



Australian Government

National Health and Medical Research Council

NHMRC additional levels of evidence and grades for recommendations for developers of guidelines

PILOT PROGRAM 2005 - 2007

Introduction

The Health Advisory Committee (HAC) is a principal committee of the National Health and Medical Research Council (NHMRC) in Australia.

Over recent years, HAC has developed a suite of handbooks to support organisations involved in the development of evidence-based clinical practice guidelines (www.nhmrc.gov.au/publications/synopses/cp65syn.htm). However, it has been identified that the levels of evidence used by the NHMRC for intervention studies is restrictive for guideline developers, especially where the areas of study do not lend themselves to randomised controlled trials. It is expected that this issue will be addressed, along with a number of others, when the handbooks are reviewed in the future.

Levels of evidence

However, because of the urgency of the problem, as an interim measure, a table has been developed that assigns levels of evidence according to the type of research question. As well as the current NHMRC levels of evidence for interventions, 'interim' levels have been applied to studies of diagnosis, prognosis, aetiology and screening. Additional information is also provided in the form of tablenotes, as well as a study design glossary and a summary of how the NHMRC dimensions of evidence should be used (see **Part A**).

Grades of recommendations

A grading system for recommendations has also been developed as an interim measure to assist guideline developers in assessing the entire body of evidence (rather than an individual study) and indicating the strength of the recommendation (see **Part B**). An assessment matrix is provided that lists the components to be considered when judging the body of evidence, together with the range of grades. The overall grade of recommendation is based on a summation of the grading of individual components of the body of evidence assessment.

Feedback

The 'interim' levels of evidence and grading system for recommendations do not have official NHMRC status, but are being piloted until mid-2006 with feedback being sought until 30 June 2007 on their usability and applicability. Your feedback is encouraged <http://www.nhmrc.gov.au/consult/docfeedback.htm>.

Those NHMRC guidelines that are developed using the interim framework should include a preamble at the front of the document explaining that the guidelines were developed using the pilot

process, blending the official NHMRC levels with the ‘interim’ levels of evidence and grading system for recommendations.

Authors:

Kristina Coleman, Sarah Norris, Adele Weston - Health Technology Analysts Pty Ltd
Karen Grimmer, Susan Hillier - Division of Health Sciences, University of South Australia
Tracy Merlin - Adelaide Health Technology Assessment (AHTA), Department of Public Health, University of Adelaide
Philippa Middleton, Rebecca Tooher - ASERNIP-S
Janet Salisbury - Biotext

Acknowledgements:

Feedback has been provided during this document’s development phase from the following:
Paul Glasziou – Oxford University, United Kingdom
Brian Haynes – McMaster University, Canada
Andrew Oxman – Oslo, Norway (GRADE Working Group)
Nicki Jackson – Deakin University
Sally Lord and Les Irwig – University of Sydney

Implementing NHMRC dimensions of evidence including new ‘interim’ levels of evidence

This document clarifies how evidence should be assessed using the NHMRC dimensions of evidence (as set out in the NHMRC toolkit series on developing clinical practice guidelines). In addition to the current NHMRC levels of evidence for intervention studies, this document provides new ‘interim’ levels of evidence for some common research questions.

The results of each included study must be assessed according to the following three dimensions of evidence:

1. Strength of evidence

- a. Level of evidence: the design of the included study is assessed according to its place in a hierarchy. The hierarchy reflects the effectiveness of the study design to answer a particular research question. Effectiveness is based on the probability that the design of the study has reduced or eliminated the impact of bias on the results. See page 6 - *How to use the evidence: assessment and application of scientific evidence (NHMRC 2000b)*. The NHMRC levels of evidence for intervention studies, together with the new ‘interim’ levels of evidence for questions on diagnosis, prognosis, aetiology and screening are provided in [Table 1](#).
- b. Quality of evidence: each included study is critically appraised as to its methodological quality. The study is assessed according to the likelihood that bias, confounding and/or chance have influenced the results. The NHMRC toolkit *How to review the evidence: systematic identification and review of the scientific literature (NHMRC 2000a)* has various example checklists that can be used. In cases where the checklist may be lacking sufficient detail, several recent checklists are suggested that can be used (at the GAR consultant’s discretion) to supplement and/or replace the original checklist - see [Table 2](#).
- c. Statistical precision: the primary outcomes of each included study are critically appraised to determine whether the effect is ‘real’ as opposed to being due to chance (using the p-value and/or confidence interval). See page 17 - *How to use the evidence: assessment and application of scientific evidence (NHMRC 2000b)*.

2. Size of effect

This dimension is useful for assessing the clinical importance, as opposed to statistical significance, of the primary outcomes of each included study. This is calculated on the basis of the size of the effect and its corresponding 95% confidence interval. See pages 17-23 - *How to use the evidence: assessment and application of scientific evidence (NHMRC 2000b)*.

3. Relevance of evidence

This dimension assesses the relevance of the results of each individual study with respect to:

- a. Outcomes: the appropriateness of the outcomes. Are they relevant to the patient? See pages 23-27 - *How to use the evidence: assessment and application of scientific evidence (NHMRC 2000b)*.
- b. Population: are the outcomes of the study based on a similar population and therefore generalisable or applicable to the population of interest?
- c. Intervention: are the outcomes of the study a consequence of a similar intervention and therefore generalisable or applicable to the intervention of interest?

Once each included study is assessed according to the three dimensions of evidence, a grade of recommendation for the whole body of evidence can be determined (see **Part B** of this document).

Table 1. Designations of levels of evidence* according to type of research question (including tablenotes)

Level	Intervention §	Diagnosis **	Prognosis	Aetiology †††	Screening
I *	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among consecutive patients with a defined clinical presentation ††	A prospective cohort study ***	A prospective cohort study	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among non-consecutive patients with a defined clinical presentation ††	All or none §§§	All or none §§§	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparative study with concurrent controls: <ul style="list-style-type: none"> • Non-randomised, experimental trial † • Cohort study • Case-control study • Interrupted time series with a control group 	A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	Analysis of prognostic factors amongst untreated control patients in a randomised controlled trial	A retrospective cohort study	A comparative study with concurrent controls: <ul style="list-style-type: none"> • Non-randomised, experimental trial • Cohort study • Case-control study
III-3	A comparative study without concurrent controls: <ul style="list-style-type: none"> • Historical control study • Two or more single arm study ‡ • Interrupted time series without a parallel control group 	Diagnostic case-control study ††	A retrospective cohort study	A case-control study	A comparative study without concurrent controls: <ul style="list-style-type: none"> • Historical control study • Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard) ‡‡	Case series, or cohort study of patients at different stages of disease	A cross-sectional study	Case series

Tablenotes

* A systematic review will only be assigned a level of evidence as high as the studies it contains, excepting where those studies are of level II evidence.

§ Definitions of these study designs are provided on pages 7-8 *How to use the evidence: assessment and application of scientific evidence* (NHMRC 2000b).

† This also includes controlled before-and-after (pre-test/post-test) studies, as well as indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C).

‡ Comparing single arm studies ie. case series from two studies.

** The dimensions of evidence apply only to studies of diagnostic accuracy. To assess the effectiveness of a diagnostic test there also needs to be a consideration of the impact of the test on patient management and health outcomes. See *MSAC (2004) Guidelines for the assessment of diagnostic technologies*. Available at: www.msac.gov.au.

§§ The validity of the reference standard should be determined in the context of the disease 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 index test. The validity of the reference standard can be determined through quality appraisal of the study. See Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003, 3: 25.

†† Well-designed population based case-control studies (eg. population based screening studies where test accuracy is assessed on all cases, with a random sample of controls) do capture a population with a representative spectrum of disease and thus fulfil the requirements for a valid assembly of patients. However, in some cases the population assembled is not representative of the use of the test in practice. In diagnostic case-control studies a selected sample of patients already known to have the disease are compared with a separate group of normal/healthy people known to be free of the disease. In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are 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.

‡‡ Studies of diagnostic yield provide the yield of diagnosed patients, as determined by an index test, without confirmation of the accuracy of this diagnosis by a reference standard. These may be the only alternative when there is no reliable reference standard.

*** At study inception the cohort is either non-diseased or all at the same stage of the disease.

§§§ All or none of the people with the risk factor(s) experience the outcome. For example, no smallpox develops in the absence of the specific virus; and clear proof of the causal link has come from the disappearance of small pox after large-scale vaccination.

††† If it is possible and/or ethical to determine a causal relationship using experimental evidence, then the 'Intervention' hierarchy of evidence should be utilised. If it is only possible and/or ethical to determine a causal relationship using observational evidence (ie. cannot allocate groups to a potential harmful exposure, such as nuclear radiation), then the 'Aetiology' hierarchy of evidence should be utilised.

Note 1: Assessment of comparative harms/safety should occur according to the hierarchy presented for each of the research questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms are rare and cannot feasibly be captured within randomised controlled trials; physical harms and psychological harms may need to be addressed by different study designs; harms from diagnostic testing include the likelihood of false positive and false negative results; harms from screening include the likelihood of false alarm and false reassurance results.

Note 2: When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding research question eg. level II intervention evidence; level IV diagnostic evidence; level III-2 prognostic evidence.

Hierarchies adapted and modified from: NHMRC 1999; Bandolier 1999; Lijmer et al. 1999; Phillips et al. 2001.

Table 2. Assessment of individual study quality

Research question	Location of NHMRC checklist *	Additional/ supplemental quality assessment tool
Intervention	Page 45	QUADAS (Whiting et al., 2003)
Diagnosis	Page 62	
Prognosis	Page 81	GATE checklist for prognostic studies (NZGG, 2001)
Aetiology	Page 73	UK National Screening Committee Guidelines (2000)
Screening	Page 45	

* Included in *How to review the evidence: systematic identification and review of the scientific literature* (NHMRC 2000a).

Study design glossary (alphabetic order)

Adapted from NHMRC 2000; Glasziou et al. 2001; Elwood 1998

All or none - all or none of a series of people (case series) with the risk factor(s) experience the outcome. For example, no smallpox develops in the absence of the specific virus; and clear proof of the causal link has come from the disappearance of small pox after large scale vaccination. This is a rare situation.

A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among consecutive patients with a defined clinical presentation – a cross-sectional study where a consecutive group of people from an appropriate (relevant) population receive the test under study (index test) and the reference standard test. The index test result is not incorporated in (is independent of) the reference test result/final diagnosis. The assessor determining the results of the index test is blinded to the results of the reference standard test and vice versa.

A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among non-consecutive patients with a defined clinical presentation – a cross-sectional study where a non-consecutive group of people from an appropriate (relevant) population receive the test under study (index test) and the reference standard test. The index test result is not incorporated in (is independent of) the reference test result/final diagnosis. The assessor determining the results of the index test is blinded to the results of the reference standard test and vice versa.

Case-control study – people with the outcome or disease (cases) and an appropriate group of controls without the outcome or disease (controls) are selected and information obtained about their previous exposure/non-exposure to the intervention or factor under study.

Case series – a single group of people exposed to the intervention (factor under study).

Post-test – only outcomes after the intervention (factor under study) are recorded in the series of people, so no comparisons can be made.

Pre-test/post-test – measures on an outcome are taken before and after the intervention is introduced to a series of people and are then compared (also known as a ‘before-and-after study’).

Cohort study – outcomes for groups of people observed to be exposed to an intervention, or the factor under study, are compared to outcomes for groups of people not exposed.

Prospective cohort study – where groups of people (cohorts) are observed at a point in time to be exposed or not exposed to an intervention (or the factor under study) and then are followed prospectively with further outcomes recorded as they happen.

Retrospective cohort study – where the cohorts (groups of people exposed and not exposed) are defined at a point of time in the past and information collected on subsequent outcomes, eg. the use of medical records to identify a group of women using oral contraceptives five years ago, and a group of women not using oral contraceptives, and then contacting these women or identifying in subsequent medical records the development of deep vein thrombosis.

Cross-sectional study – a group of people are assessed at a particular point (or cross-section) in time and the data collected on outcomes relate to that point in time ie. proportion of people with asthma in October 2004. This type of study is useful for hypothesis-generation, to

identify whether a risk factor is associated with a certain type of outcome, but more often than not (except when the exposure and outcome are stable eg. genetic mutation and certain clinical symptoms) the causal link cannot be proven unless a time dimension is included.

Diagnostic case-control study – the index test results for a group of patients already known to have the disease (through the reference standard) are compared to the index test results with a separate group of normal/healthy people known to be free of the disease (through the use of the reference standard). In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are 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. *Note: this does not apply to well-designed population based case-control studies.*

Historical control study – outcomes for a prospectively collected group of people exposed to the intervention (factor under study) are compared with either (1) the outcomes of people treated at the same institution prior to the introduction of the intervention (ie. control group/usual care), or (2) the outcomes of a previously published series of people undergoing the alternate or control intervention.

Interrupted time series with a control group – trends in an outcome or disease are measured over multiple time points before and after the intervention (factor under study) is introduced to a group of people, and then compared to the outcomes at the same time points for a group of people that do not receive the intervention (factor under study).

Interrupted time series without a parallel control group – trends in an outcome or disease are measured over multiple time points before and after the intervention (factor under study) is introduced to a group of people, and compared (as opposed to being compared to an external control group).

Non-randomised, experimental trial - the unit of experimentation (eg. people, a cluster of people) is allocated to either an intervention group or a control group, using a non-random method (such as patient or clinician preference/availability) and the outcomes from each group are compared.

- This can include:
- (1) a controlled before-and-after study, where outcome measurements are taken before and after the intervention is introduced, and compared at the same time point to outcome measures in the (control) group.
 - (2) an indirect comparison, where two randomised controlled trials compare different interventions to the same comparator ie. the placebo or control condition. The outcomes from the two interventions are then compared indirectly (ie. A vs B and B vs C, to indirectly compare A vs C).

Pseudorandomised controlled trial - the unit of experimentation (eg. people, a cluster of people) is allocated to either an intervention (the factor under study) group or a control group, using a pseudo-random method (such as alternate allocation, allocation by days of the week or odd-even study numbers) and the outcomes from each group are compared.

Randomised controlled trial – the unit of experimentation (eg. people, or a cluster of people¹) is allocated to either an intervention (the factor under study) group or a control group, using a

¹ Known as a cluster randomised controlled trial

random mechanism (such as a coin toss, random number table, computer-generated random numbers) and the outcomes from each group are compared.

Study of diagnostic yield – these studies provide the yield of diagnosed patients, as determined by the index test, without confirmation of the accuracy of the diagnosis (ie. whether the patient is actually diseased) by a reference standard test.

Systematic review – Systematic location, appraisal and synthesis of evidence from scientific studies.

Two or more single arm study – The outcomes of series of people receiving an intervention (case series) from two or more studies are compared.

Bibliography

Bandolier editorial. Diagnostic testing emerging from the gloom? *Bandolier*, 1999;70. Available at: <http://www.jr2.ox.ac.uk/bandolier/band70/b70-5.html>

Elwood M. (1998) *Critical appraisal of epidemiological studies and clinical trials*. Second edition. Oxford: Oxford University Press.

Glasziou P, Irwig L, Bain C, Colditz G. (2001) *Systematic reviews in health care. A practical guide*. Cambridge: Cambridge University Press.

Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 1999;282(11):1061-6.

MSAC (2004). *Guidelines for the assessment of diagnostic technologies*. Medical Services Advisory Committee. www.msac.gov.au

NHMRC (1999). *A guide to the development, implementation and evaluation of clinical practice guidelines*. Canberra: National Health and Medical Research Council.

NHMRC (2000a). *How to review the evidence: systematic identification and review of the scientific literature*. Canberra: National Health and Medical Research Council.

NHMRC (2000b). *How to use the evidence: assessment and application of scientific evidence*. Canberra: National Health and Medical Research Council.

NZGG (2001). *Handbook for the preparation of explicit evidence-based clinical practice guidelines*. Wellington: New Zealand Guidelines Group. Available at: <http://www.nzgg.org.nz>

Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M (2001). *Oxford Centre for Evidence-Based Medicine Levels of Evidence (May 2001)*. Oxford: Centre for Evidence-Based Medicine. Available at: http://www.cebm.net/levels_of_evidence.asp

UK National Screening Committee (2000). *The UK National Screening Committee's criteria for appraising the viability, effectiveness and appropriateness of a screening programme*. In: Second Report of the UK National Screening Committee. London: United Kingdom Departments of Health. Pp. 26-27. Available at: <http://www.nsc.nhs.uk/>

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3(1): 25. Available at: <http://www.biomedcentral.com/1471-2288/3/25>

How to assess the body of evidence and formulate recommendations

The NHMRC has developed a process for assessing the body of evidence and formulating recommendations that should assist guideline developers and help to ensure that different guidelines are consistent in their development of evidence-based recommendations.

On the following pages are a suggested text which should be added to the methods/process section of the guidelines document to explain how recommendations have been formulated and a form for assessing the body of evidence in order to make a recommendation.

How to use the *Assessing the body of evidence form*

NOTE: The form is intended to be used for each clinical question identified prior to systematic review of the relevant literature. Prior to completing the form, each individual study relevant to the clinical question should be critically appraised and the relevant data synthesised (as per the minimum standards). The form should be used as the basis of discussion regarding the five key components important in grading the recommendations (see below for explanation).

1. **Grade each of the five components** and note any important issues arising in the discussion and grading on the form.
2. **Write an evidence statement** (page 2 of the form) summarising briefly the assessment of the five separate components.
3. The grades for each of the components and the accompanying descriptor (excellent, good, satisfactory, poor) should be written in the relevant boxes.
4. **Determine the overall grade for the body of evidence** by summing the individual component grades.

REMEMBER: A recommendation cannot be graded A or B unless the volume and consistency of evidence components are both graded either A or B.

5. **Formulate a recommendation** based on this body of evidence. The recommendation should address the original clinical question and ideally be written as an action statement.

Assessing a body of evidence and grading the recommendation arising from it

Suggested explanatory material

The application of a grade to a recommendation is based on an assessment of all the included studies for that recommendation (the ‘body of evidence’). The five components that are considered in judging the body of evidence are:

- volume of evidence (which includes the number of studies sorted by their methodological quality and relevance to patients)
- consistency of the study results
- the potential clinical impact of the proposed recommendation (including the balance of risks and benefits, the relevance of the evidence to the clinical question, the size of the patient population and resource issues)
- the generalisability of the body of evidence to the target population for the guideline
- the applicability of the body of evidence to the Australian healthcare context.

Each of these components is initially graded according to the matrix below.

Applying evidence in real clinical situations is not usually straightforward and thus the body of evidence supporting a recommendation is rarely entirely one grade for all important components. For example, a body of evidence may contain a large number of studies with a low risk of bias that are consistent but may not be directly applicable to the target population or Australian healthcare context or may not be expected to have a very large clinical impact. Alternatively, a body of evidence may only consist of one or two randomised trials with small sample sizes that have a moderate risk of bias but have a very large clinical impact and are directly applicable to the Australian healthcare context and target population. The grading process is designed to allow for this mixture of components while still reflecting the overall strength of the body of evidence supporting a recommendation.

Body of evidence assessment matrix

Component	A	B	C	D
	Excellent	Good	Satisfactory	Poor
Volume of evidence	several level I or II studies with low risk of bias	one or two level II studies with low risk of bias or a SR/multiple level III studies with low risk of bias	level III studies with low risk of bias, or level I or II studies with moderate risk of bias	level IV studies, or level I to III studies with high risk of bias
Consistency	all studies consistent	most studies consistent and inconsistency may be explained	some inconsistency reflecting genuine uncertainty around clinical question	evidence is inconsistent
Clinical impact	very large	substantial	moderate	slight or restricted
Generalisability	population/s studied in body of evidence are the same as the target population for the guideline	population/s studied in the body of evidence are similar to the target population for the guideline	population/s studied in body of evidence different to target population for guideline but it is clinically sensible to apply this evidence to target population*	population/s studied in body of evidence different to target population and hard to judge whether it is sensible to generalise to target population
Applicability	directly applicable to Australian healthcare context	applicable to Australian healthcare context with few caveats	probably applicable to Australian healthcare context with some caveats	not applicable to Australian healthcare context

* e.g. results in adults that are clinically sensible to apply to children OR psychosocial outcomes for one cancer that may be applicable to patients with another cancer

Overall grade of recommendation

The overall grade of recommendation reflects the strength of the evidence supporting it. It is based on a summation of the grading of individual components of the body of evidence assessment. **A recommendation cannot be graded A or B unless the volume and consistency of evidence components are both graded either A or B.** A standardised form has been used to assess the body of evidence for each clinical question requiring a recommendation in this guideline.

NHMRC grades of recommendation are provided to assist users of the clinical practice guideline in making clinical judgements and indicate the strength of the recommendation. Grade A and B recommendations are generally based on a body of evidence which can be trusted to guide clinical practice, whereas Grade C and D recommendations must be applied carefully to individual clinical and organisational circumstances and should be followed with care.

Grade of recommendation	Description
A	Body of evidence can be trusted to guide practice
B	Body of evidence can be trusted to guide practice in most situations
C	Body of evidence provides some support for recommendation(s) but care should be taken in its application
D	Body of evidence is weak and recommendation must be applied with caution

NHMRC - Assessing the body of evidence form

Key question:		Evidence table ref:
<i>(Circle appropriate grade for each component)</i>		
1. Volume of evidence <i>(quantity, level, methodological quality and relevance to patients of the body of evidence for this question, based on critical appraisal of each individual study according to Minimum Requirements)</i>		
	A	Excellent (several level I or II studies with low risk of bias)
	B	Good (one or two Level II studies with low risk of bias or SR/multiple Level III studies with low risk of bias)
	C	Satisfactory (Level III studies with low risk of bias or Level I or II studies with moderate risk of bias)
	D	Poor (Level IV studies or Level I to III studies with high risk of bias)
2. Consistency <i>(the degree of consistency demonstrated by the available evidence. Where there are conflicting results indicate how the group formed a judgement as to the overall direction of the evidence)</i>		
	A	Excellent (all studies consistent)
	B	Good (most studies consistent and inconsistency can be explained)
	C	Satisfactory (some inconsistency, reflecting genuine uncertainty around question)
	D	Poor (evidence is inconsistent)
3. Clinical impact <i>(the potential impact of recommendation ie. size of patient population, relevance of outcomes to the question, balance of risks and benefits, relative benefit over other management options, resource and organisational implications)</i>		
	A	Excellent (very large clinical impact)
	B	Good (substantial clinical impact)
	C	Satisfactory (moderate clinical impact)
	D	Poor (slight or restricted clinical impact)
4. Generalisability <i>(how reasonable is it to generalise from the results of the studies used as evidence to the target population for this guideline?)</i>		
	A	Excellent (directly generalisable to target population)
	B	Good (directly generalisable to target population with some caveats)
	C	Satisfactory (not directly generalisable to the target population but could be sensibly applied)
	D	Poor (not directly generalisable to target population and hard to judge whether it is sensible to apply)
5. Applicability <i>(the extent to which the body of evidence is directly applicable to Australian healthcare context)</i>		
	A	Excellent (directly applicable to Australian healthcare context)
	B	Good (applicable to Australian healthcare context with few caveats)
	C	Satisfactory (probably applicable to Australian healthcare context with some caveats)
	D	Poor (not applicable to Australian healthcare context)

6. Other factors

Indicate here any other factors that you took into account when assessing the evidence base (for example, issues that might cause the group to downgrade or upgrade the recommendation)

EVIDENCE STATEMENT

Please summarise the development group's synthesis of the evidence relating to the key question, taking all the above factors into account. Please indicate any dissenting opinions.

	Component	Descriptor	Grade
	Volume of evidence		
	Consistency		
	Clinical impact		
	Generalisability		
	Applicability		

RECOMMENDATION

What recommendation (s) does the guideline development group draw from this evidence?

	The overall grade is the summation of the grades for individual components. A recommendation cannot be graded A or B unless the volume and consistency of evidence are both either A or B.		
	GRADE OF RECOMMENDATION		