing between non-inferiority and superiority is that non-inferiority can result in a simpler economic evaluation of cost-minimisation. This simplifies the subsequent economic evaluation because it does not require any claim for a change in clinical management or an improvement in health outcomes if funding the proposed investigative medical service would not result in an increase in overall expenditure.

Similarly, failure to demonstrate a superior investigative claim would weaken any case for an increase in expenditure. A superior investigative claim 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 can only be translated into a change in health outcomes through a linked analysis. In relation to accuracy studies that are relevant to inform an investigative conclusion, review whether the authors of these studies engineered the study design to demonstrate superiority of the proposed investigative medical service over the main comparator, or whether the hypothesis was that a specific measure of accuracy must surpasses a pre-specified value. In other words, the null hypotheses of each of these studies informs whether the accuracy studies under consideration have been designed to be either a superiority or non-inferiority study. Sample size in each study is also important to achieve a high probability of finding a statistically significant result.

An investigative medical service that is directly more harmful would need to have superior analytical performance to justify its use. If the investigative conclusion in the assessment report is that the proposed investigative medical service is no worse than the main comparator but is significantly less harmful, a judgement will need to be made whether to proceed to an incremental cost-effectiveness analysis (via a linked analysis) depending on the nature of the potential harms. 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 a conclusion on comparative harms.

If the investigative conclusion is that the proposed investigative medical service is superior to the comparator investigative medical service, it is optional to link the investigative conclusion to a change in clinical management and thus to an improvement in health outcomes (see Step 2 in ‘Introduction’). The information requests for this pathway are described in Section C.

If the investigative conclusion is such that the proposed investigative medical service is non-inferior to the comparator investigative medical service, then a cost-minimisation analysis only applies (see to Part III, Section D).

# Section C Translation Issues

## Introduction

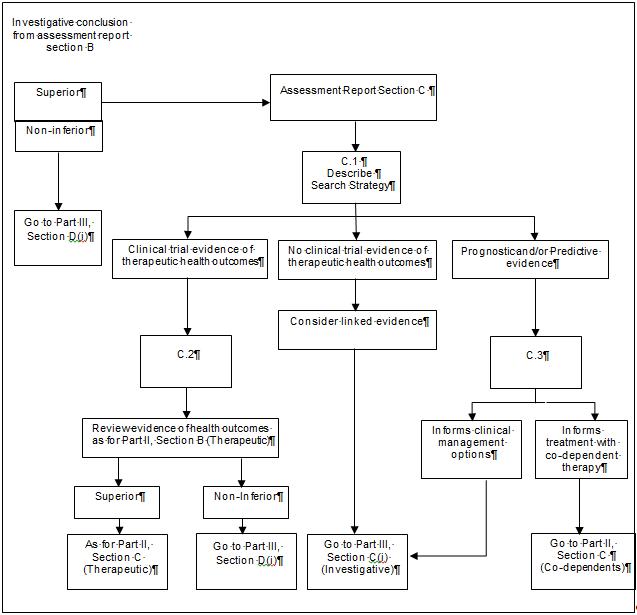
The purpose of assessment report Section C is to guide presentation of the evidence used to link the conclusion of the investigative evaluation from assessment report Section B to the economic evaluation (assessment report Section D).

As described in the introduction to Section B, if an investigative conclusion of superiority is reached, the second step in the flow of logic is to determine whether there is evidence of health outcomesrelating to the proposed investigative medical service. This requires development of a search strategy and systematic review of the literature, which might reveal one of three types of evidence:

* **Clinical trial evidence of therapeutic 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’).
* **No clinical trial evidence of therapeutic health outcomes:** given the scarcity of clinical trials or studies specifically designed to test the linkage between an investigative service and a therapeutic health outcome, it is likely that no such trials or studies will be found. In this case, applicants need to consider a linked evidence approach to evaluate potential health outcomes from use of the investigative medical service. (However, clinical trial evidence is considered to be a more convincing source of evidence.)
* **Evidence of prognosis or predictive value:** if the health outcomes from the use of the investigative medical service are improved prognosis or predictive value, the literature research will identify supporting evidence to support modification of other clinical options based on the results of the initial investigations. If the prognosis or predictive value of the initial investigation supports use of a co-dependent therapy, specific information requirements apply.

These pathways are shown in Figure C.1 with cross-reference to the relevant Sub-sections of this Section and to other Sections, as appropriate.

Figure C.1 Pathways for assessment report Section C of a standard assessment for MSAC



**Note:** ‘Clinical trial evidence of therapeutic health outcomes’ includes trials or studies specifically designed to prove the linkage between an investigative service and a therapeutic outcome. See Sub-section C.2 for further details.

## C.1Description of search strategies

| 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 clinical trial evidence of health outcomes. |
| --- |

### Search strategies

The primary objective of the search strategies is to locate *all* clinical trials 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:

1. a search of the published literature;
2. a search of registers of clinical trials; and
3. 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 trials or studies that compare (directly or indirectly) 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 investigative medical service, go to Sub-section C.2.

If there is no clinical trial evidence, other forms of evidence might be considered to link the investigative conclusion to a change in clinical management and a health outcome. Section C(i)provides the information requests for presentation of linked evidence.

If studies of prognosis or predictive value are found, go to Sub-section C.3.

## C.2 Presentation of clinical trial evidence of therapeutic health outcomes

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

### Clinical trial evidence of therapeutic health outcomes

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. 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 the proposed treatment or the main comparator.

The clinical trial evidence needs to be presented alongside the presentation of the investigative claim. This is for two reasons: first, to assist in any subsequent analysis of a competing investigative service that is claimed to be superior; and second, to provide corroborating evidence to support a claim of superiority (in much the same way as surrogate outcome data provide corroborating evidence to support a claim of superiority in terms of more directly patient-relevant outcomes for a therapeutic medical service).

The principles for presenting clinical trial evidence relating to an investigative medical service 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, however, some terminology might be different (e.g. intention to test/screen versus intention to treat).

**C.3 Presentation of evidence of prognosis and/or predictive value**

| INFORMATION REQUESTS   * State whether the information generated as a result of providing the investigative medical service under consideration is of prognostic value and/or predictive value (treatment effect modification) and present key evidence supporting this. * For an investigative medical service (e.g. a biomarker) that is co-dependent on a pharmaceutical or therapeutic medical service, follow the information requests in Part II, Section C (Co-dependents). |
| --- |

### Prognostic information

An investigative medical service might only generate prognostic information about overall health outcomes independent of therapeutic medical services offered (i.e. it does not also generate predictive information in the form of treatment effect modification). In this case, the prognostic impact of the information generated by the investigative medical service can be, for example, to allow better targeting of a broader range of existing treatment options, or to provide other clinical information of interest.

The question of prognosis 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. 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.

The HTA Groups will run an independent literature search to confirm whether the relevant prognostic evidence has been retrieved for the purposes of this Sub-section.

### Co-dependency

An investigative medical service that generates information that predicts the effect of a specific treatment (such as a pharmaceutical or a therapeutic medical service) on particular patients is known as a predictiveinvestigative medical service (i.e. it predicts whether there will be a response to treatment). Such a predictive investigative medical service is co-dependent with the therapeutic service. In this case, follow the information requests in Part II, Section C (CoD) (*Co-dependents*) in these Guidelines.

# Section C (CoD) *(Co-dependents)* Translation Issues

## Introduction

The purpose of assessment report Section C *(CoD)* is to guide presentation of the evidence of therapeutic health outcomes for an investigative medical service (particularly genetic tests or biomarkers) in cases where there is either a paired co-dependent pharmaceutical (PBS-listed) or a paired therapeutic medical service (MBS-listed). This evidence is used to link an investigative conclusion of superiority to the economic evaluation (assessment report Section D).

If a new pharmaceutical appears to be better than the current pharmaceutical, when a certain biomarker is present or absent, this can be due to the fact that the specified patient subgroup will always do better, independent of the effect of the drug*.* As mentioned above, such a case would not meet the definition of a simple paired co-dependent technology, because the co-dependent technology would not necessarily require a joint review by both PBAC and MSAC. For a pharmaceutical or therapeutic medical service to be considered co-dependent, it must interact with the identified biomarker to improve patient health outcomes. That is, there must be some form of improved treatment effect occurring as a consequence of the pharmaceutical targeting something that is intrinsic to the biomarker, or for which the biomarker is a proxy. Sometimes, a simple paired co-dependent technology may have evidence of both prognostic impact (tied to a specific biomarker) and treatment effect modification.

In all co-dependent technologies that are evaluated, the relationship between the investigative medical service (e.g. biomarker) and treatment needs to be made explicit, such that it is clear whether treatment effect modification or prognostic impact is operating in the relationship. Should both characteristics be operating, then the net treatment effect (i.e. treatment effect modification controlling or adjusting for the prognostic impact) would need to be determined as part of the evaluation of the simple paired co-dependent technology.

## C.1(CoD) ‘Overlap’ information

| INFORMATION REQUESTS   * Provide all the information identified as ‘overlap’ issues in the Draft Information Requests for Assessing a Pair of Co-Dependent Technologies. |
| --- |

When an investigative medical service is paired with a co-dependent therapy, the applicant does not need to submit a full analysis of the therapeutic evidence of health outcomes as this is submitted in full in the paired application (either to PBAC or MSAC). 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.

The information requests for assessing a pair of co-dependent technologies are set out in detail in *Draft Information Requests for Assessing a Pair of Co-Dependent Technologies* (Department of Health, 2010)*.* These requests are identified as relating specifically to the investigative medical service, specifically to the therapy (usually a drug), or to both. Requests that relate to both technologies are called ‘overlap’ questions and these are the requests that need to be addressed in this Section of the assessment report.

A list of these questions is presented in Table C.1.1( *CoD*).

Table C.1.1(CoD) ‘Overlap’ questions to answer when submitting a co-dependent assessment report to MSAC

| Information request | Rationale | Request no. in MSAC 2010a |
| --- | --- | --- |
| Define the biomarker(s) (e.g. specific genetic DNA mutation)? | Describe the nature of the genetic DNA biomarker; for example, single nucleotide polymorphisms (SNPs), mutation or copy number variation (CNV). Where relevant, include the following elements describing the context for a biomarker:   * the general clinical area * the specific use of the biomarker * the critical parameters that define when and how the biomarker should be used.   Describe exactly what the test is identifying in cases where there is no ‘specific mutation’; for example, an expression microarray of tumour tissue that identifies cancer with activation of a particular pathway, and therefore susceptibility to a certain drug, but does not identify a specific mutation as such. Categorise the mutation as either germline or somatic. If the mutation is classified as a germline mutation, then consider issues related to heritability (e.g. testing of relatives and genetic counselling would need to be considered), and assess the ethical and medico-legal implications of testing. | 4, 8 |
| What is the biological rationale for targeting that biomarker(s) with the drug? | Present the initial evidence that was relied on to select the biomarker. 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 the selection of the population sample studied in the biomarker qualification. Present the criteria used for selection of candidate genes (e.g. candidate by position, by function, based on expression profiling data). Justify, using molecular biological or pharmacological principles, the plausibility of treatment effect modification (or interaction) between the biomarker itself and the drug, or alternatively between the drug and another factor for which the biomarker is a proxy. Advise whether this rationale precedes the specification of the data collection that forms the primary source of evidence. | 9 |
| Do any other biomarkers predict variation in the comparative treatment effect (between using the drug and not using the drug)? In the case of another biomarker that is a genetic mutation:   * Have details on the specific mutation and the nature of the mutation been provided? * Is the effect of treatment on this other mutation consistent with the effect under consideration? |  | 10 |
| What is the prevalence of a true positive biomarker in the population likely to receive the test? | The source population would be those who are eligible according to the requested MBS item descriptor and PBS restriction, and would follow the corresponding clinical pathway to the point of being offered the test — or the drug in the absence of the test. An estimate of the prevalence of a true positive biomarker is relevant to calculating the performance of a test in terms of its negative and positive predictive value. Indicate where there is no ‘gold’ standard to determine this true positive status of the biomarker and use an alternative appropriate methodology to estimate it.  State whether the biomarker occurs in other settings unrelated to the focus of the assessment. The same genetic test result can carry very different implications in different settings. For example, the same mutation may be used to identify non-familial cancers for one particular cancer, and for drug selection in another type of cancer. | 11 |
| Can the proposed drug be used with other specific tests for that biomarker, other than the test proposed? What methodologies are available to test for the marker? | If other tests are MBS-listed, this would move to a more complex scenario. | 19 |
| Is there direct evidence of prognostic impact associated with different biomarker status? | This is used to discriminate prognostic impact as an alternative (or in addition) to treatment effect modification. It requires a comparison of outcomes in patients receiving usual care conditioned on the presence or absence of biomarker-positive status. | 20 |
| Is the direct evidence presented and selected in a comprehensive and unbiased manner? | For example, present a systematic review of randomised trials of the proposed drug targeting this biomarker with inclusion/exclusion criteria delineated and a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart indicating how trials were selected and reasons why any potentially relevant trials were excluded.  (This also applies to the evidence of effectiveness for the therapy, as conditioned by the test or biomarker result.) | 21, 36 |
| Is the direct evidence of good quality? | Assess bias, confounding, the impact of chance on results, and whether the analyses were pre-specified and/or exploratory. Use a study design critical appraisal checklist to cover all issues likely to affect the internal validity of the presented trial results. Confounding may occur as a consequence of imbalance in biomarker status in the drug and usual care trial arms in the case where biomarker status is also an indicator of prognostic impact.  (This also applies to the evidence of effectiveness for the therapy, as conditioned by the test or biomarker result.) | 22, 37 |
| Does the direct evidence provided show a clinically important and statistically significant impact on patient-relevant outcomes? | Assess both effectiveness and safety. Describe outcomes in the studies (primary and secondary) and statistical methods used. Provide an extended assessment of comparative harms. Assess the balance of benefits and harms and interpret findings from the body of evidence.  Evidence for the effectiveness for the therapy, as conditioned by the test or biomarker result, relates to factors intrinsic to the proposed therapy and to factors intrinsic to the proposed test. | 23, 38 |
| Is the direct evidence provided applicable to the requested MBS and PBS populations? | Describe patient characteristics in the trials and indicate whether they are relevant to the Australian situation. Indicate whether the requested technologies were provided in a setting similar to the Australian setting of use. | 24 |
| Will knowledge of the test result cause a change in management of the patient by the treating clinician? Are there instances where management would not change, despite the text indicating the biomarker is present? | There may be ‘leakage’ issues identified through an evaluation of the change in management part of a linkage. Often a test is done to rule out a drug (e.g. to avoid potential drug-related adverse events or the development of drug resistance), but the drug is given anyway, or, alternatively, the test is used to select a specific drug, but the drug is not provided. As companion tests in a co-dependent pairing will often be used to guide drug therapy decisions, this would need to be explicitly addressed. Once listed, these issues could be informed by data that compare the numbers of test ‘positive’ results and scripts filled for the drug. | 33 |
| Is there evidence available of treatment effect modification or significant interaction between biomarker status and treatment outcomes? For example, is there evidence of substantial variation in a measure of relative treatment effect between the proposed drug and usual care trial arms after stratifying for biomarker status? | Treatment effect modification in this setting identifies a relationship between the biomarker and the drug, which is likely to be unique or limited to companion tests assessing a particular biomarker and drugs with a particular mechanism of action.  Cross reference to response to request 2, above. | 34 |
| Is there evidence of better targeting to patients likely to respond most by using the prognostic impact of the biomarker on the baseline risk of disease progression? For example, is there evidence of minimal variation in a measure of relative treatment effect between the proposed drug and usual care trial arms (where the biomarker status helps to identify patients at greatest risk of an event, which helps to maximise the absolute treatment effect)? | If a drug’s result is due to better targeting to those patients who 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 nondrug treatment options) than are likely to apply for treatment effect modification. It is possible for both treatment effect modification and prognostic impact to coexist. In this case, to assess the unique contribution of the drug therapy, an assessment of the drug’s effect must be made relative to usual care and adjusted for the background prognostic impact that is operating in both the drug and usual care arms, and that is also flagged by that particular biomarker. By contrast, if the drug’s apparent improvement in result is simply due to the fact that a certain patient subgroup (flagged by a specific biomarker status) will always do better, then it does not identify a basis to pair co-dependent technologies. | 35 |
| Is the evidence supporting the pairing of the co-dependent technologies applicable to the intended PBS and MBS populations? |  |  |

**a** *Draft Information Requests for Assessing a Pair of Co-Dependent* *Technologies* (Department of Health 2010).

## C.2 (CoD) Time course of the test results

| INFORMATION REQUEST   * Where test results may change over time, provide sufficient detail to clarify the relationship and time frames between test results and the appropriateness of treatment. |
| --- |

Many primary cancers change over time, given that cancer is often a dynamic process and secondary tumours may have different mutation characteristics from primary tumours. In this instance, the sample collected at one stage may not be an accurate reflection of the genotype mutation status when treatment is started. In other words, the genotype mutation status of a stored sample taken at the time of original diagnosis may not reflect the genetic mutation status at the time the test result is relied upon.

Genetic status may also potentially alter once a cancer is being treated by a therapeutic medical service, because a resistant clone can theoretically grow at the expense of other clones within the cancer. Hence, it is important to determine whether the association between biomarker and drug sensitivity, and between biomarker and disease, persist after treatment, or whether the association lost as the tumour gains resistance.

Another factor to consider is *mosaicism*, which is well documented in familial cancer syndromes.

# Section D Economic evaluation for the main indication

## Introduction

The purpose of assessment report Section D is to present an economic evaluation of substituting the proposed 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 medical service is therapeutically superior to the main comparator; or
* the proposed medical service is non-inferior (equivalent) to the main comparator; or
* the proposed medical service is inferior to, but significantly less expensive than, the main comparator.

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

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 comparison randomised trials; see the *Therapeutic Guidelines* (Part II, Section B) or ‘stepped-to-modelled’ (i.e. direct comparison randomised trial results with pre-modelling; see the *Therapeutic Guidelines* (Part II, 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 treatment effects on health outcomes, resource use and cost effects relevant to the requested listing.

For economic evaluations that rely on incremental treatment effects based on results from either indirect comparisons of randomised trials or comparisons based on non-randomised studies (see the *Therapeutic Guidelines* Part III, Section B), consider adapting the stepped approach described here to provide a ‘modelled’ evaluation to improve the transparency of the economic evaluation (see also the *Therapeutic Guidelines* (Part III, Section C)).

Figure D.1 Key information requests for assessment report Section D of a standard assessment for MSAC

|  |
| --- |
| Key information requests for assessment report section D of a standard assessment for MSAC |

## D.1 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 comparison randomised trials presented in Section B (*Therapeutic*) or Section C (*Investigative*); * a stepped economic evaluation (i.e. derived from direct comparison randomised trials presented in Section B (*Therapeutic*) or C (*Investigative*) using variables reported in Section C (*Therapeutic*) or C (*Investigative*); and * a modelled economic evaluation based on an indirect comparison of randomised trials or non-randomised studies. * 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 randomised trials presented in C (*Investigative)* or Section B (*Therapeutic*)). If the direct comparison randomised trial(s) recruited patients directly representative of those for whom listing is sought, trialled the proposed 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 direct comparison randomised trials 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 D.1.1. This classification should be based on the differential effectiveness and safety of the 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 comparison randomised trials and will be most influenced by the results of these types of trials as the most rigorous source of data. However, MSAC has considered and will continue to consider all levels of evidence.

Where there are trade-offs between incremental effectiveness and incremental safety; that is, where there is reduced effectiveness but improved safety (see Table D.1.1) 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 D.1.1 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 foregone possible: need other supportive factors | Health foregone: need other supportive factors | ? Likely CUA |
| **Uncertain a** | Health foregone 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 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 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 application or 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.

#### Therapeutically superior service

If the proposed medical service has been shown to be therapeutically 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 D.1.1):

* **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 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 evaluation 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 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 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 therapeutic medical services 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 D.1.2 shows the type of economic evaluation that should be presented for each classification from Table D.1.1.

Table D.1.2 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: |  |
| 1. Overall, there are net benefits to patients as the benefits from improved effectiveness outweigh the harms from reduced safety and/or changed risk profile. | 1. Cost-consequences, cost-effectiveness, cost-utility, cost–benefit |
| 1. 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 | 1. 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). |
| 1. Overall, there are net harms to patients as the harms from reduced safety and/or changed risk profile outweigh the benefits from improved effectiveness. | 1. 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: |  |
| 1. Overall, there are net benefits to patients as the benefits from improved safety and/or changed risk profile outweigh the harms from reduced effectiveness. | 1. Cost-consequences, cost-effectiveness, cost-utility, cost–benefit. |
| 1. 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. | 1. 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). |
| 1. Overall, there are net harms to patients as the harms from reduced effectiveness outweigh the benefits from improved safety and/or changed risk profile | 1. No economic evaluation needs to be presented; MSAC is unlikely to recommend government subsidy of this service. |
| The proposed service is less effective than the comparator and is no worse than the comparator in terms of safety. | 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 D.1.2, 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 D.3 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 D.1.1. 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 D.3, 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 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.

## D.2 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 D.2.1).

Table D.2.1 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, to the *Therapeutic Guidelines* (Sub-section C.4) 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 (using cross-references, as appropriate, to the *Therapeutic Guidelines* (Sub-section C.4) 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 the Therapeutic Guidelines (Sub-section C.4) 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 D.6).

Table D.2.2 suggests a format that will summarise these characteristics and circumstances for which sensitivity analysis shows that the variable is important.

Table D.2.2 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 D.2.3 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 D.4.

Table D.2.3 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 T3 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 T3 greater that 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). | | |

## D.3 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 application or assessment report. The application or 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 D.4 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 Protocol, 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 D.4 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 D.4 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:

* Word 2003®;
* Excel 2003®, including [@RISK](mailto:?@RISK)®, but not necessarily including all advanced features and plug-ins (e.g. Crystal Ball® and customised macros developed using Visual Basic);
* Endnote®;
* STATA® (where this software is used please provide log-file, to allow the evaluators to check what confounders were actually controlled for etc.); and
* TreeAge Pro Suite®.

Economic evaluations constructed using any of these may be submitted without earlier arrangement with the HTA Team.

Reference lists should be provided in an Endnote library to allow ease of sorting by the evaluators.

Economic evaluations or reference lists must not be provided in PDF.

### 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

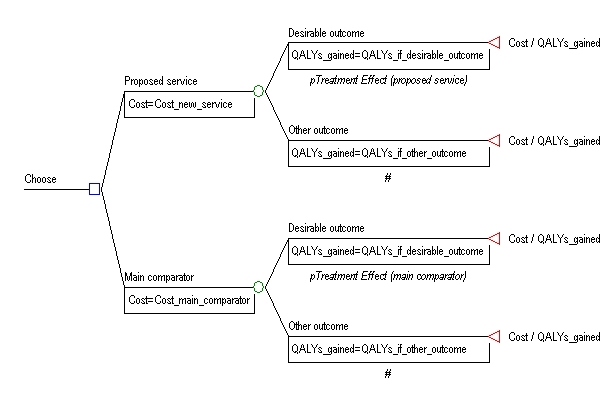
The description of the economic evaluation should include:

* 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 D.3.1.

Figure D.3.1 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 ‘protocol-driven’ 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 D.3.1.

Table D.3.1 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:   * 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). | Introduction to Section D and Section D.5 |
| Comparator | Currently available service that is most likely to be replaced by the new service | Section C |
| Type of economic evaluation | Cost-effectiveness analysis | Section D.1 |
| Source of evidence | Systematic review | Section C |
| 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 D.3 |
| 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 D.3 |
| 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 D.6, 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 application or assessment report should specify whether first-order and/or second-order distributions are sampled); and
* Markov models (the application or 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, and the treatment algorithms used in the studies presented (using cross-references, as appropriate, to Sections B and C). 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 A.5;
* 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 A.5; 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* (Section C ) 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, by reference to 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.

## D.4 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 D.5 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 D.1).

For the results of trials and pre-modelling studies conducted to provide variables for the economic evaluation, cross-refer to relevant responses to Sub-sections B.6 and C.4 of the *Therapeutic 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 (e.g. if using data or opinion that differs from the evidence on incremental treatment effects provided in response to Sub-sections B.6 and C.4 of the *Therapeutic Guidelines* ). 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 sensitivity analyses (see also Sub-section D.6).

#### 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 (ITT) 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).

### 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 D.4.1. 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 D.4.1 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 — costs from Medicare Australia a |
| Initial specialist attendance | Visit | $x | $y (MBS) | MBS item 104 — costs from Medicare Australia a |
| Subsequent specialist attendance | Visit | $x | $y (MBS) | MBS item 105 — costs from Medicare Australia 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 — costs from Medicare Australia 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 from Medicare Australia can be obtained by application to the MSAC Secretariat.

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 1). 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 D.6).

### 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. For example, refer to any analysis presented in Section C of the assessment report to transform an outcome as measured in the direct comparison randomised trials into an outcome presented in the economic evaluation. This includes transforming a modelled final outcome from a measured extent of treatment effect in the trials (see Sub-section C.2).

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 D.1 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.

#### Time-to-event data (extrapolated)

Present the calculations of the integrals between the two Kaplan–Meier curves from within the horizon of the median duration of follow-up in the trial(s), with appropriate discounting of any patient-relevant events occurring beyond 12 months of starting the treatment. Similarly, but separately, present the corresponding calculations based on the methods justified in response to Sub-section C.2 to extrapolate beyond the horizon of the median duration of follow-up in the direct comparison randomised trial(s).

Where patients transit uni-directionally in a modelled economic evaluation from one mutually exclusive health state to the next, more than one time-to-event analysis can be applied in the same economic evaluation (‘partitioned survival’). A particular application of this in economic evaluations of late-stage cancer treatment has involved the quality-adjusted time without symptoms of the disease or toxicity (Q-TWiST) health state. Time with toxicity is measured using mean time-to-treatment cessation for each arm of the trial; time in the Q-TWiST health state is measured as the difference between mean time-to-disease progression and mean time-to-treatment cessation for each arm of the trial; and time with symptoms of the disease is measured as the difference between mean time-to-death and mean time-to-disease progression. These health states are assigned utilities to then calculate QALYs gained.

### 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.

## D.5 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 D.1 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 D.5.1.

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 D.5.2 and D.5.3) 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 D.5.2).

Second, present a table that partitions the costs according to their health states across all cycles of the models (see Table D.5.3).

Table D.5.1 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 |
| *Pharmaceutical products* | | | | |
| PBS drug form and strength | $x | $y | $x – $y | z% |
| Non-PBS drug form and strength | $x | $y | $x – $y | z% |
| *Medical services* | | | | |
| Type of medical practitioner attendance | $x | $y | $x – $y | z% |
| *Hospital services* | | | | |
| Hospitalisation admission | $x | $y | $x – $y | z% |
| Outpatient clinic | $x | $y | $x – $y | z% |
| Emergency department | $x | $y | $x – $y | z% |
| *Diagnostic and investigational services* | | | | |
| Type of service | $x | $y | $x – $y | z% |
| *Allied health care services* | | | | |
| 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 D.5.2 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 D.5.3 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 | $x1 | $y1 | $x1 – $y1 | z1% |
| Health state 2 | $x2 | $y2 | $x2 – $y2 | z2% |
| Etc. | $x etc. | $y etc. | $x etc. – $y etc. | z etc.% |
| **Total** | **$x** | **$y** | **$x – $y** | **100%** |

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

Present 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 D.5.4).

Table D.5.4 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 | x1 | y1 | x1 – y1 | z1% |
| Health state 2 | x2 | y2 | x2 – y2 | z2% |
| Etc. | x etc. | y etc. | x etc. – y etc. | z etc.% |
| **Total** | **x** | **y** | **x – y** | **100%** |

#### 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 D.6) 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 Subsections A.2 and A.5).

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 D.1 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 D.5.5 and D.5.6 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 D.5.5 is aligned to the structure of Table D.2.1. More flexibility is warranted in considering the impact of extrapolating and transforming the treatment effect (Step 3). Table D.5.6 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 D.5.6 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 D.5.5 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 Sub-section  B.6) |  |
| 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 D.5.6 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 D.5.5) | (From corresponding row of Step 2 in Table D.5.5) | (From corresponding row of Step 2 in Table D.5.5) |
| For any trial-based outcome relied on in the economic evaluation *with any* ***extrapolation*** *from the time horizon of the trial(s)* ***only*** 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) (**‘*transformation of nature of outcome’ only****)*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) (****‘value transformation’ only****)* | (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 ***combining*** *any extrapolation* from the time horizon of the trial(s) ***with*** *any transformation* of the trial-based outcome(s) |  |  | (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.

## D.6 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 D.2, 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 D.6.1 gives advice on presenting this analysis in a format that is comparable to Tables D.2.1 and D.5.5.

Table D.6.1 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

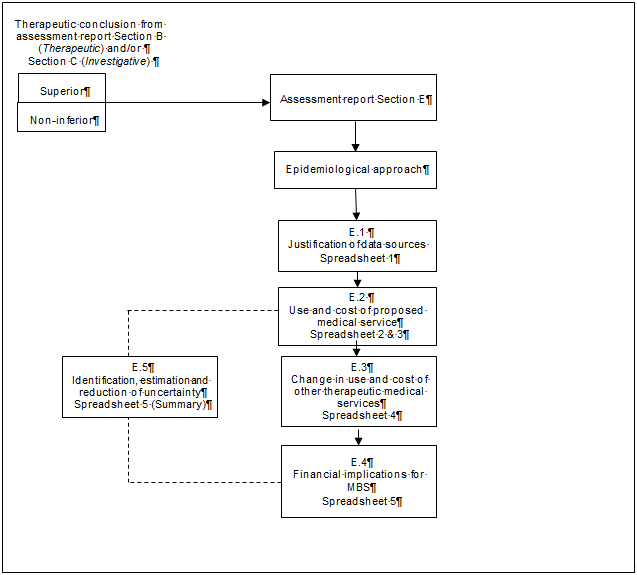
The purpose of this Section is to generate the most likely utilisation and financial estimates by requesting a set of budget impact analyses. These analyses are relevant to both the MSAC and the Australian Government. In the event of a positive recommendation by MSAC, the Australian Government needs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

Figure E.1 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 2003 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 response to Sub-section B.8) 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 E.2) and of other medical services in the context of the main indication (see Sub-sectionE.3).

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 Subsection E.5).

Figure E.1 Key information requests for assessment report Section E of a standard assessment for MSAC

****

Sub-sections E.2–E.4 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.

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 2003 workbook for each indication. As a final step in each of Sub-sections E.4 and E.5, these results can be aggregated across the indications.

## E.1 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 2003 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 E.1.1, however, there might be other suitable data sources.[[7]](#footnote-7) 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.

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) |
| * 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) |
| * 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 E.5). 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).

## E.2 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 2003 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 2003 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 2003 workbook.

## E.3 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 2003 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 B.7 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 E.2. 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 Medicare Australia. 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 2003 workbook.

## E.4 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  E.3 from the corresponding estimates calculated in Sub-section  E.2. * Use Spreadsheet 5 of the standardised Excel 2003 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 2003 workbook.

## E.5 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 uncertainly 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 2003 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 2003 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 include equity principles, ‘rule of rescue’ and other relevant factors that can affect MSAC’s assessment of proposed therapeutic 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.

## F.1 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 assumption 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 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 B(i)   
Clinical evaluation for the main indication: 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 2x2 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 B.7(a) 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.

# Section C(i) Linking an investigative conclusion of superiority to the economic evaluation when there are no clinical trials of therapeutic health outcomes

## Introduction

The purpose of assessment report Section C(i) is to guide presentation of the evidence used to link an investigative conclusion of superiority to therapeutic health outcomes (and hence to the economic evaluation Section D of the assessment report) in cases when there are no clinical trials that compare the health outcomes of clinical management including the investigative medical service with that of the main comparator (i.e. clinical management without the investigative service).

For those investigative medical services shown to have investigative superiority and for which there are clinical trials that compare therapeutic effectiveness with and without the investigative medical service, applicants should follow the information requests set out in the *Therapeutic Guidelines* (Part II, Section Band Section C).*.*

Figure C.1(i)shows a summary of the key information requests for this Section to be included in an assessment report.

Figure C.1(i) Key information requests for assessment report Section C of a standard assessment report for MSAC where there is an investigative conclusion of ‘superior’ from assessment report Section B but no clinical trials of therapeutic effectiveness

Key information requests for assessment report section C (Investigative) of an standard assessment report for MSAC where there is an investigative conclusion of ‘superior’ from assessment report section B (Investigative) but no clinical trials of therapeutic effectiveness 

**Note: ‘**No clinical trials of health outcomes’ means that after a full search of the literature, not trials or studies have been found that compare (directly or indirectly) the health outcomes of clinical management including the investigative medical service with that of the main comparator (i.e. clinical management without the investigative service).

## C(i).1Basis 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 from the Final Protocol as guidance. * 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 shown to have investigative superiority and for which no direct evidence of 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? This involves linking evidence of analytical validity with evidence that the result generated by the investigative medical services changes treatment practice, and with evidence that the alternative therapeutic medical services have different effectiveness and safety profiles.

A full, linked evidence approach is only meaningful when the evidence for analytical validity 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 flowing from the results of 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. This is further expanded in Part II, Section C*.*

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 must be presented alongside, and in addition, to evidence of the investigative performance. A linked analysis for these categories of investigative medical service is not sufficient. See Appendix 8 for further details.

## C(i).2Evidence for a change in clinical management

| 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. |
| --- |

Presenting the evidence for a change in clinical management is based on two key questions:

* 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, despite the result of the investigative medical service indicating the presence or absence of the target condition or clinical information of interest?

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. However, the proportion of patients in whom such management changes will occur cannot be predicted from accuracy data alone.

There might be ‘leakage’ issues identified through an evaluation of the ‘change in management’ part of a linkage. 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.

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

## C(i).3Evidence for change in health outcomes

| INFORMATION REQUESTS   * If evidence on health outcomes for each therapeutic medical service option has been identified, and it has been shown in Sub-section C(i).1 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 Part II, Section C (CoD). |
| --- |

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. 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. 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.

The HTA Group assigned to provide a commentary on the assessment report will run an independent literature search to confirm whether the relevant evidence has been retrieved for the purposes of this Sub-section.

# 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 medical service for the main comparator in the context of the listing requested. As already described in Sub-section B.8 and shown in Figures D.1 and D(i).1, the economic evaluation of the proposed medical service initially depends on whether the therapeutic conclusion shows:

* the proposed medical service is superior to the main comparator; or
* the proposed 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) 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.

Figure D(i).1 Key information requests for assessment report Section D of a standard assessment for MSAC in which the therapeutic conclusion from assessment report Section B is non-inferior

|  |
| --- |
| Key information requests for assessment report section D of a standard assessment for MSAC in which the therapeutic conclusion from assessment  report section B is non-inferior |

## 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 D.1). 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 (see Part I, 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 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 | The extent and nature of assumptions compared with the extent and nature of data-sourced evidence are important considerations.  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.  Issues which may impact the decision of MSAC include (but are not limited to) uncertainty related to:   * 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 | Affordable access is a central policy principle of the MBS and is considered alongside the economic evaluation.  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. |
| Presence of effective alternatives | This distinguishes between:   * an active comparator or placebo for add-on treatment; and * a placebo for no active intervention.   It also helps to define the clinical need for the proposed medical service. |
| Severity of medical condition treated | This depends on any restriction requested.  The emphasis here is only on the nature and extent of disease as it is currently managed (see Part II, Sub-section A.2). |
| Ability to target treatment with the proposed medical service precisely and effectively to patients likely to benefit most | 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).  Claims of benefits greater than the average result from the ITT analysis should be supported by appropriate trial evidence. |

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 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, Sub-section A.2), and the main comparator(s) and clinical management algorithms based on what is likely to change (see Part II, Sub-section A.5);
* interpret the clinical importance and patient relevance of the outcome measures reported in the trials (see Part II, Sub-sections B.5 and B.8);
* modify the patterns of resource use and, very rarely, the clinical outcomes measured in randomised trials conducted in different settings, such as in other countries (see Part II, Sub-section C.2);
* predict which resources would be used and how often each would be used to manage outcomes reported in the randomised trials, but not followed up (see Part II, Sub-section  C.2);
* identify the proportion of patients with the medical condition who would meet the eligibility criteria established by the requested restriction (see Part II, Sub-section E.2);
* predict the proportion of patients within this eligible population who would take the proposed medical service (see Part II, Sub-section E.2);
* predict the rates of uptake of the proposed medical service (see Part II, Sub-section E.2); and
* predict the extents of substitutions, increases and decreases of other medical services that are MBS-listed (see Part II, Sub-section E.3).

A2.3 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.

A2.4 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, Sub-sections D.6 and E.5) 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, Sub-sections B.6 or C.2 or any other input into the economic evaluation in Part II, Sub-section D.4, compare the results and justify the modification. Present a summary table that compares multiple sources or multiple variables. Table A2.1 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:   * 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.  On a question-by-question basis, assess:   * the extent to which each question is neutral or biased; and * the extent to which each question is open or closed.   To allow an independent assessment 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. |
| 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. |
| Whether all experts agreed with each response, and, if not: | 1. 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. 2. 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. |
| 1. the approach used to finalise the estimate; and |
| 1. the approach used to present the variability in the opinions. |

**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 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.

A4.2 Service delivery information

As part of the information provided in assessment report Section B, ensure that the service delivery relativity used in the trials is appropriate.

A4.3 Non-inferiority threshold

As part of the information provided in assessment report Section B, 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 (see Part III, Section B(i) for further background on these issues).

A3.4 Method of analysis

Also as part of the information provided in assessment report Section B, indicate whether the analysis of each trial 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 intention-to-treat (ITT) basis (which is the generally preferred basis for an analysis).

If one or more specifically designed non-inferiority direct comparison randomised trials are available, also describe the primary analysis of non-inferiority in detail for each such trial, 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 trials 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 Part II, Sub-section B.6, 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 (see Part II, Sub-section B.8). 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 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 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 in response to Part III, Sub-section B(i). 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 justified in response to Part III, Sub-section B(i). 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 assessment report Section B, 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 assessment report Section B, 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.

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 D.1.

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 (see Part II, Sub-sections B.5 and B.6).

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. Those recommended in Part II, Sub-section B.5 have been assessed according to the criteria for such instruments identified. 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:

1. It promotes comparability across cost-utility analyses;
2. It minimises bias by eliminating the need for an analyst intermediary;
3. It can appropriately minimise observer bias by assessing the subjective outcome of health-related quality of life under appropriate blinded conditions;
4. 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;
5. 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;
6. 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;
7. It estimates some of the distribution and heterogeneity variation of health states in a population;
8. It maintains a fixed period of assessment to which the MAUI applies;
9. 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;
10. 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);
11. 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;
12. 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 the each of the main MAUIs is well developed;
13. 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.
14. 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
15. 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:

1. 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;
2. 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
3. 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. Using a MAUI in the context of a randomised trial (see Part II, Sub-section B.5) avoids this problem. 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 C.1. 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 assessment report Section B or modified in assessment report Section C. 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 Sub-section A5.2 of 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 C.2. 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 C.2, identify and justify the estimates to be used as variables in the economic evaluation presented in Sub-section D.5 for the base case and Sub-section D.6 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 (the latter two have been assessed as acceptable for the MAUIs recommended in Sub-section B.5). 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 (see Sub-section B.5), 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 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 C.1. 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 C.2. 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 C.2, identify and justify the estimates to be used as variables in the economic evaluation presented in Sub-section D.5 for the base case and Sub-section D.6 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 Sub-section D.1).

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 Sub-section D.4 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 Sub-section D.4 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 Measures taken by the investigators to minimise bias in non-randomised studies

This appendix is designed as a useful guide to help MSAC and the applicant review the scientific rigour of the evidence by assessing the measures taken by the investigators to minimise bias in non-randomised studies that have been conducted on therapeutic medical service that is primarily provided as a result of providing the investigative medical service under consideration. It is not intended to discourage the presentation of data.

Categorise studies into the study types defined below. Then, for each methodological topic listed for the relevant study type, choose the description that best fits each study. If the submission includes a number of studies of the same type, tabulate the responses. In each case, the methodological descriptions are arranged in a descending order of quality (i.e. 1 being the worst).

As for the assessment of randomised trials in Part II, Section B and Part III, Section B(i) of the *Therapeutic Guidelines*, the purpose of these assessments is to provide the sponsor and MSAC with a clear idea of which 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. Submissions should therefore be particularly careful to justify using the results of studies with less scientific rigour in an economic evaluation in place of trials with greater scientific rigour.

There may be other aspects of particular non-randomised studies that might affect the results of such studies and their comparability with different studies of the same type. If these aspects are likely to be important, they should also be identified.

**Note**: In each case, if there is insufficient information available to classify the study, assign it to category 1.

A7.1 Classical observational designs

Controlled cohort studies

In this study type, assignment of the groups of individuals to treatment is not random. However, individuals receiving the proposed medical service are followed forward in time from their first exposure and control individuals are followed forward in time from their enrolment in the study. Cohort studies can be concurrent or historical. In the former, the study is planned and conducted prospectively. In the latter, existing records are used to define treatment status and determine the outcomes.

Possibility of confounding

It is important that there are no substantial differences at the baseline between treated and control participants in respect of factors that could influence the outcome(s) being studied. Identify which of the following best describes the differences in baseline factors:

1. There were significant differences in baseline factors between treated and control participants that have been shown to influence the study outcome(s), and these were not adjusted for in the main analysis;
2. There were significant differences in baseline factors between treated and control participants that might have influenced the study outcome(s), and these were not adjusted for in the main analysis; or
3. There were no differences in baseline factors between treated and control participants that might have influenced the study outcome(s), or any differences were adjusted for in the main analysis.

Adequacy of follow-up

It is important that an attempt is made to summarise the study outcomes for all participants who were included in the study. Identify which of the following best describes the adequacy of follow-up in the study:

1. There were significant numbers of drop-outs with no assessment of study outcome(s) in the participants who dropped out, and drop-out rates differed between treated and control groups;
2. There were some drop-outs with no assessment of study outcome(s) in the participants who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups; or
3. Study outcome(s) were assessed in all or nearly all treated and control participants.

Blinding of outcomes assessment

It is important that the observer responsible for measuring the study outcome is unaware of whether the participant belongs to the treated or control group. Identify which of the following best describes the blinding of outcomes assessment:

1. There was no attempt to blind the observer(s) to the treatment or control status of the study participants, or any attempt made was inadequate to keep the observer(s) fully blind to the treatment or control status of the study participants; or
2. The observer(s) were kept fully blinded to the treatment or control status of the study participants.

Case-control studies

In this study type, participants are defined by the presence (cases) or absence (controls) of the study outcome, and their previous use of the proposed medical service is compared.

Selection of cases

It is most important that cases are selected independently of their treatment status. Identify which of the following best describes the selection of cases:

1. The process of referral and selection of cases was likely to have been influenced by the participants’ previous use of the proposed medical service, and knowledge of the association between use of the proposed medical service and study outcome; or
2. The process of referral or selection of cases was not influenced by the participants’ previous use of the proposed medical service or knowledge of the association between use of the proposed medical service and study outcome.

Selection of controls

The purpose of the control group is to provide an estimate of the odds of exposure in participants who are free from the disease in question in the source population. Identify which of the following best describes the selection of controls:

* 1. The controls were not drawn from the same source population as the cases; or
  2. The controls were drawn from the same source population as the cases (community controls).

Possibility of confounding

It is important that there are no substantial differences between cases and controls in respect of factors that could influence the outcome being studied, other than the risk of exposure to the proposed medical service. Identify which of the following best describes the comparability of cases and controls:

1. There were significant differences in factors between cases and controls that have been shown to influence the study outcome, and these were not adjusted for in the main analysis;
2. There were differences in factors between cases and controls that might have influenced the study outcome, and these were not adjusted for in the main analysis; or
3. There were no differences in factors between cases and controls that might have influenced the study outcome, or any differences were adjusted for in the main analysis.

Possibility of measurement bias

It is important that assessment of treatment status (or exposure) is made in an unbiased way. Identify which of the following best describes the assessment of treatment status:

1. The measurement of previous use of the proposed medical service (or exposure) was made using an unstructured interview or questionnaire by an observer who was aware of the case or control status of the participants;
2. The measurement of previous use of the proposed medical service or exposure was made using a structured interview or questionnaire by an observer who was aware of the case or control status of the participants; or
3. The measurement of previous use of the proposed medical service (or exposure) was made using a structured interview or questionnaire by an observer who was unaware of the case or control status of the participants or the definition of exposure preceded the outcome.

A7.2 Quasi-experimental designs

‘Before and after’ studies

In this type of study, participants are observed before and after using a medical service. It is really only possible to use this design if the manifestations of the illness being treated are both chronic and reversible. Typically, this will be an opportunistic study, rather than planned. In addition to the sources of bias that affect the previously mentioned observational designs, this study type has particular problems related to time (or order) effects, resulting from the participants being observed over a period, and the lack of a contemporaneous control group. There may be changes in disease severity, symptomatology or resource use that occur independently of any treatment, and it is impossible to assess these properly without a contemporaneous control group. It is highly likely that participants would be switched to the new treatment because they have not been doing well on the old treatment, and thus their symptoms would tend to be most severe at the time of switching. Regression to the mean will make the new treatment seem better than the old one, in terms of both apparent treatment responses and resource provision.

Selection of participants

1. The participants were selected retrospectively from case notes, and the investigators were probably aware of the responses to the old treatment at the time of selection.
2. The study was planned, prospective data collection was undertaken in both study periods, and selection of the participants was made without knowledge of the treatment responses.

Possibility of confounding

1. There were within-participant differences in factors between the two study periods that were likely to influence the study outcome(s), and these were not adjusted for in the main analysis.
2. There were no within-participant differences in factors between the two study periods that were likely to influence the study outcome(s), or any differences were adjusted for in the main analysis.

Adequacy of follow-up

1. Drop-out rates differed between the ‘before’ and ‘after’ study periods, with no assessment of study outcome(s) in the participants who dropped out.
2. There were no drop-outs in either study period (this implies prospective data collection in both periods), or study outcome(s) were assessed in all participants who were commenced on treatment.

Blinding of outcomes assessment

1. The observer(s) responsible for outcome assessment was aware of which treatment the study participants had been receiving.
2. The observer(s) responsible for outcome assessment was kept fully blinded to the treatment being received by the study participants.

Case-series with historical controls

Typically, this type of study is carried out by a clinical department that has introduced a new management procedure and wishes to compare the results with those of patients treated previously in the department using the old management procedure. Therefore, this type of study shares the same problems of order effects as ‘before and after’ studies but does not involve the same individuals in both arms.

Selection of participants

1. The participants were selected retrospectively from case notes, and the investigators were probably aware of the responses to the old treatment at the time of selection.
2. The study was planned, prospective data collection was undertaken in both study periods, and selection of the participants was made without knowledge of the treatment responses.

Possibility of confounding

1. There were differences in factors between participants in the two study periods that were likely to influence the study outcome(s), and these were not adjusted for in the main analysis.
2. There were no differences in factors between participants in the two study periods that were likely to influence the study outcome(s), or any differences were adjusted for in the main analysis.

Adequacy of follow-up

1. Drop-out rates differed between the two study periods, with no assessment of study outcome(s) in the participants who dropped out.
2. There were no drop-outs in either study period, or study outcome(s) were assessed in all participants who began the treatment.

Blinding of outcomes assessment

1. The observer(s) responsible for outcome assessment were aware of which treatment the study participants had been receiving.
2. The observer(s) responsible for outcome assessment were kept fully blinded to the treatment being received by the study participants.

Comparison of the results of two or more single-arm studies

In addition to all the problems noted earlier with ‘before and after’ studies or case-series with historical controls, this approach has the added disadvantage that the outcome assessments were made by different investigators in different settings. It is not possible to compare the results of such studies with any confidence. Assess comparisons involving single arms extracted from randomised trials (when compared without a common reference) as comparisons of the results of two or more single-arm studies.

Selection of participants

1. In the studies for either or both alternatives, the participants were selected retrospectively from case notes, and the investigators were probably aware of the responses to the old treatment at the time of selection.
2. The studies for both alternatives were planned, prospective data collection was undertaken for all consecutive patients in the study period, and selection of the participants was made without knowledge of the treatment responses.

Possibility of confounding

1. There were differences in factors between participants in the study populations for the two alternatives that were likely to influence the study outcome(s), and these were not adjusted for in the main analysis.
2. There were no differences in factors between participants in the study populations for the two alternatives that were likely to influence the study outcome(s), or any differences were adjusted for in the main analysis.

Adequacy of follow-up

1. Drop-out rates differed between the studies for the two alternatives, with no assessment of study outcome(s) in the participants who dropped out.
2. There were no drop-outs in the studies for either alternative, or study outcome(s) were assessed in all participants who were commenced on treatment.

Blinding of outcomes assessment

1. In the studies for one or both of the alternatives, the observer(s) responsible for outcome assessment were aware of which treatment the study participants had been receiving.
2. In the studies for both alternatives, the observer(s) responsible for outcome assessment were kept fully blinded to the treatment being received by the study participants.

# 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 principles[[8]](#footnote-8) 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 condition should be a public health problem (e.g. it should be a suitable condition or disease to screen);
* The natural history of the condition should be well understood;
* There should be a recognisable latent (asymptomatic) or early symptomatic stage of the condition;
* The screening program must have a well-defined population to be screened (e.g. age, sex) and well-defined frequency of screening. Case finding should be a continuing process and not a ‘once-and-for-all’ project. If the test is advocated as part of a cluster or sequence of tests, its contribution to the overall validity of the cluster or sequence needs to be determined;
* The benefits of screening should outweigh harms (psychological and physical). Treating the condition at an early stage should be of more benefit than delayed treatment. A screening investigative medical service must have sound epidemiological evidence underpinning it where impact of uncertainty has been considered thoroughly. The impact of people’s values (utilities) and preferences must be considered (see acceptability, below). As the recipients of screening are usually people who are not exhibiting signs of illness, it is important that the screening test itself is very unlikely to cause harm;
* The benefits of screening should outweigh costs — that is, the screening should be relatively cost-effective. The cost should be economically balanced in relation to expenditure on health care as a whole (as shown in the economic evaluation);
* There should be agreed policy on whom to treat as patients; and
* The investigative medical service at the centre of the screening program being considered should be reliable (i.e. repeated tests give the same results) and acceptable to the population. The methods of the investigative medical service should be described in sufficient detail to permit its exact replication and to ensure that results are valid (i.e. reflect the true status). Ideally, the screening will be highly sensitive (i.e. detects the condition if present, with few false negatives) and highly specific (i.e. is negative if the condition is present, with few false positives), although this is not always possible.

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 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

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 be 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.

1. Please note that templates will not be placed on the MSAC website until Guidelines are finalised [↑](#footnote-ref-1)
2. *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. [↑](#footnote-ref-2)
3. *Clinical validity* relates to whether an investigative medical service answers the clinical question being asked and refers to how an investigative medical service predicts the target condition or clinical information of interest under consideration (or, in the case of genetic testing, the degree of association between the genotype and clinical phenotype). Clinical validity is related to the prevalence of the target condition or clinical information of interest. [↑](#footnote-ref-3)
4. Comparative *harms* relates to the possible harms that a patient might endure when using the proposed medical service over the main comparator. Direct comparison randomised trials are often inadequately designed to compare harms. Therefore, a wider basis of assessment of comparative harms from other sources (i.e. beyond the results of direct comparison randomised trials) is encouraged to complement rather than replicate the assessment of comparative harms. [↑](#footnote-ref-4)
5. 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***.** [↑](#footnote-ref-5)
6. ACCE gets its name from [analytic validity](http://www.cdc.gov/genomics/gtesting/ACCE/); [clinical validity](http://www.cdc.gov/genomics/gtesting/ACCE/); [clinical utility](http://www.cdc.gov/genomics/gtesting/ACCE/); and associated [ethical, legal and social implications](http://www.cdc.gov/genomics/gtesting/ACCE/). [↑](#footnote-ref-6)
7. 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 [↑](#footnote-ref-7)
8. 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-fwork/$File/screening-framework.pdf> [↑](#footnote-ref-8)