



**“Development of Clinical Practice
Guidelines for the management of
Cutaneous Melanoma and Melanoma in
special sites” 2006**

Handbook

for Chapter Leaders and Expert Working Groups

Australian Cancer Network



The University of Sydney
AUSTRALIA



**Clinical
Oncological
Society of
Australia**



**The
Cancer
Council**
Australia

This handbook was prepared for the Australian Cancer Network by

Phoebe Holt PhD
Senior Associate

Michael Frommer MB BS DObstRCOG MPH FAFPHM FAFOM
Director

Sydney Health Projects Group,
School of Public Health,
University of Sydney,

February 2006

Acknowledgments

We greatly appreciate the advice that we received from Philippa Middleton, of the Australasian Cochrane Centre, Adelaide. Ms Middleton is the NHMRC Guideline Assessment Register advisor allocated to the development of the new clinical guidelines for the management of melanoma.

We also thank Ms Rosa Reitano and Dr John Bennett, of the Sydney Health Projects Group, for their help in the production of this Handbook.

Contents

- 1 Introduction.....4**

- 2 Steps in the preparation of NHMRC clinical practice guidelines:6**
 - 2.1 Structure your research questions 6
 - 2.2 Develop a search strategy 9
 - 2.3 Search the literature 11
 - 2.4 Select and sort the literature 14
 - 2.5 Critically appraise, and summarise literature selected for review..... 15
 - 2.6 Assess the body of evidence and formulate recommendations 26

- 3 Checklist of minimum requirements for the approval of NHMRC guidelines.....31**

- 4 Summary of literature appraisal templates33**

- 5 NHMRC publications providing further information about guideline development.....39**

- 6 Further down the track – beyond the recommendations40**

- Glossary of study design terms42**

- References45**

- Appendices47**

1 Introduction

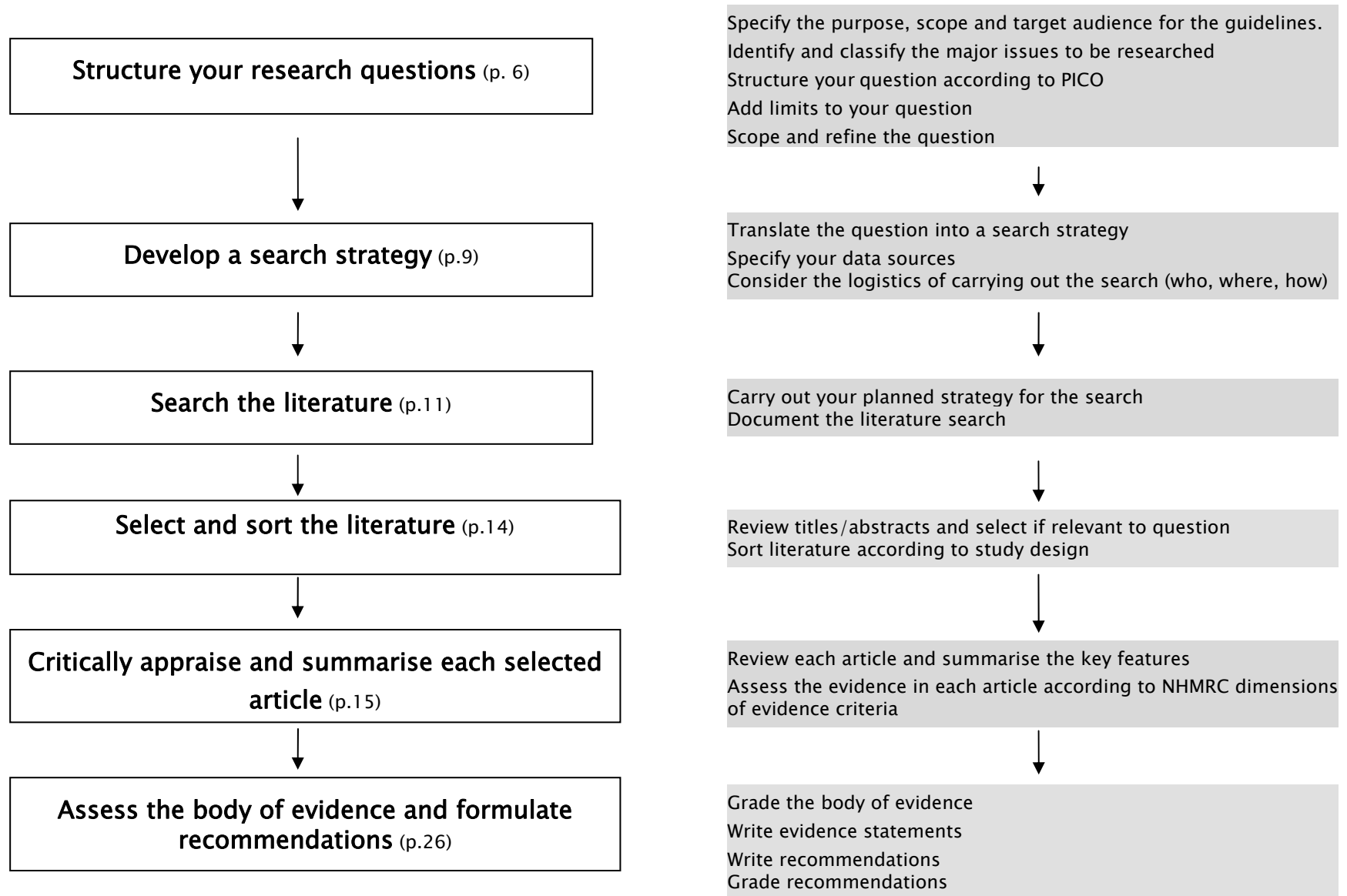
This handbook provides a summary of the key steps and requirements for members of the Expert Working Groups to prepare a set of NHMRC guidelines. Guidelines submitted to the NHMRC for approval must satisfy NHMRC legislative requirements and designated standards of quality and process.

This resource complements a series of publications on guideline development produced by the NHMRC. These publications are available on-line at the NRMRC website. See Section 5 for details.

This handbook provides a series of checklists and generic templates, which will assist users to document their search strategy, review and summarise the evidence, and formulate their guideline recommendations.

The major steps in the guideline development process are outlined in Flow Chart 1.

Flowchart 1. Steps in the preparation of NHMRC clinical practice guidelines



2 Steps in the preparation of NHMRC clinical practice guidelines:

2.1 Structure your research questions

Well-formulated research questions are crucial to a successful systematic review and to the development of clinical practice guidelines.

There are six major components to developing your questions.

A. Specify the purpose, scope and target audience for the guidelines.

The guidelines group as a whole should reach consensus on the purpose, scope and target audience for the guidelines. For example:

- Will the guidelines address prevention of melanoma through to palliation?
- Will the literature searches be limited to studies published since the 1997 melanoma guidelines?
- Will the target audience for the guidelines be GPs, surgeons, dermatologists, other medical specialists, nurses, or consumers?

Answers to these questions will influence which clinical questions are posed and how they are framed.

B. Identify and classify the major issues to be researched

Identify the major areas deserving attention in the guidelines based upon:

- “Potential for reducing clinically significant and unexplained variations in services and procedures used in the prevention, diagnosis, treatment, management or outcomes related to the clinical condition.
- Number of individuals affected by the condition
- Adequacy of scientific evidence on which to develop a guideline
- Amenability of a particular condition to prevention
- Cost to the patients.” (AHCP, 1993)

Classify the major issues you wish to research. The NHMRC recommends that guideline developers categorise their research issues/questions according to the following groupings:

1. Interventions
2. Frequency or rate of a condition or disease
3. Diagnostic test performance
4. Aetiology and risk factors
5. Prediction and prognosis
6. Economics.

Each category of research issues/questions will influence the literature search strategy, and the appraisal criteria, employed. For example, “interventions” are most likely to involve RCTs, cohort and case control studies, while “diagnostic studies” are more likely to employ cross-

sectional studies. Similarly, a researcher appraising “interventions” will have to consider randomisation, follow-up and blinding of patient and clinicians, while a researcher appraising “diagnostic studies” will need to look for independent blind comparisons with a “gold standard”, and appropriate selection of patients. (NHMRC, 1999)

C. Structure your question according to the PICO formula

Developments in evidenced based medicine have led to the specification of the “anatomy of a question”. Richardson (1995) proposed four main factors to consider when generating a study idea or clinical question: **P**atients, **I**ntervention, **C**omparison, and **O**utcome (PICO).

Patients/Population

Define the patient or population and their characteristics that may affect outcome including:

- classification
- previous treatment
- sex
- age
- general health
- medical history
- concomitant medications
- conditions of interest
- setting of interest. (AO, 2006)

Intervention (prevention, prognostic factor or diagnostic test),

Define the intervention/prevention, prognostic factor or diagnostic test. For example:

- *“an intervention:* such as a new implant or new surgical technique
- *prognosis:* factors that may influence outcome such as smoking status or age; or a prediction rule
- *diagnosis:* diagnostic procedure of interest such as a new technology.” (AO, 2006)

Comparison group(s)

Define the “control group to which the intervention group will be compared. Sometimes your question will not have a control group, such as in the case where you are interested in safety or handling characteristics of a new implant or procedure.” (AO, 2006)

Outcome(s)

Decide on “what outcomes are important to your question. Aim for the most important outcomes”. Examples could include remission, complications, repeat surgery, quality-of-life, or death. (AO, 2006)

D. Examples of questions broken down into their PICO components

"Is laparoscopic cholecystectomy more cost-effective than biliary lithotripsy in patients with gallstones?"

Population: Adult patients with gallstones

Intervention: Laparoscopic cholecystectomy

Comparison: Biliary lithotripsy

Outcomes: Safety, efficiency, costs, equity of access, patient satisfaction, quality-of-life.
(Booth and Fry-Smith, 2003)

"Will DVT prophylaxis with heparin, as compared with no method of DVT prophylaxis, produce fewer complications for patients undergoing colorectal surgery?"

Population: Adult patients undergoing colorectal surgery

Intervention: DVT prophylaxis with heparin

Control: No method of DVT prophylaxis

Outcomes: Fewer complications (e.g. DVT, pulmonary embolism, other thrombolisms).
(Birch, 2003)

E. Add limits to your question

Another method of focussing your question may include:

- "The study designs in which one is interested (E.G. randomised controlled trials or equivalent high level evidence)
- The context in which relevant studies have been undertaken, or whether there are factors that will limit its applicability to the question being asked
- Language restrictions (E.G. English only – note potential for language bias)
- Any date restrictions on the period the literature review will cover (E.G. if updating an existing review)." (Booth and Fry-Smith, 2003)

F. Scope and refine the question

Some form of preliminary searching (scoping) is usually required to validate the initial question.

Scoping searches will indicate:

- "How much literature has been published on this topic (including recent or ongoing assessments from other HTA agencies)?"
- What is the quality of the literature (E.G. is there evidence from RCTs)?"
- What are the terms used for this topic (i.e. definitions, indexing terms or keywords)?"
- Which databases are likely to provide the highest yield of relevant items?"
- What is this component of the assessment likely to cost (i.e. in terms of database search costs, literature retrieval charges and staff time for the review)?"

Strictly speaking, the results of these "pilot" searches should not be included in the review. The final search, reflecting the amended review protocol and revised inclusion and exclusion criteria, should retrieve the relevant items from scoping searches as a matter of course. Existing systematic reviews are also a means of scoping questions." (Booth and Fry-Smith, 2003)

2.2 Develop a search strategy

A Translate the question into a search strategy

The construction of a search strategy should be based on the components of the research question, i.e. PICO: populations, interventions, comparisons, outcomes, along with the study designs being considered (Booth and Fry, 2003). To summarise:

1. Specify your search concepts (e.g. population, intervention, control, and outcome)
2. Identify search terms for each concept. For each key word consider synonyms, abbreviations, related terms, differences in spelling, old and new terminology, brand and generic names, lay and medical terminology
3. Combine search terms relevant within each concept (e.g. intervention) using OR
4. Combine concepts sets using AND (i.e. keywords for Population AND keywords for Intervention AND keywords for Comparison AND keywords for Outcomes)
5. Consider using relevant limits (e.g. language, years, study designs) and specify these in your inclusion and exclusion criteria. (Glanville, 2004)

An example of combining search terms (NHS, 2005). Combine words within each column using 'OR' and combine words across the columns using 'AND'.

Example

Patient/Population and/or Problem	Intervention	Comparison/Control Intervention (if applicable)	Outcome (or effects)
<i>acute stroke</i>	<i>blood pressure reduction</i>	<i>no treatment</i>	<i>secondary prevention</i>
OR?	OR?	OR?	OR?
<i>cerebrovascular accident</i> <i>cerebrovascular event</i> <i>ischaemic stroke</i>	<i>anti-hypertensive agents</i> <i>hypertension - drug therapy</i> <i>diuretics, atenolol etc</i> <i>lowering blood pressure</i>	<i>placebo</i>	<i>secondary prophylaxis</i> <i>reduce mortality</i> <i>risk reduction</i>

Finally, the amalgamation of all four concepts should be combined further with key words and MeSH¹ terms relevant to (1) economic evaluation (i.e. economic evaluation OR cost benefit analysis OR costs, cost analysis OR economics OR cost effectiveness analysis OR cost utility analysis) and (2) socioeconomic position (see [Appendix 1](#) for key words/search terms). These searches are a minimum requirement for approval of the guidelines by the NHMRC.

¹ MeSH (Medical Subject Headings) Thesaurus is published by the US Library of Medicine

B Specify your data sources

The major types of data sources include:

- General databases: e.g. Medline, Embase
- Specialist topic databases: e.g. Cancerlit, National Health Service (NHS) Economic Evaluation Database, Health Economic Evaluations Database (HEED)
- Systematic review databases e.g. Cochrane library
- Format specific data bases: e.g. dissertation abstracts
- Grey literature: e.g. reports, proceedings etc
- Ongoing research e.g. research registers and personal communications
- Internet. (Glanville, 2004)

For a more detailed list of databases and other literature sources see [Appendix 1](#).

Minimum requirements for the use of data sources in the development of NHMRC clinical practice guidelines are:

- Medline, Clinical Evidence, Cochrane Library (CDSR, Central, HTA, DARE) – and any other databases (such as CINAHL, PsychLit) as appropriate to the topic.
- Reviews of reviews in the topic area
- At least two other forms of searching (i.e. internet, expert sources, checking of reference lists and hand searching) as appropriate for topic. (NHMRC, 2005)

C Consider the logistics of carrying out the search (who, where, when)

Things to consider in your search strategy include:

- Who will conduct the search?
- Will the searcher(s) require supervision?
- Will the searcher(s) have access to computers, the internet and printers?
- Who will submit requests for inter library loans?
- Is the library prepared for a long list of interlibrary requests?
- Who will assess the titles or abstracts generated by the search for relevance?
- Who will check the quality of the search?

Further information on developing a search strategy, are available from:

NHMRC Toolkit No. 1. *How to review the evidence: systematic identification and review of the scientific literature. 1999* (www.nhmrc.gov.au/publications/synopses/cp65syn.htm)

2.3 Search the literature

A. Carry out your planned strategy for the search

In the process of carrying out your planned search:

- **Ensure** you keep a printout of all search histories. This will assist you to remember which words you combined, which were MESH terms, which were free-text, which terms were used in each database, and which limits you used (.e.g. years for the search).
- Forward any articles you come across that are relevant to colleagues answering other questions in your Chapter group.
- Jot down any problems that arise in the process of conducting the search.

B. Document the literature search

The Expert working groups will be asked to provide details on:

- Electronic databases searched
- Terms used to search the databases
- Search Inclusion/Exclusion criteria
- Period searched (dates covered by the search)
- At least two other methods of searching
- Abbreviations
- Methods used to assess the quality of the search. (NHMRC,1998)

When reporting on database searches it is important to include:

- the database used (e.g. Medline)
- the database provider (e.g. OVID).
- Search strategy (search terms used), keep a print out
- date of the search ,
- dates covered by the search (e.g. 1990– March 2006). (Glanville, 2004: NHS, 2005)

When reporting on website searches include:

- Name of the resource
- Publisher of the resource
- Web address (URL)
- Search terms used
- Date accessed. (NHS, 2005)

When reporting on the use of search engines (e.g. Google) include:

- Search engine used
- Web address
- Search terms used
- Date searched. (NHS, 2005)

Your search information can be presented in Template 1 which is based upon a modified template designed by the Diabetes Australia Consortium for the NHMRC Type 2 diabetes mellitus guidelines.

Template 1. An example of a template to document the literature search

Macrovascular Disease Guideline Search Strategy and Yield

Electronic data bases searched and years covered in search

Medline (1966–2006) OVID, Embase (1988–2006) OVID, CINAHL (1982–2006) OVID, Cochrane (1993–2006) CD

Other search methods

Reference lists at the end of review articles were hand searched

Relevant articles were solicited from expert colleagues and organisations

Local and international clinical practice guidelines were reviewed for relevant references

Search inclusion/exclusion criteria

Limited to English Language and Human research

Search terms

See Table 1. below

Abbreviations

The database searched has been indicated next to each set of keywords using the following abbreviations.

M = Medline, EM = EMBASE, CO = Cochrane and CI = CINAHL.

The symbol “/” after a word suggests that this is a MESH term

The abbreviation “.mp” next to a word means that it was searched for as a free text word

NIDDM = non–insulin–dependent diabetes mellitus

CVD = cardiovascular diseases

PVD = peripheral vascular diseases

Identified = number of titles arising from the search using the MESH terms or free text words

Relevant = those articles considered relevant to the questions being asked after viewing titles or abstracts

Articles identified by other strategies = articles identified by hand searching, other searches from other questions, other clinical practice guidelines or from colleagues

(adapted from National Evidence Based Guidelines for the Management of Type 2 diabetes Mellitus 2001)

Template 1. Documenting the Search

	Questions	Key Words	No. Articles Identified	No. Relevant Articles	Articles Identified by Other Strategies (e.g.)	Total for Review	Total reviewed
1	Does the absence of palpable pedal pulses predict the development of complications of diabetes in the asymptomatic patient?	NIDDM/ AND pulse/ M Pedal pulse.mpM *Pulse/ AND *Medical history taking/ 89-99M NIDDM/ and pulse/CI NIDDM/ and pulse rate/ EM 88-99 Pedal.pulse.mp EM NIDDM.mo and pulse.mo CO NIDDM.mp AND pedal pulse.mp CO	18 25 4 1 6 14 7 0	3 6 1 1 0 3 0 0	1	15	7
2							
3							

2.4 Select and sort the literature

Appraise the literature for relevance to your question. Consider the following steps in the process and the individuals who may be able to complete the task:

- Review titles from the search (experts or research staff or students or others with relevant knowledge/skills)
- Print abstracts (research staff/students as above)
- Review abstracts (typically experts or staff/students under supervision by experts)
- Identify articles for review and those for background information (typically experts or staff/students under supervision by experts)
- Obtain articles from internet, library or interlibrary loans (research staff/students, clerical staff)
- Sort studies by type of study e.g. interventions, prognosis, diagnosis (experts, or research staff)
- Sort them by study design e.g. systematic review, randomised controlled trial, cohort, case control, case series, descriptive (experts, or research staff)
- Enter the number of articles for review for each set of search terms in the “Table of search terms and yield” (Template 1, Section 2.3)
- Prepare folders to file searches, background papers and reviewed articles for each question addressed
- Enter selected articles for review into the guideline master list
- Assess the quality of the search and the appraisal e.g.
 - Double cull a sample of the abstracts/titles (i.e. two independent people)
 - Send searches, search results and selected studies to your colleagues or the ACN.

2.5 Critically appraise, and summarise each selected article

A Review each article and summarise the key features

Each article selected should be reviewed and the key features documented. The NHMRC has recently provided an example of the types of headings which could be used to summarize an article in a table. This table, with minor modifications, is presented over page (p.16) as Template 2. The headings in this summary table for the literature may need to be modified according to different study designs under review.

These summaries of the each article will be invaluable to the Chapter Leaders/experts who have to write up and assess the body of evidence for each guideline recommendation. The Health Advisory Committee of the NHMRC may also request to view these summaries if they have a question about the guideline development process, or the guideline recommendations.

B Assess the evidence in each article according to NHMRC dimensions of evidence criteria

Each article needs to be assessed with respect to the following NHMRC dimensions of evidence:

- strength of evidence which includes (1) level of evidence, (2) quality of evidence, and (3) statistical precision
- size of the effect
- relevance of the results of each individual study with respect to outcomes, population and intervention.

A template for recording the results of these assessments of evidence is presented as Template 3 on Page 17.

The following Sections 2.5C, 2.5D, 2.5E, 2.5F and 2.5G provide information on how to rate the NHMRC dimensions of evidence.

Further detailed information on assessing the strength of evidence, the size of effect and the relevance of the results is available in the NHMRC publications listed in Section 5 and in particular:

NHMRC Toolkit No. 3. *How to use the evidence: assessment and application of scientific evidence 2000* (www.nhmrc.gov.au/publications/synopses/cp69syn.htm)
NHMRC Toolkit No. 5. *How to compare the costs and benefits: evaluation of the economic evidence 2001* (www.nhmrc.gov.au/publications/synopses/cp73syn.htm)
NHMRC Toolkit No. 6. *Using socioeconomic evidence in clinical practice guidelines 2002* (www.nhmrc.gov.au/publications/synopses/cp89syn.htm)

Template 3 Assessment of the evidence in each article

<i>Name of study</i>	<i>Study Type</i>	<i>Level of evidence</i>	<i>Quality of evidence rating</i>	<i>P Value</i>	<i>Size of the effect rating</i>	<i>Relevance of evidence rating</i>
Abbott RD (1987)	Cohort	II	Medium	p=0.003	1	4
Curb JD (1996)	RCT	II	High	p=0.04	2	1
Flack JM (1995)	Populati on	II	Medium	p=0.52	3	3
Fuller jj(1999)	Systema tic Review	I	High	p=0.01	2	1
Gaede P (1999)	RCT	II	High	p=0.92	3	1

(adapted from National Evidence Based Guidelines for the Management of Type 2 diabetes Mellitus 2001)

C. Levels of evidence

The level of evidence in each article should be assessed according to the criteria provided in Table 1. Your assigned level of evidence for each study should be recorded in Template 3 (p.17) under the heading “level of evidence”.

Table 1. Designations of levels of evidence* according to type of research question (NHMRC, 2005)

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 pseudo-randomised 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 pseudo-randomised 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

Key to Table 1 (from NHMRC, 2005)

* 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 (i.e. utilise A vs. B and B vs. C, to determine A vs. C).

‡ Comparing single arm studies i.e. 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 (i.e. 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.

D. Quality of evidence

The following checklists provided by the NHMRC should to be considered when rating the quality of evidence in each article. Each article should be given a rating of high, medium or low for quality of evidence. This rating should be placed in Template 3 (page 17) in the column entitled quality of evidence.

Table 2. Quality criteria for randomised controlled trials, cohort studies, case-control studies and systematic reviews

Study type	Quality criteria	YES/NO
Randomised controlled trials	<p>Was the study double blinded?</p> <p>Was allocation to treatment groups concealed from those responsible for recruiting the subjects?</p> <p>Were all randomised participants included in the analysis?</p>	
Cohort studies	<p>How were subjects selected for the 'new intervention'?</p> <p>How were subjects selected for the comparison or control group?</p> <p>Does the study adequately control for demographic characteristics, clinical features and other potential confounding variables in the design or analysis?</p> <p>Was the measurement of outcomes unbiased (i.e. blinded to treatment group and comparable across groups)?</p> <p>Was follow-up long enough for outcomes to occur?</p> <p>Was follow-up complete and were there exclusions from the analysis?</p>	
Case-control studies ^b	<p>How were cases defined and selected?</p> <p>How were controls defined and selected?</p> <p>Does the study adequately control for demographic characteristics and important potential confounders in the design or analysis?</p> <p>Was measurement of exposure to the factor of interest (eg the new intervention) adequate and kept blinded to case/control status?</p> <p>Were all selected subjects included in the analysis?</p>	
Systematic reviews ^c	<p>Was an adequate search strategy used?</p> <p>Were the inclusion criteria appropriate and applied in an unbiased way?</p> <p>Was a quality assessment of included studies undertaken?</p> <p>Were the characteristics and results of the individual studies appropriately summarised?</p> <p>Were the methods for pooling the data appropriate?</p> <p>Were sources of heterogeneity explored?</p>	

^aBased on work of Schulz et al (1995) and Jadad et al (1996).

^bBased on quality assessment instruments developed and being tested in Australia and Canada.

^cBased on articles by Greenhalgh (1997) and Hunt and McKibbin (1997).

Table 3. Checklist for appraising the quality of studies of interventions

	YES/NO
1. Method of treatment assignment	
a Correct, blinded randomisation method described OR randomised, double-blind method stated AND group similarity documented	
b Blinding and randomisation stated but method not described OR suspect technique (eg allocation by drawing from an envelope)	
c Randomisation claimed but not described AND investigator not blinded	
d Randomisation not mentioned	
2. Control of selection bias after treatment assignment	
a Intention to treat analysis AND full follow-up	
b Intention to treat analysis AND <15% loss to follow-up	
c Analysis by treatment received only OR no mention of withdrawals	
d Analysis by treatment received AND no mention of withdrawals OR more than 15% withdrawals/loss-to-follow-up/post-randomisation Exclusions	
3. Blinding	
a Blinding of outcome assessor AND patient and care giver	
b Blinding of outcome assessor OR patient and care giver	
c Blinding not done	
4. Outcome assessment (if blinding was not possible)	
a All patients had standardised assessment	
b No standardised assessment OR not mentioned	

Source: modified from I Chalmers, Cochrane Handbook; available on the Cochrane Library CDROM.

Table 4. Checklist for appraising the quality of studies of diagnostic accuracy

	YES/NO
Descriptive information about the study	
▪ Study identification	
▪ What is the study type?	
▪ What tests are being evaluated?	
▪ What are the characteristics of the population and study setting?	
▪ Is the incremental value of the test being compared to other routine tests?	
Has selection bias been minimised?	
▪ Were patients selected consecutively?	
Was follow-up for final outcomes adequate?	
▪ Is the decision to perform the reference standard independent of the test results (i.e. avoidance of verification bias)?	
If not, what per cent were not verified?	
Has measurement bias been minimised?	
▪ Was there a valid reference standard?	
▪ Are the test and reference standards measured independently (i.e. blind to each other)?	
▪ Are tests measured independently of other clinical and test information?	
▪ If tests are being compared, have they been assessed independently (blind to each other) in the same patients or done in randomly allocated patients?	
Has confounding been avoided?	
▪ If the reference standard is a later event that the test aims to predict, is any intervention decision blind to the test result?	

Source: modified from Cochrane Methods Working Group on Diagnostic and Screening Tests.

Table 5. Checklist for appraising the quality of studies of aetiology and risk factors

This set of criteria should be used for appraising studies of the extent to which the characteristics or behaviour of a person, an environmental exposure or the characteristics of a disease alter the risk of an outcome. Note: Items marked [Ⓞ] minimise bias.	
Information about the study	
Study identification.	
What is the study type?	
What risk factors are considered?	
What outcomes are considered?	
What other factors could affect the outcome(s)?	
What are the characteristics of the population and study setting?	
Evaluation criteria for the study	YES/NO
Are study participants well-defined in terms of time, place and personal characteristics? [Ⓞ]	
What percentage of individuals or clusters refused to participate? [Ⓞ]	
Are outcomes measured in a standard, valid and reliable way? [Ⓞ]	
Are risk factors and outcomes measured independently (blind) of each other? [Ⓞ]	
Are all important risk factors included in the analysis? [Ⓞ]	
What percentage of individuals or clusters recruited into the study are not included in the analysis (i.e. loss to follow-up)? [Ⓞ]	
Overall assessment of the study	
How well does the study minimise bias? What is the likely direction in which bias might affect the study results?	
Include other comments concerning areas for further research, applicability of evidence to target population, importance of study to policy development.	

Source: modified from Liddle et al (1996).

Table 6. Checklist for appraising prognostic studies

	YES/NO
A Has selection bias been minimised?	
B Have adequate adjustments been made for residual confounding?	
C Was follow-up for final outcomes adequate?	
D Has measurement of misclassification bias been minimised?	

Table 7. Checklist for appraising economic evaluation studies

Appraisal item for internal validity	YES/NO
1. Was the study question well defined?	
2. Were appropriate health care options chosen and clearly described?	
3. Was an appropriate study type used?	
4. Was the effectiveness of the health care options established?	
5. Were the cost estimates related to baseline population risk?	
6. Were all the relevant costs and consequences identified for each health care option?	
7. Were costs and consequences measured accurately?	
8. Were costs and consequences valued credibly?	
9. Was differential timing considered?	
10. Was incremental analysis performed?	
11. Was a sensitivity analysis performed?	
12. Were modelling techniques used in a clear and reasonable way?	

The NHMRC has provided additional checklists to assess the quality of studies, if the earlier tables are not sufficient.

Table 8. Assessment of individual study quality

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

E. Statistical precision

Statistical precision can be measured through the magnitude of the P value and the precision of the estimate of the treatment effect (width of the confidence intervals). Where possible, the P value(s) for each study should be placed in Template 3 (p.17) under the heading “statistical precision”. (NHMRC, 2000)

F. Size of the effect

Size of the effect should be rated according the criteria in Table 9 from the NHMRC and placed in Template 3 (p.17) under the heading “size of the effect”.

Table 9. Classifying size of the effect

Ranking	Clinical importance of benefit
1.	A clinically important benefit for the full range of plausible estimates The confidence limit closest to the measure of no effect (the ‘null’) rules out a clinically unimportant effect of the intervention
2.	The point estimate of effect is clinically important BUT the confidence interval includes clinically unimportant effects
3.	The confidence interval does not include any clinically important effects
4.	The range of estimates defined by the confidence interval includes clinically important effects BUT the range of estimates defined by the confidence interval is also compatible with no effect, or a harmful effect

Key points for considering the size of the effect

1. The size of the effect is important because it relates to the clinical importance of the effect.
2. The size of the effect should be expressed in both relative and absolute terms (i.e. as relative risks and absolute risk reductions or NNT for a range of baseline risks).
3. The size of the effect and the certainty with which it is known should both be assessed (see Table 10).

G. Relevance of the evidence

Relevance of the evidence should be rated according to the criteria in Table 10 from the NHMRC and placed in Template 3 (p.17) under the heading “relevance of the evidence”.

Table 10. Classifying the relevance of the evidence

Ranking	Relevance of the evidence
1.	Evidence of an effect on patient-relevant clinical outcomes, including benefits and harms, and quality of life and survival.
2.	Evidence of an effect on a surrogate outcome that has been shown to be predictive of patient-relevant outcomes for the same intervention.
3.	Evidence of an effect on proven surrogate outcomes but for a different intervention.
4.	Evidence of an effect on proven surrogate outcomes but for a different intervention and population.
5.	Evidence confined to unproven surrogate outcomes.

Key points for considering patient-relevant outcomes

1. The goal of decision making in health care is to choose the intervention(s) (which may include doing nothing) that is (are) most likely to deliver the outcomes that patients find desirable (Eddy 1990a).
2. Surrogate outcomes (such as blood pressure measurements or levels of serum cholesterol) may be reasonable indicators of whether there has been some effect. However, they should not be the basis for clinical decisions unless they reliably predict an affect on the way the patient feels; otherwise they will not be of interest to the patient or their carers (Eddy 1990a).
3. All possible outcomes that are of most interest to patients (particularly harms) should be identified and evaluated.
4. It may be useful to classify the relevance of the evidence using the categories shown in Table 11.

2.6. Assess the body of evidence and formulate guideline recommendations

The NHMRC has developed a process for assessing the body of evidence and formulating recommendations that should assist guideline developers and to ensure that guidelines are consistent in their development of evidence-based recommendations. The following is an extract from a recent NHMRC publication which details how to assess the body of literature and formulate recommendations (NHMRC, 2005). The NHMRC have also provide a suggested checklist and form for recording the evaluation of body of evidence and grading recommendations (see Template 4 on pages 31–32).

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

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 subsequent recommendation

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 presented in Table 12.

Table 12. 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 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 but it is clinically sensible to for guideline 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 children OR psychosocial outcomes for one cancer that may be applicable to patients with another cancer.

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.

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 (see Template 4 p29).


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.

Table 13. Grading recommendations

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

Template 4. 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 i.e. 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)	
Other factors				
<i>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)</i>				
EVIDENCE STATEMENT				
<i>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.</i>				
		Component	Descriptor	Grade
		Volume of evidence		
		Consistency		
		Clinical impact		
		Generalisability		
		Applicability		
RECOMMENDATION		<p>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.</p> 		
<i>What recommendation (s) does the guideline development group draw from this evidence?</i>				
		GRADE OF RECOMMENDATION		

3 Checklist of minimum requirements for the approval of NHMRC clinical guidelines

The NHMRC has recently outlined minimum requirements for the approval of guidelines by HAC/NHMRC.

1 How to review the evidence: systematic identification and review of the scientific literature.

Systematic review

TICK if achieved !	Minimum Requirement
	Structured research questions should be developed by the multidisciplinary group (the questions will vary depending on the scope of the guidelines, ideally should be broad rather than narrow, but must be clear, focused and closely define the boundaries of the topic, PICO criteria)
	Discuss and state the inclusion criteria/basis of inclusions and exclusions (restrictions on language, study design, etc)
	Document the search strategy
	Searching:
	<ul style="list-style-type: none"> • minimum number of databases – Medline, Clinical Evidence, Cochrane Library (CDSR, Central, HTA, DARE) – and any other databases (such as CINAHL, PsychLit) as appropriate to the topic.
	<ul style="list-style-type: none"> • search for Reviews of reviews in the topic area
	<ul style="list-style-type: none"> • use at least two other forms of searching (i.e. internet, expert sources, checking of reference lists and hand searching) as appropriate for topic
	Delineate the search period (e.g. start from when the technique/intervention was first practised, or when the database first collated citations)
	Appraise studies using critical appraisal checklist – as set out in the NHMRC toolkit publication: <i>How to review the evidence: systematic identification and review of the scientific literature</i>
	Develop appropriate tables (templates to be developed)

2. How to use the evidence: assessment and application of scientific evidence

TICK if achieved !	Minimum Requirement
	Documented approach to applying NHMRC dimensions of evidence for studies to all primary outcomes (including NHMRC levels of evidence and quality assessment) – as set out in the NHMRC Toolkit publication: <i>How to use the evidence: assessment and application of scientific evidence – Section 1</i>
	Access the applicability of the evidence – as set out in the NHMRC Toolkit publication: <i>How to use the evidence: assessment and application of scientific evidence – Section 2</i>

3. How to compare the costs and benefits: evaluation of economic evidence

Note: The developers are not required to conduct their own economic evaluation.

TICK if achieved !	Minimum Requirement
	Ensure both the cost-effectiveness and feasibility of any recommendation/s are developed following assessment of the economic evidence as set out in the NHMRC Toolkit publication: <i>How to compare the costs and benefits: evaluation of the economic evidence</i> . In cases where this is not possible, the guidelines must contain a statement that confirms that every effort had been made to discover the existence of such an analysis

4. Using socioeconomic evidence in clinical practice guidelines

TICK if achieved !	Minimum Requirement
	Ensure evidence of socio-economic differences in prevention or treatment outcomes is explicitly considered in formulation of the recommendation/s and where no such evidence exists, the search efforts are documented to the standard of the NHMRC Toolkit: <i>Using socioeconomic evidence in clinical practice guidelines</i>

5. How to put the evidence into practice: implementation and dissemination strategies

TICK if achieved !	Minimum Requirement
	Include a dissemination and/or implementation plan as set out in the NHMRC Toolkit publication: <i>How to put the evidence into practice: implementation and dissemination strategies</i>

6. How to present the evidence for consumers: preparation of consumer publications

Note 1: Companion documents (e.g. for general practitioners or consumers) are not mandatory.

Note 2: If guideline developers wish for the companion documents to display the NHMRC logo, they must seek agreement from HAC within six months of the primary document being approved by the NHMRC.

TICK if achieved !	Minimum Requirement
	Document the evidence of pilot testing or focus group testing that is undertaken for the companion documents as set out in the NHMRC Toolkit publication: <i>How to present the evidence for consumers: preparation of consumer publications</i>

4 Summary of literature appraisal templates

In this section we present four templates that satisfy the documentation requirements of the NHMRC according to their published standards. Details on how to complete these templates are presented in Section 2 of this handbook.

Template 1. Documenting the literature search

Melanoma Guideline Search Strategy and Yield <hr/>
Electronic data bases searched and years covered in search
Other search methods
Search inclusion/exclusion criteria
Search terms
Abbreviations
Identified = number of titles arising from the search using the MESH terms or free text words
Relevant = those articles considered relevant to the questions being asked after viewing titles or abstracts
Articles identified by other strategies = articles identified by hand searching, other searches form other questions, other clinical practice guidelines or from colleagues

Template 1 continues over page


Template 2. Summary of reviewed articles

Bibliographic reference	Study type	Level of evidence	Number of patients	Patient characteristics	Intervention	Comparison	Length of follow-up	Outcome measure	Results	Economic/ socio-economic information	Source of funding	Additional comments

Extracted/adapted from Draft NHMRC GAR minimum requirements and protocols

Template 4. Assessing the body of evidence

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 i.e. 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)	
Other factors				
<i>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)</i>				
EVIDENCE STATEMENT				
<i>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.</i>				
		Component	Descriptor	Grade
		Volume of evidence		
		Consistency		
		Clinical impact		
		Generalisability		
		Applicability		
RECOMMENDATION		<p>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.</p> 		
<i>What recommendation (s) does the guideline development group draw from this evidence?</i>				
		GRADE OF RECOMMENDATION		

5 NHMRC publications which provide further information on guideline development

The following NHMRC publications offer further information on developing NHMRC guidelines. All are available on the NHMRC website, and include:

A guide to the development, implementation and evaluation of clinical practice Guidelines, NHMRC (1999)

(www.nhmrc.gov.au/publications/synopses/cp30syn.htm)

NHMRC Toolkit No. 1. *How to review the evidence: systematic identification and review of the scientific literature. 1999* (www.nhmrc.gov.au/publications/synopses/cp65syn.htm)

NHMRC Toolkit No. 2. *How to present the evidence for consumers: preparation for consumer publications 1999* (www.nhmrc.gov.au/publications/synopses/cp66syn.htm)

NHMRC Toolkit No. 3. *How to use the evidence: assessment and application of scientific evidence 2000* (www.nhmrc.gov.au/publications/synopses/cp69syn.htm)

NHMRC Toolkit No. 4. *How to put the evidence into practice: implementation and dissemination strategies 2000* (www.nhmrc.gov.au/publications/synopses/cp71syn.htm)

NHMRC Toolkit No. 5. *How to compare the costs and benefits: evaluation of the economic evidence 2001* (www.nhmrc.gov.au/publications/synopses/cp73syn.htm)

NHMRC Toolkit No. 6. *Using socioeconomic evidence in clinical practice guidelines 2002* (www.nhmrc.gov.au/publications/synopses/cp89syn.htm)

Further information on NHMRC standards for guideline development is provided in:

- *NHMRC Standards and procedures for externally developed guidelines* (Updated February 2005) (www.nhmrc.gov.au/publications/synopses/nh56syn.htm)
- NHMRC additional levels of evidence and grades for recommendations for developers of guidelines: Pilot Program 2005–2006 (www.nhmrc.gov.au/publications/_files/levels_grades05.pdf)
- *NHMRC Guidelines Assessment Register (GAR) minimum requirements and protocols* (2005) (see Section 4.0).

6 Further down the track – Beyond the recommendations

Expert working groups may be asked to provide input to the following:

A. Public consultation

Once the draft guidelines have been developed, they will undergo a public consultation process where individuals and organisations will be encouraged to review the guidelines and provide comments and suggestions for improvement.

The Expert Working Group for each of the guideline chapters may be asked to respond to questions or comments raised in the public consultation. This may involve changes to the guideline document.

B. A plan for the dissemination, implementation and evaluation of the guidelines

Another requirement for the approval of a set of NHMRC clinical practice guidelines is the development of a dissemination, implementation and evaluation plan. The Expert working groups will be asked for their advice on:

- How to make the guidelines accessible
- Informing the target audience of the guidelines' availability
- Appropriate strategies for their implementation e.g.
 - media marketing;
 - the use of opinion leaders and 'champions';
 - endorsement by clinical groups;
 - practice visits from influential experts;
 - education of patients;
 - educational materials;
 - seminars and conferences;
 - reminder systems incorporated in clinicians' daily work;
 - continuing quality assurance and data feedback;
 - local adaptation and incorporation;
 - local involvement in evaluation; and
 - incentives
- The use of information technology
- A systems approach
- Evaluation
 - an assessment of guideline dissemination;
 - an assessment of whether or not clinical practice is moving towards the guidelines' recommendations;
 - an assessment of whether or not health outcomes have changed;
 - an assessment of whether or not the guidelines have contributed to any changes in clinical practice or health outcomes;

- an assessment of the guidelines' impact on consumers' knowledge and understanding; and
- an economic evaluation of the guideline process. (NHMRC, 1998)

C. A strategy for regular review of the guidelines

The NHMRC (1998) specifies that “once the guidelines have been completed, a date and strategy should be set for their revision. The NHMRC recommends that revision take place at least every three to five years and more often where the subject matter or circumstances are prone to rapid change. (NHMRC 1995)

A number of steps are required for revision:

- A multidisciplinary group, from disciplines similar to those of the guideline development group, should assess the guidelines to see whether there is any new evidence that should be incorporated;
- The group should also assess what has been learnt from the evaluation of the dissemination and implementation strategies and incorporate suggested improvements in the further dissemination and implementation of guidelines;
- The group should draw on current practice and experience and on national data that have been informed by the guidelines. Revision should be a coordinated activity, extending beyond the academic literature on clinical practice and health outcomes to incorporate experience, local knowledge and regional and national data; and
- Guidelines should state the date of their development and the anticipated revision date”.

Glossary of terms

Study design glossary (in alphabetic order)

Adapted from NHMRC 2000; Glasziou et al. 2001; Elwood 1998 and printed in NHMRC Additional levels of evidence and grades for recommendations for developers of guidelines: Pilot Program 2005–2006 (www.nhmrc.gov.au/publications/_files/levels_grades05.pdf)

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).

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.

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 i.e. 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 (i.e. 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 i.e. the placebo or control condition. The outcomes from the two interventions are then compared indirectly (i.e. A vs. B and B vs. C, to indirectly compare A vs. C).

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').

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.

Pseudo-randomised controlled trial

The unit of experimentation (e.g. people, a cluster of people¹) 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 random mechanism (such as a coin toss, random number table, computer-generated random numbers) and the outcomes from each group are compared.

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, e.g. 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.

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 (i.e. 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.

References

- AHCPR, *AHCPR Program Note: clinical practice guideline development*. US Department of Health and Human Services, Public Health Service, August 1993
- AO Foundation. Searching the literature to answer a clinical question, 2006 (accessed January 2006 at www.aofoundation.org/portal/wps/portal)
- Birch DW, Eady A, Robertson D, Pauw SD, Tandan V. User's guide to the surgical literature: how to perform a literature search. *Canadian Journal of Surgery* 46(2):136–141, 2003
- Booth A, Fry-Smith AS. *EText on Health Technology Health Care Resources: Developing the research question*. US National Library of Medicine NIH, 2003 (accessed January 2006 at www.nlm.nih.gov/nichsr/ehta/chapter2.html)
- Glanville J. *Literature searching to inform systematic reviews*, Centre for Dissemination and Reviews University of York, 2004 (accessed January 2006 at www.york.ac.uk/inst/crd/03litsearching.pdf)
- Diabetes Australia Consortium. *National Evidence based guidelines for Type 2 diabetes mellitus*, Diabetes Australia, 2001
- NHMRC. *A guide to the development, implementation and evaluation of clinical practice guidelines*. NHMRC, Canberra. 1999
- NHMRC. *How to review the evidence: systematic identification and review of the scientific literature*. NHMRC, Canberra. 2000
- NHMRC. *How to present the evidence for consumers: preparation for consumer publications*. NHMRC, Canberra. 2000
- NHMRC. *How to use the evidence: assessment and application of scientific evidence*. NHMRC, Canberra. 2000
- NHMRC. *How to put the evidence into practice: implementation and dissemination strategies*. NHMRC, Canberra. 2000
- NHMRC. *How to compare the costs and benefits: evaluation of the economic evidence*. NHMRC, Canberra. 2001
- NHMRC. *Using socioeconomic evidence in clinical practice guidelines*. NHMRC, Canberra. 2002
- NHMRC. Standards and procedures for externally developed guidelines (Updated February 2005)

Richardson WS, Wilson MC, Nishikawa et al (1995) The well-built clinical question: a key to evidence based decisions. *ACP Journal Club*, **123** (3): A12–A13.

NHS Thames Valley Health Libraries Network. *The literature review process: recommendations for researchers*. Version 1.0. Issued 2005, Reviewed 2006
(www.oxfordradcliffe.nhs.uk/research/researchers/news/documents/LiteratureSearchingGuidelinesChecklist.pdf) Accessed January 2006

Appendices

Appendix 1. Markers and search terms for socioeconomic position

Marker	Search Terms	Information generated
SEP	socioeconomic position; SEP; socioeconomic status; SES; class; social class; disadvantage; differentials; gradient; socio-demographic; social determinants	SEP/SES generally, together with articles using other markers, but where the specific details of the marker(s) chosen do not appear in the title of abstract
Income	income; poverty; wealth; insurance	Differentials in risk, treatment and outcome related to level of income – often as a proxy for SEP. Note that in many United States (US) studies, insurance status is often used as a proxy for income and employment (see below)
Education	education; educational status; literacy	Differentials in risk, treatment and outcome related to level of education. Search results regarding risk modification will require sorting as the general term ‘education’ will also provide papers on health education programs, as well as papers on differentials related to levels of education.
Occupation	occupation	The relationship between occupation and risk, treatment and outcome. Differentials between classes of occupation, for example, professional versus labour-based occupations.
Employment	employment (status); unemployment; insurance	Differentials in risk, treatment and outcome related to employment – often as a proxy for SEP
Ethnicity	ethnicity; culture; race; Indigenous; Aboriginal	Differentials in risk, treatment and outcome related to race/ethnicity. Mostly literature on African American versus Caucasian versus Hispanic differentials in the US. Use Indigenous OR Aboriginal AND Australian to specifically search for literature related to Aboriginal and Torres Strait Islander peoples.
Housing	housing; tenant; tenure	The relationship between housing and risk, treatment and outcome. Level of public versus private housing, levels of vacant housing, overcrowding and housing turnover are used as a proxy for socioeconomic position of neighbourhoods, rather than individuals.
Area of Residence	urban; rural; remote; geographical; regional variation; urban differential	Differentials in risk, treatment and outcomes related to area or residence – usually the urban/rural differential.
Lifestyle	smoking; married; nutrition; physical activity; exercise; alcohol; tobacco; drugs; overweight; obese	Lifestyle, usually considered as risk factors.
Gender	gender; sex; sexuality	Differences in risk, treatment and outcomes between males and females.

1) <u>Resource Checklist</u> (continued)	b) Other core resources:	Searched	N/A	Unavail-able
<p>Level 2: Recommended resources Some of these resources may help to locate unpublished literature including theses & conference proceedings.</p> <p>They are useful if you need a greater level of confidence that you are not duplicating other research. However, access to some of these sources may be limited.</p> <p>* Resources marked with a * are not available through the NHS core content collection but may be available through a University Library.</p>	<p>LEVEL 2 : RECOMMENDED RESOURCES</p> <ul style="list-style-type: none"> • Bandolier • BioMed Central • Clinical Evidence • Evidence-based gateways e.g. BIOME (Includes OMNI & NMAP); TRIP database • Drug information <ul style="list-style-type: none"> a) Drug and Therapeutics Bulletin b) National electronic Library for Medicines c) National Prescribing Centre (including MeReC bulletins) d) UK Medicines Information e) Pharmaceutical and device manufacturers 			
<p>Level 3: Additional resources These resources will further increase the comprehensiveness of your search. N.B. you may want to weigh up the likelihood of locating significantly new information before investing the time needed to explore these methods.</p>	<p>LEVEL 3 : ADDITIONAL RESOURCES</p> <ul style="list-style-type: none"> • British Library's Electronic Table of Contents (Zetoc) – journals conference proceedings • Contacting centres of excellence and experts in the field • Dissertation Abstracts * • Index to Theses* • Library catalogues (to locate theses etc.) • Networked Digital Library of Theses and Dissertations (NDLTD) • References cited by retrieved research papers • Science Citation Index (SCI) * • Social Sciences Citation Index (SSCI) * • System for Information on Grey Literature in Europe (SIGLE)* • Web search engines, e.g. Google (contact your librarian for advice in focusing and evaluating the quality of your results) • World Health Organisation Regional Databases 			

