Development of Clinical Practice Guidelines using Cancer Council Australia’s Cancer Guidelines Wiki

Handbook
2012

for section authors and the guideline working party

Clinical Guidelines Network
Cancer Council Australia
This handbook is an update of the original 2006 document “Development of Clinical Practice Guidelines for the Management of Cutaneous Melanoma and Melanoma in special sites” that 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,

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.

This 2012 revision was prepared by the Clinical Guidelines Network team:

Professor Ian Olver  
CEO, Cancer Council Australia

Christine Vuletich  
Manager, Clinical Guidelines Network, Cancer Council Australia

Jutta von Dincklage  
Project Manager, Wiki Development, Cancer Council Australia

Laura Holliday  
Project Officer, Clinical Guidelines Network, Cancer Council Australia

Alisha Dorrigan  
Project Officer, Clinical Guidelines Network, Cancer Council Australia

Alice Winter-Irving  
Project Officer, Clinical Guidelines Network, Cancer Council Australia

August 2012
Contents

Introduction.................................................................................................................................................. 4

Steps in the preparation of Cancer Council Australia clinical practice guidelines:
Flowchart summary .......................................................................................................................................... 5

1 Establishment of working party and guideline objectives ................................................................. 6

2 Developing key clinical questions........................................................................................................ 6

3 Developing the search strategy and completing the literature search ............................................. 8

4 Reviewing and summarising the literature .......................................................................................... 12

5 Assessing the body of evidence, writing the clinical question content and formulating the evidence statements and recommendations ................................................................. 34

6 Uploading guideline content to Cancer Council Australia Cancer Guidelines Wiki.. 39

7 Public consultation and guideline dissemination.................................................................................. 39

8 Incorporating new evidence and updating guideline content ............................................................ 40

Appendix 1 NHMRC guideline development publications .................................................................. 42

References .................................................................................................................................................. 43
Introduction

Cancer Council Australia aims to produce concise, clinically relevant and up-to-date electronic guidelines. This online handbook illustrates the steps in the development of Cancer Council Australia’s web-based clinical practice guidelines. It provides information to assist working party members and staff members to develop concise clinical questions, construct sound search strategies, systematically search the literature, critically appraise, summarise the evidence and formulate guideline recommendations. It is also a resource for all stakeholders to find out details about our clinical practice guideline development methodology.

Cancer Council Australia’s electronic guidelines are based on the NHMRC guideline requirements, designated standards of quality, process and grading system for recommendations.

The major steps in the Cancer Council Australia guideline development process are outlined in Flow Chart 1 on the following page.
Flowchart 1. Steps in the development of Cancer Council Australia’s clinical practice guidelines

1. Establish a working party and guideline objectives
   Joint responsibility of CCA and working party
   - Establish a small management committee to oversee guideline production. Multidisciplinary members are invited to join the greater working party. Specify the purpose, scope, target audience and timelines for the guidelines.

2. Develop clinical questions
   Responsibility of working party
   - Structure the clinical questions according to PICO. Add limits and exclusion criteria to the questions.

3. Develop a search strategy and search the literature
   Responsibility of CCA
   - Translate each clinical question into a search strategy. Conduct pilot searches and communicate with working party to refine the search strategy. Carry out the planned strategy and document the literature search. Relevant articles are uploaded to the CCA Cancer Guidelines Wiki, full text articles are obtained and sent to working party authors.

4. Critically appraise and summarise the literature
   Responsibility of working party
   - Assess the evidence in each article according to NHMRC dimensions of evidence criteria using the CCA Cancer Guidelines Wiki critical appraisal form.

5. Assess the body of evidence, formulate recommendations and write guideline chapters
   Responsibility of working party
   - Write guideline text, evidence summaries, recommendations and/or practice points in MS Word. Grade recommendations using the body of evidence assessment matrix.

6. Upload of question content to CCA Cancer Guidelines Wiki and WP internal review
   Responsibility of CCA and working party
   - CCA team uploads guideline content with references inserted to the CCA Cancer Guidelines Wiki. WP members’ internal review of the question content and recommendations in preparation for release of draft guidelines for public consultation process.

7. Public consultation and guideline dissemination
   Responsibility of CCA, invited colleges, societies and greater public
   - Guideline draft content undergoes an initial public consultation period of 30 days to allow relevant colleges, societies, individuals and the public to make submissions. After 30 days comments are reviewed by the working party authors and guideline content is updated. Guidelines are available online, advertised through various media channels and dissemination is tracked via Google analytics. A spaced education program is currently being developed to ensure guideline implementation and uptake.

8. Incorporate new evidence and update chapters
   Joint responsibility of CCA and working party
   - New PubMed and Embase evidence is constantly being screened by CCA as well as the literature alerts submitted via the Cancer Guidelines Wiki by stakeholders and contributors. Relevant articles are uploaded to the CCA Cancer Guidelines Wiki for critical appraisal and incorporation into guideline content by working party authors. Annual WP meeting held to review all changes made by authors.
1 Establishment of working party and guideline objectives
Joint responsibility of Cancer Council Australia and the working party

A. Establishment of working party to develop the guidelines

A decision is made to develop or revise a clinical practice guideline. Cancer Council Australia is often commissioned by external health and government bodies who contribute funding for the project.

A small guideline management committee is established to oversee the production of the guideline. This consists of Cancer Council Australia staff (CEO, Cancer Council Australia; Manager, Clinical Guidelines Network, Project Officer, Clinical Guidelines Network; Project Manager, Wiki Development) and a selected group of health professionals. Other clinical experts and consumer representatives are invited to join the greater multidisciplinary guideline working party. A member is nominated to chair the guideline project.

B. Determining the purpose, scope, target audience and timelines for the guidelines

The guideline working party determines the purpose, scope, target audience and expected timelines for completion for the guidelines. For example:

- What are the guidelines aiming to address? E.g. epidemiology, prevention, diagnosis, treatment, multidisciplinary care, palliative care, follow-up or cost effectiveness.
- Will the literature searches be limited to high level evidence only (meta-analyses and randomised controlled trials), published studies or include grey literature also? Will there be a lower date limit?
- Who is the target audience for the guidelines? E.g. GPs, surgeons, medical oncologists, radiation oncologists, other medical specialists, nurses or consumers.
- When are the guidelines expected to be completed? What are the expected completion dates for each stage? E.g. literature searching, critical appraisal of the literature, writing guideline chapters and public consultation.

2 Developing key clinical questions
Responsibility of the working party

A. Identification and classification of the major issues to be addressed and structuring these as clinical questions according to the PICO formula

The first step in evidence-based medicine is to formulate focussed and answerable clinical questions. A question must be specific and concrete in order to be searchable in literature databases and capable of being answered after critical appraisal of the literature.1 Cancer Council Australia finds the ‘PICO’ approach helpful in formulating focused clinical questions.

Patients/Population/Problem
What type of patient or population are you interested in? Considerations for defining the patient/population and their characteristics include:
- diagnosis
- cancer stage
- previous treatment
- demographic variables such as gender and age
**Intervention**
Define the intervention. For example:
- New drug
- Surgery or surgical technique
- Radiotherapy
- Diagnostic test

**Comparator**
Define the control group (if applicable) to which the intervention group will be compared. For example:
- Placebo
- Reference standard

**Outcome**
What outcomes are important to the clinical question? For example:
- Survival
- Disease free survival
- Remission
- Complications
- Repeat surgery
- Quality of life
- Toxicity
- Reduction of symptoms
- Patient reported outcomes

**Examples of questions broken down into their PICO components**

Is minimally invasive lobectomy as effective as open lobectomy for treatment of operable stage I non-small-cell lung cancer in adults?

**Population:** Adult patients with stage I non-small-cell lung cancer
**Intervention:** Laparoscopic lobectomy
**Comparison:** Open lobectomy
**Outcomes:** Safety, efficiency, costs, patient satisfaction, quality of life

What is the clinical benefit of stereotactic radiotherapy/radiosurgery for brain metastasis from non-small-cell lung cancer compared with conventional radiotherapy?

**Population:** Adult patients with brain metastasis from non-small-cell lung cancer
**Intervention:** Stereotactic radiotherapy/radiosurgery
**Control:** Conventional radiotherapy
**Outcomes:** Survival, reduction in symptoms, quality of life, toxicity

**B. Addition of limits and exclusion criteria to the clinical questions**

Limits and search inclusion/exclusion criteria are essential for ensuring that articles retrieved by the literature search are relevant and help answer the clinical question. The literature searcher and question authors define what inclusion/exclusion criteria apply for each clinical question. For example:
- Methodology – limit the search to the highest levels of evidence available (e.g meta-analyses, systematic reviews and randomised controlled trials), stipulate a minimum number of patients and/or arms in a trial for inclusion, exclude low level evidence studies (non-systematic review, conference abstracts, case series, case studies, commentary)
- Language - usually limited to articles written in English
• Date – will there be a lower date limit?

3 Developing the search strategy and completing the literature search
Responsibility of Cancer Council Australia

A Translation of each clinical question into a search strategy
A search strategy is constructed from the components of the clinical question (i.e. PICO) along with the study designs and limits being considered. The steps are as follows:

1. The individual search components (population, intervention, control, outcome, methodology and limits) are specified

2. Search terms for each concept are identified. For each key word synonyms, abbreviations, related terms, differences in spelling, old and new terminology, brand and generic names, lay and medical terminology are considered. Index terms unique to each database are identified e.g. Medical Subject Headings (MeSH) terms for Medline and PubMed and Emtree terms for Embase. When there is no adequate index term a combination of textwords is used to cover this concept.

3. Search terms within each component (e.g. intervention) are combined using the Boolean operator ‘OR’

4. Component sets are combined using the Boolean operator ‘AND’ (i.e. search terms for population AND search terms for intervention AND search terms for comparison AND search terms for outcomes and search terms for methodology and limits)

The NHS example below shows how words are combined within each column using the Boolean operator ‘OR’ and words across the columns using the Boolean operator ‘AND’.

![Example](image)

B  Pilot searches and refining the literature search strategy

Some form of preliminary searching (scoping/pilot searching) is usually required to validate the initial clinical question. Pilot searches will indicate:

- How much literature has been published on this topic?
- What is the quality of the literature? Is there high level evidence available, or will the methodology filter need to be expanded?
- What are the best search terms to use for this topic? (indexing terms, textwords or a combination)
- Which databases are likely to provide the highest yield of relevant items?

In the process of carrying out pilot searches it may be necessary for the CCA Project Officer to seek advice from the relevant working party author in order to:

- Clarify the interpretation of the clinical question
- Identify relevant index terms, text words and synonyms
- Ensure the search is on topic and adequately answering the clinical question
- Justify the inclusion/exclusion of certain PICO components
- Restrict a sensitive search to yield fewer articles e.g. by limiting the search to high level evidence only
- Expand a specific search to yield more articles e.g. by removing one of the components of the search such as removing the primary cancer filter from a clinical question concerned with metastasis

The results of these pilot searches are used to help tailor the final search. The final search strategy reflects the amended search terms and revised inclusion and exclusion criteria and should retrieve the relevant items from scoping searches as a matter of course. The search strategy is fully documented on the wiki platform.

C. Carrying out the planned search strategy and documenting the literature search

The CCA Project Officer records the following details for each database that is searched:

- Electronic databases searched (e.g. Embase)
- The database provider (e.g. Elsevier)
- Search terms used
- Search inclusion/exclusion criteria
- Period searched (dates covered by the search)
- List of relevant articles retrieved by each search (see also D Selection, sort and dissemination of the literature search results)
- Reasons for exclusion of irrelevant articles
- Type of search (initial literature search, update alert [see also section on updates])
- PubMed is CCA’s primary database because abstracts are publicly available and free of charge. Therefore PubMed is searched first, followed by Embase, The Cochrane Library then other relevant databases.

Example of how one database literature search is being recorded in the wiki:

**Clinical question**

Clinical question is minimally invasive lobectomy as effective as open lobectomy for treatment of operable stage I NSCLC?

**Database**

PubMed

**Search Command**

```plaintext
```

**Search Type**

Initial search

**Date**

2011-02-01

**Number of articles retrieved**

123

**Number of irrelevant articles**

102

**Number of relevant articles**

21

**Reasons for exclusion**

- Nonsystematic reviews
- Segmentectomy/limited resection compared to lobectomy
- VATS lobectomy alone with no comparator
- VATS lobectomy converted to thoracotomy
- Adjunctive treatment
- Irrelevant

**Relevant Articles**

- Citation: Cho S, Do YW, Lee EB 2011
- Citation: Dentinger CE, Fernandez F, Meyers BF, Pratt W, Zoled J, Patterson GA et al. 2010
- Citation: Scott WJ, Matesotti RS, Egleston BL, Oseni S, Flathearty JF 2010
- Citation: Scott WJ, Allen MS, Darling G, Meyers B, Decker PA, Putnam JB et al. 2010
- Citation: Handy JR Jr, Asapth JH, Dowville EC, Ott GY, Grunkemeier GL, Wu Y 2010
- Citation: Yang X, Wang S, Qu J 2009
- Citation: Flores RM, Ikekezu U, Rzak N, Dycoco J, Bains MS, Downey RJ et al. 2011
D Selecting, sorting and disseminating the literature search results

The Cancer Council Australia Project Officer screens the literature results for a database search for relevance to the clinical question and completes the following steps:

Reviews titles and abstracts (and full text if needed) from the search against the exclusion criteria
- Obtains full text articles through subscriptions to Sydney University e-Library or the Clinical Information Access Wiki (CIAP). Occasionally neither source has the full text article available for free and the article must be purchased

- Sorts the articles by study design e.g. systematic review, randomised controlled trial, cohort, case control or case series and assemble this information in an Excel spreadsheet along with the citation and whether or not a full text article is available

- Sends the link of the search yield and the full text articles to the working party member responsible for that question

A summary of each literature search can be generated interactively per question. It can be accessed in the appendix on each clinical question page. See an example of a literature search summary table below:

<table>
<thead>
<tr>
<th>Database</th>
<th>Date</th>
<th>Retrieved</th>
<th>Excluded</th>
<th>Reasons for exclusion</th>
<th>Relevant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian Medical Association</td>
<td>2011-03-02</td>
<td>6</td>
<td>6</td>
<td>Recurrent NSCLC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inoperable</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Irrelevant</td>
<td></td>
</tr>
<tr>
<td>Cochrane Library</td>
<td>2011-03-02</td>
<td>17</td>
<td>16</td>
<td>Surgery alone</td>
<td>[1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unresectable NSCLC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Palliative radiotherapy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Small cell lung cancer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Irrelevant</td>
<td></td>
</tr>
<tr>
<td>Embase</td>
<td>2011-03-02</td>
<td>217</td>
<td>215</td>
<td>Duplicate studies</td>
<td>[2] [3]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nonsystematic reviews</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inoperable NSCLC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Postoperative radiotherapy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Induction radiotherapy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Surgery vs radiotherapy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mixed stage study</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With less than 5 stage III patients</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Retrospective reviews</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prognostic factors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Phase III clinical</td>
<td></td>
</tr>
</tbody>
</table>
4 Reviewing and summarise the evidence
Responsibility of the working party

A Review each article, summarise the key features and assess the evidence

Each relevant article is critically appraised by the relevant working party author using the CCA Cancer Guidelines Wiki critical appraisal form. This form has been adapted from the NHMRC dimensions of evidence criteria to ensure that critical assessment of the literature is succinct and straightforward while still maintaining academic rigor. The critical appraisal form is used to record the key features of an article (study design, number of patients, inclusion of an economic evaluation) and assess the evidence (level, quality, relevance of the evidence and size of the effect).

Please note: If the authors becomes aware of an important article that was missed by the Cancer Council Australia literature search, s/he notifies us of the missing article on the question page by using the “Submit new evidence” button and following the prompts. The submitted article will be screened by the Project Officer to ensure that it meets the inclusion criteria for the clinical question (on 02 8063 4142 or laura.holliday@cancer.org.au). If the criteria are fulfilled, a critical appraisal will be assigned and added to the body of evidence for that clinical question.
Critically appraise an article: Critical appraisal: Bradley JD, Paulus R, Graham MV, Ettinger DS, Johnstone DW, Pilepich MV et al. 2005 2

Please critically appraise this article using the form below


Step 1: Decide if the article is relevant to the clinical question

Is this article relevant to the clinical question?  

- If YES —> please continue with step 2
- If NO —> please skip the critical appraisal exclude the article in step 4

Step 2: Background information

Please provide the following background details for the article

Study design:  

Total number of patients:  
Please enter the total number of patients or if the study assesses numerous stages please enter the number of patients in the relevant stage.

Does the article include an economic evaluation?  

Step 3: Assess the evidence

Please assess the level of the evidence, quality and relevance of the article

Level of evidence:  

Quality of the evidence:  
Reasons for decision:

Relevance of the evidence:

Reasons for decision:

Size of the effect:

Reasons for decision:

Step 4: Include or exclude article

Please decide whether to include or exclude the article in your topic/clinical question

Final step: Enter appraisal status and save your appraisal

Once you have completed the critical appraisal please change the status of the box below from "pending" to "completed" and click "save page".

Please note if you are midway through an appraisal you can still save your answers without marking the appraisal as completed.
The question icon in the critical appraisal form link to help material on how to rate the NHMRC dimensions of evidence using the CCA Cancer Guidelines Wiki critical appraisal form (see below).

**Critical appraisal help form: study design glossary**

**Study design categories**

Each study design is grouped into six major categories for critical appraisal and quality assessment purposes. Below is a summary of each category and the study designs that are included within each (please see the glossary of terms for a further individual explanations of each study design). Each main study design category has a different quality appraisal targeted towards the particular study design.

1. **Systematic reviews**
   A literature focussed review that systematically locates, appraises and summarises all relevant evidence pertaining to a defined clinical question. Systematic reviews may also include a meta-analysis that quantitatively combines the results of other studies with similar research hypotheses to evaluate therapeutic effectiveness and/or plan new studies.

   Included studies:
   - Systematic review
   - Meta-analysis

2. **Randomised controlled trials**
   Study participants are allocated to either an intervention or control group using a *proper* randomisation method (such as a random number table or a computer generated random number list). The outcomes from each group are then compared.

   Included studies:
   - Randomised controlled trial
   - Randomised cross-over trial
   - Phase II randomised controlled trial
   - Phase III randomised controlled trial

3. **Pseudo-randomised trials, non-randomised trials and cohort studies**
   Outcomes for groups of people exposed to an intervention, or the factor under study, are compared to outcomes for groups of people not exposed. This includes cohort studies, non-randomised trials and pseudo-randomised trials. Pseudo-randomisation is where patients are allocated to an intervention group or control group using a pseudo-randomisation method such as alternating allocation, allocation based on days of the week or odd/even medical record numbers.

   Included studies:
   - Pseudo-randomised/quasi-randomised controlled trial
   - Non-randomised controlled trial
   - Non-randomised cross-over trial
   - Non-randomised phase II
   - Phase IV clinical trial
   - Cohort study
   - Historical control study
   - Two or more single arm study
   - Adjusted/unadjusted indirect comparison
   - Controlled pre-test and post-test outcome study
   - Interrupted time series with a control group
4. **Case-control studies**
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.

Included studies:
- Case-control study
- Diagnostic case-control study
- Population based case-control study

5. **Case series**
A single group of patients (series) who are exposed to an intervention, risk factor or undergo a clinical test that is under investigation. Case series can be retrospective or prospective and do not have a control or reference study group.

Included studies:
- Case series
- Cross-sectional study
- All or none study
- Single arm phase I clinical trial
- Single arm phase II clinical trial
- Diagnostic test accuracy study
- Interrupted time series (without control)
- Diagnostic yield study

6. **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. the 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 e.g. genetic mutation and certain clinical symptoms) the causal link cannot be proven unless a time dimension is included.

7. **Other**
We have included the category “other” for any study designs that do not fit any of the main categories.
Glossary of terms

Adjusted indirect comparisons
An indirect comparison where two randomised controlled trials compare different interventions to the same control (e.g. placebo). The outcomes from the two interventions are then compared indirectly (utilise A vs. B and B vs. C to determine A vs. C with statistical adjustment for B).

All or none
All or none of a series of people (case series) with the risk factor(s) experience the outcome. The data should relate to an unselected or representative case series which provides an unbiased representation of the prognostic effect. 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 smallpox after large scale vaccination. This is a rare situation.

Case series
A single group of people exposed to the intervention.

Case series with post-test outcomes
A single group of people are exposed to the intervention and only outcomes after the intervention are recorded so no comparisons can be made.

Case series with pre-test/post-test outcomes
A single group of people are exposed to the intervention and outcomes before the intervention (pre-test) are compared with outcomes after the intervention (post-test). Also know as a ‘before and after study’.

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.

Clinical practice guideline
An evidence based or consensus document developed to help health care professionals and patients make decisions about screening, prevention or treatment of a specific health condition.

Clinical trial phase I
Researchers test a new drug or treatment in a small group of people for the first time to evaluate its safety, determine a safe dosage range and identify side effects. The design is usually a single arm study with 12-15 patients in cohorts of three and completed in several months. The prime end point is safety.

Clinical trial phase II
The drug or treatment is given to a larger group of people to see if it is effective and to further evaluate its safety. The prime end point is efficacy. The design can be a single arm study with 40-50 patients usually completed within one year or a randomized phase II design with 50-100 patients in each arm which is not powered to make a direct comparison between the arms but allows candidate treatments to be compared prior to selecting for phase III or simply to ensure that the control arm is similar to historical controls and therefore the results in the new treatment are not biased by the Will Rogers phenomenon (stage migration).

Clinical trial phase III
The drug or treatment is given to large groups of people to confirm its effectiveness, monitor side effects, compare it to commonly used treatments and collect information that will allow the drug or treatment to be used safely. These are usually randomised studies of 100 to 150 in each arm powered to
detect a clinically meaningful difference between arms. The major end point is often the survival difference between the new and the standard arms.

**Clinical trial phase IV**
Studies are done after the drug or treatment has been marketed to gather information on the drug’s effect in various populations and any side effects associated with long term use. A post marketing study for efficacy and late toxicities or other toxicities when available to a less select population than is in the studies.

**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.

**Controlled pre test/post test study**
Pre-test and post-test outcomes for a group of people exposed to the intervention are compared with pre-test and post-test outcomes for the control group of people not exposed to the intervention.

**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. the 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 e.g. genetic mutation and certain clinical symptoms) the causal link cannot be proven unless a time dimension is included.

**Cross-over trial**
All subjects receive one treatment and then switch to the other treatment halfway through the study. Crossover studies are often used to study rare diseases where the lack of subjects would make a conventional trial underpowered.

**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.*

**Diagnostic test accuracy study**
A cross-sectional study to investigate the accuracy of a diagnostic test. A group of patients undergo the new diagnostic test and the reference or gold standard diagnostic test. The level of agreement between the investigated test and the gold standard diagnostic test is reported in terms of the sensitivity and specificity, or the likelihood ratio.

**Historical control study**
Outcomes for a prospectively collected group of people exposed to the intervention (or factor under study) are compared with either: (i) the outcomes of people treated at the same institution prior to the introduction of the intervention (i.e. control group/usual care) or (ii) 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 (or 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 who do not receive the intervention.

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 (or factor under study) is introduced to a group of people and compared (as opposed to being compared to an external control group).

Meta-analysis
A quantitative study that combines the results of other studies with similar research hypotheses (usually clinical trials) synthesising summaries and conclusions to evaluate therapeutic effectiveness and/or plan new studies.

Non-randomised experimental trial
People are allocated to either the intervention group or the control group using a non-random method (such as patient or clinician preference/availability) and the outcomes from each group are compared.

Population based 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 e.g. population based screening studies where test accuracy is assessed on all cases with a random sample of controls.

Prospective cohort study
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
People are allocated to either an intervention 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
Study participants are allocated to either an intervention group or a control group using a random mechanism (such as a coin toss, random number table or computer-generated random numbers) and the outcomes from each group are compared.

Retrospective cohort study
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
A literature focussed review that systematically locates, appraises and summarises all relevant evidence pertaining to a defined health question.

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

Unadjusted indirect comparisons
An indirect comparison where two randomised controlled trials compare different interventions to the same control (e.g. placebo). The outcomes from the two interventions are then compared indirectly (utilise A vs. B and B vs. C to determine A vs. C but without statistical adjustment for B).
<table>
<thead>
<tr>
<th>Level</th>
<th>Intervention</th>
<th>Diagnosis</th>
<th>Prognosis</th>
<th>Aetiology</th>
<th>Screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Meta-analysis or a systematic review of level II studies</td>
<td>Meta-analysis or a systematic review of level II studies</td>
<td>Meta-analysis or a systematic review of level II studies</td>
<td>Meta-analysis or a systematic review of level II studies</td>
<td>Meta-analysis or a systematic review of level II studies</td>
</tr>
<tr>
<td>II</td>
<td>Randomised controlled trial or a phase III/IV clinical trial</td>
<td>Diagnostic test accuracy study or a well designed population based case-control study</td>
<td>Prospective cohort study</td>
<td>Prospective cohort study</td>
<td>Randomised controlled trial or a phase III/IV clinical trial</td>
</tr>
<tr>
<td>III-1</td>
<td>Pseudo-randomised controlled trial or a meta-analysis/systematic review of level III-1 studies</td>
<td>Diagnostic test accuracy study, a well designed population based case-control study or a meta-analysis/systematic review of level III-1 studies</td>
<td>All or none or a meta-analysis/systematic review of level III-1 studies</td>
<td>All or none or a meta-analysis/systematic review of level III-1 studies</td>
<td>Pseudo-randomised controlled trial or a meta-analysis/systematic review of level III-1 studies</td>
</tr>
<tr>
<td>III-2</td>
<td>Comparative study with concurrent controls: • Phase II clinical trial • Non-randomised, experimental trial • Controlled pre test/post test study • Adjusted indirect comparisons • Interrupted time series with a control group • Cohort study • Case-control study or a meta-analysis/systematic review of level III-2 studies</td>
<td>Comparison with reference standard that does not meet the criteria required for level II and III-1 evidence or a meta-analysis/systematic review of level III-2 studies</td>
<td>Analysis of prognostic factors amongst untreated control patients in a randomised controlled trial or a meta-analysis/systematic review of level III-2 studies</td>
<td>Retrospective cohort study or a meta-analysis/systematic review of level III-2 studies</td>
<td>Comparative study with concurrent controls: • Phase II clinical trial • Non-randomised, experimental trial • Cohort study • Case-control study or a meta-analysis/systematic review of level III-2 studies</td>
</tr>
<tr>
<td>III-3</td>
<td>A comparative study without concurrent controls: • Phase I clinical trial • Historical control study • Two or more single arm study • Unadjusted indirect comparisons • Interrupted time series without a parallel control group • or a meta-analysis/systematic review of level III-3 studies</td>
<td>Diagnostic case-control study or a meta-analysis/systematic review of level III-3 studies</td>
<td>Retrospective cohort study or a meta-analysis/systematic review of level III-3 studies</td>
<td>Case-control study or a meta-analysis/systematic review of level III-3 studies</td>
<td>Comparative study without concurrent controls: • Phase I clinical trial • Historical control study • Two or more single arm study or a meta-analysis/systematic review of level III-3 studies</td>
</tr>
<tr>
<td>IV</td>
<td>Case series with either post-test or pre-test/post-test outcomes or a meta-analysis/systematic review of level IV studies</td>
<td>Study of diagnostic yield (no reference standard) or a meta-analysis/systematic review of level IV studies</td>
<td>Case series or cohort study of patients at different stages of disease or a meta-analysis/systematic review of level IV studies</td>
<td>Cross-sectional study or a meta-analysis/systematic review of level IV studies</td>
<td>Case series or a meta-analysis/systematic review of level IV studies</td>
</tr>
</tbody>
</table>

Critical appraisal help form: level of evidence criteria according to clinical question

---

1. Meta-analysis
2. Systematic review
3. RCT
4. Case-control study
5. Cohort study
6. Case series
7. Pretest/post-test study
8. Before/after study
9. Cross-sectional study
10. Historical control
11. Diagnostic test accuracy
Explanatory notes

1 Definitions of these study designs are provided on pages 7-8 of National Health and Medical Research Council. How to use the evidence: assessment and application of scientific evidence. Canberra: NHMRC, 2000. 

2 These levels of evidence apply only to studies of assessing the accuracy of diagnostic or screening tests. To assess the overall effectiveness of a diagnostic test there also needs to be a consideration of the impact of the test on patient management and health outcomes. The evidence hierarchy given in the ‘Intervention’ column should be used to assess the impact of a diagnostic test on health outcomes relative to an existing method of diagnosis/comparator test(s). The evidence hierarchy given in the ‘Screening’ column should be used to assess the impact of a screening test on health outcomes relative to no screening or opportunistic screening.

3 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 (e.g. cannot allocate groups to a potential harmful exposure, such as nuclear radiation), then the ‘Aetiology’ hierarchy of evidence should be utilised.

4 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. Systematic reviews of level II evidence provide more data than the individual studies and any meta-analyses will increase the precision of the overall results, reducing the likelihood that the results are affected by chance. Systematic reviews of lower level evidence present results of likely poor internal validity and thus are rated on the likelihood that the results have been affected by bias, rather than whether the systematic review itself is of good quality. Systematic review quality should be assessed separately. A systematic review should consist of at least two studies. In systematic reviews that include different study designs, the overall level of evidence should relate to each individual outcome/result, as different studies (and study designs) might contribute to each different outcome.

5 For a diagnostic test accuracy study with an independent, blinded comparison with a valid reference standard, 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.

6 Well-designed population based case-control studies (e.g. 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 or spectrum effect because the spectrum of study participants will not be representative of patients seen in practice.

7 At study inception the cohort is either non-diseased or all at the same stage of the disease. A randomised controlled trial with persons either non-diseased or at the same stage of the disease in both arms of the trial would also meet the criterion for this level of evidence.

8 All or none of the people with the risk factor(s) experience the outcome; and the data arises from an unselected or representative case series which provides an unbiased representation of the prognostic effect. 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.

9 This also includes controlled before-and-after (pre-test/post-test) studies, as well as adjusted indirect comparisons (i.e. utilise A vs B and B vs C, to determine A vs C with statistical adjustment for B).

10 Comparing single arm studies i.e. case series from two studies. This would also include unadjusted indirect comparisons (i.e. utilise A vs B and B vs C, to determine A vs C but where there is no statistical adjustment for B).

11 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.

Note A: Assessment of comparative harms/safety should occur according to the hierarchy presented for each of the clinical questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms (and other outcomes) are rare and cannot feasibly be captured within randomised controlled trials, in which case lower levels of evidence may be the only type of evidence that is practically achievable; 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 B: When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding clinical question e.g. level II intervention evidence; level IV diagnostic evidence; level III-2 prognostic evidence.

Note C: Each individual study that is attributed a “level of evidence” should be rigorously appraised using validated or commonly used checklists or appraisal tools to ensure that factors other than study design have not affected the validity of the results.
Critical appraisal – quality rating

Each main study design category (1. Systematic reviews, 2. Randomised controlled trials, 3. Pseudo-randomised trials, 4. non-randomised trials and cohort studies, 5. Case-control studies, 6. Case series, 7. Cross-sectional study) requires to complete a quality appraisal specifically designed to assess the study quality for that particular study design.

The system will generate a suggestion on what the overall quality rating (high, medium, low) should be according to a specific algorithm.

The appraiser then selects the overall quality rating for the study in the appraisal form

```
Overall quality assessment:
Based on the answers you have given, the recommended evidence quality rating is Low

Low [ ] Medium [X] High [ ]
```

If needed, the free text field allows entry of additional comments in regards to the quality rating.

The quality appraisal forms for each study design category and the key for the overall quality recommendation, that are used in the online form are listed below for your information.
Quality assessment form: Systematic reviews and meta-analysis

1. Studies included in the systematic review or meta-analysis
   a) Was an adequate search strategy used?
      2 Very thorough – included appropriate search terms and databases
      1 Adequate – search terms and/or choice of databases could have been improved upon
      0 No or not described
   b) Were the inclusion criteria appropriate and applied in an unbiased way?
      2 Yes – pre-specified inclusion criteria applied independently by two people
      1 Adequate – inclusion criteria were pre-specified and applied by one person
      0 No – inclusion was decided in an arbitrary fashion or not described

2. Were the studies assessed for quality (relating to the minimisation of biases)?
   2 Yes – appropriate quality issues were assessed independently by two people
   1 Adequate – some problems with quality issues or assessed by one person only
   0 No – inappropriate, no quality assessment undertaken or not described

3. Were the characteristics and results of individual studies appropriately
   2 Yes - summary descriptive tables of subjects, intervention, outcomes etc are
   provided and estimates of treatment effect displayed
   1 Adequate – more information would be desirable
   0 No

The following questions are only relevant for systematic reviews that pooled data

4. Were the methods used for pooling the data appropriate?
   2 Yes
   0 No

5. Were sources of heterogeneity explored?
   2 Yes
   1 Some attempt was made
   0 No
   N/A No heterogeneity

Key to overall quality rating

High quality: a review that received 2 for the first three questions
Medium quality: a review that received 1 and 2 for the first three questions
Low quality: a review that received 0 for any of the first three questions
### Quality assessment form: Randomised Controlled Trials

1. **Was the study described as randomised?**
   - Yes - continue
   - No - end review

2. **Was the trial double-blinded?**
   - 2 = Reasonably certain double-blinded (eg identical placebo, active placebo, double-dummy, no revealing side-effects).
   - 1 = Indicates double-blinded but not sure (eg method of blinding inappropriate, tablet vs injection with no double-dummy, different treatment schedules, side-effects may unblind single-blinded (eg outcomes assessed blind, objective outcomes, no revealing side-effects).
   - 0 = Outcomes not blinded, substantial side-effects, or not reported.

3. **Concealment of treatment allocation schedule**
   - 2 = Adequately concealed (eg central randomisation, numbered or coded bottles, drugs prepared by pharmacy).
   - 1 = Inadequately concealed (eg numbered envelopes, sealed envelopes, alternation, medical record number, date of birth).
   - 0 = No concealment or unclear (eg no approach described, open randomisation lists, person doing recruitment to toss a coin).

4. **Inclusion of all randomised participants in analysis (ie intention-to-treat anal.)**
   - 2 = No exclusions or survival analysis used with all subjects included (note: follow-up may not be complete but balanced between the comparison groups).
   - 1 = Exclusions not likely to cause bias (some incomplete follow-up but balanced between comparison groups + survival analysis not used).
   - 0 = Too many exclusions, differential loss in comparison groups, or not reported.

5. **Generation of allocation sequences**
   - 1 = Adequate (eg random number table, computer random number generator, coin tossing, card shuffling).
   - 0 = Inadequate or not reported.

---

### Key to overall quality rating

- **High quality:** a review that received 2 for three main criteria (double-blinding, concealment of treatment allocation schedule, inclusion of all randomised participants in analysis (i.e. ITT))
- **Medium quality:** Received 2 and 1 for all three main criteria
- **Low quality:** Received 0 for all three criteria or 0 and 1 for all three main questions or received 0 for any of the three criteria
Quasi-experimental and cohort studies

1. Subject Selection
   (a) “New technology” group
      2. Representative of eligible patients
          Consecutive or random sample (e.g. states all patients recruited in given time frame) and
          In the case of patient selection of technology, all offered option - those who accepted formed the new technology group.
          or
          In the case of surgeon selection of technology, all patients with a particular surgeon, at a particular hospital or in a given time frame received the technology.
   1. Selected group
      Debatable whether group is representative (e.g. consecutive sample but extensive exclusion criteria)
   0. Highly selected or not described
      Selection at surgeon’s discretion (regardless of whether sample consecutive), unclear how group was selected, or not described.

   (b) Comparison group
      2. Representative of eligible patients
          Consecutive or random sample (e.g. states all patients recruited in given time frame), from same population as new technology group, and would be eligible for new technology.
          and
          In the case of patient selection of technology, all offered option - those who declined formed control group.
          or
          In the case of surgeon selection of technology, all patients with a particular surgeon, at a particular hospital or in a given time frame did not receive the technology.
   1. Selected group
      Matched with new technology group for baseline characteristics either prospectively or using historical controls, or debatable whether group is representative (e.g. consecutive sample but extensive exclusion criteria)
   0. Highly selected or not described
      Selected at surgeon’s discretion or patients not eligible for technology (e.g. technology contraindicated) (regardless of whether sample consecutive), unclear how group was selected, or not described.

2. Comparability of groups on demographic characteristics and clinical features
   2. Comparable
      Groups closely matched - comparable on age, extent of disease (e.g. number of bone metastases sites), stage of illness, performance status.
   1. Not comparable but adjusted analysis used
      Groups not comparable but adjusted analysis used, groups match on the majority of variables but not all.
   0. Not comparable and not adjusted for differences
      Not reported or not comparable
3. Measurement of outcomes
   (a) Outcome measures blind to technology used?
      2. Yes
         States outcomes were blinded to whether subject was in technology or control group.
      1. No, but objective measures used
         Measurement of outcomes not likely to be influenced by knowing which group subjects belonged to (e.g. objective outcomes such as mortality)
      0. No or not described
         Issues of blinding not described, subjective measurements used (e.g. QOL, pain, hospital length of stay), blinding not possible (e.g. different treatment schedules)

   (b) Same method of measurement used across comparison groups?
      2. Yes
         Concurrent controls, all subjects treated during the same time period.
      0. No or not described
         Controls measured at different times, locations, personnel, to technology group (e.g. historical controls, controls at different hospital to technology group).

4. Completeness of follow-up
   Follow-up complete and all patients included in the analysis?
      2. Yes
         (follow-up > 95%) or survival analysis using all patients >95% of subjects included or intention to treat.
      1. Reasonable follow-up of all groups (>80%)
         >80% subjects included.
      0. No or not described
         Considerable drop outs, differential drop out in intervention and control groups, or no information provided.

Key to overall quality rating

High quality: a review that received 2 for all criteria (measurement of outcomes excluded)

Medium quality: Received 2 and 1 for all criteria

Low quality: Received 0 for all criteria or 0 and 1 for all criteria
# Quality assessment: case-control study

1. **Subject selection**
   
   (a) **Cases**
   
   1. Selected group (e.g. volunteers)
   0. Highly selected group or not described
   
   (b) **Adequacy of case definition**
   
   1. Taken from medical records, self report without independent validation
   0. Highly selected, inappropriate or not described
   
   (c) **Controls**
   
   1. Selected group (e.g. hospital controls)
   0. Highly selected, inappropriate or not described
   
   (d) **Controls are free of outcome**
   
   1. Yes, based on medical records or self-report
   0. Not blinded or not described

2. **Comparability of groups on demographic characteristics and important potential confounders**
   
   1. Comparable (or matched)
   0. Not comparable and not adjusted for differences

3. **Ascertainment of exposure/treatment**
   
   1. Blinded to case/control status
   0. No or not described

4. **Follow-up complete and all patients included in the analysis?**
   
   1. Complete response
   0. Low response rate (>80%), non-responders not described, differential response in cases/controls, or no details provided

---

### Key to overall quality rating

- **High quality:** a review that received 2 for all criteria
- **Medium quality:** Received 2 and 1 for all criteria
- **Low quality:** Received 0 for all criteria or 0 and 1 for all criteria
Cross-sectional studies

1. Subject Selection
   2. Representative of population of interest
   1. Selected group
   0. Highly selected or not described

2. Comparability of groups analysed on demographic characteristics
   2. Comparable (or matched)
   1. Not comparable but adjusted analysis used
   0. Not comparable and not adjusted for differences

3. Participation rate
   2. High participation rate (>80%) and no important differences between participants and non-participants
   1. Moderate participation rate (65-80%) and no important differences between participants and non-participants
   0. Low participation rate (<65%), important differences between participants and non-participants or not described

Key to overall quality rating

**High quality:** a review that received 2 for three main criteria

**Medium quality:** Received 2 and 1 for all three criteria

**Low quality:** Received 0 for all three criteria or 0 and 1 for all three questions or received 0 for any of the three criteria
Case Series

1. Subject Selection
   2. Representative of eligible patients
      1. Selected group
      0. Highly selected or not described

2. Measurement of outcomes
   Outcome measures blind to pre/post intervention?
      2. Yes
      1. No, but objective measures used
      0. No or not described

4. Completeness of follow-up
   Follow-up complete and all patients included in the analysis?
      2. Yes  (follow-up > 95%)
      1. Reasonable follow-up (>80%)
      0. No or not described

Key to overall quality rating

High quality: a review that received 2 for three main criteria

Medium quality: Received 2 and 1 for all three criteria

Low quality: Received 0 for all three criteria or 0 and 1 for all three questions or received 0 for any of the three criteria
Critical appraisal help form: relevance of the evidence*

Please rate the relevance of the evidence according to the following criteria:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Relevance of the Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Evidence of an effect on patient-relevant clinical outcomes including benefits and harms, quality of life and survival.</td>
</tr>
<tr>
<td>2</td>
<td>Evidence of an effect on a surrogate outcome** that has been shown to be predictive of patient-relevant outcomes for the same intervention.</td>
</tr>
<tr>
<td>3</td>
<td>Evidence of an effect on proven surrogate outcomes but for a different intervention.</td>
</tr>
<tr>
<td>4</td>
<td>Evidence of an effect on proven surrogate outcomes but for a different intervention and population.</td>
</tr>
<tr>
<td>5</td>
<td>Evidence confined to unproven surrogate outcomes.</td>
</tr>
</tbody>
</table>

Points to considering patient-relevant outcomes:

i) 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.

ii) 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 effect on the way the patient feels; otherwise they will not be of interest to the patient or their carers.

iii) All possible outcomes that are of most interest to patients (particularly harms) should be identified and evaluated.

** ‘surrogate outcome’ refers to reasonable indicators of whether there has been some effect e.g. blood pressure measurements or levels of serum cholesterol.


Critical appraisal help form: size of the effect‡

Please evaluate the size of the effect according to the following criteria:

<table>
<thead>
<tr>
<th>Size of the Effect</th>
<th>Clinical Importance of Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>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.</td>
</tr>
<tr>
<td>2</td>
<td>The point estimate of effect is clinically important BUT the confidence interval includes clinically unimportant effects.</td>
</tr>
<tr>
<td>3</td>
<td>The confidence interval does not include any clinically important effects.</td>
</tr>
<tr>
<td>4</td>
<td>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.</td>
</tr>
<tr>
<td>5</td>
<td>Not assessable. Statistical significance (p value or CI) not reported and cannot be calculated from the data.</td>
</tr>
</tbody>
</table>

Points for considering the size of the effect:
i) The size of the effect is important because it relates to the clinical importance of the effect
ii) The size of the effect should be expressed in both relative and absolute terms (ie as relative risks and absolute risk reductions or NNT for a range of baseline risks)
iii) The size of the effect and the certainty with which it is known should both be assessed

Points to considering patient-relevant outcomes:
i) 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
ii) 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 effect on the way the patient feels; otherwise they will not be of interest to the patient or their carers
iii) All possible outcomes that are of most interest to patients (particularly harms) should be identified and evaluated

As a guide where there is no confidence interval:
1: Point estimate is clinically important and p value < 0.01. Assume narrow confidence interval that is unlikely to include clinically unimportant effects
2: Point estimate is clinically important and 0.01 < p value < 0.05. Assume wide confidence interval and therefore may include clinically unimportant results
3: Point estimate is not clinically important and p < 0.05. Assume confidence interval does not include clinically important effects
4: Difference not statistically significant (p > 0.05). CI will be compatible with no effect but may also include clinically important effects or a harmful effect.
5: (extra ranking): Not assessable. Statistical significance (p value or CI) not reported and cannot be calculated from the data

Further detailed information on assessing the strength of the evidence, the relevance of the results and the size of effect is available in the NHMRC publications listed in Appendix 1 and in particular:


C. Body of evidence

For the articles that the author evaluate as relevant and include them as part of the underlying body of evidence for his/her clinical question, a ‘body of evidence’ table is automatically generated. The table summarises and groups the critical appraisals of the literature. This table enables to see at a glance the evidence for a clinical question. It is sortable by various factors such as size of the effect, quality, relevance and level of evidence by clicking the arrow button next to each heading, please see the image below. The authors use this table when they assess the evidence, formulate recommendations and write the clinical question contents (please note that authors do not have to cite every article that they included in the body of evidence in their clinical question text or evidence statements)
The ‘body of evidence’ table can be accessed in the appendices at the bottom of each clinical question page.
5 Assessing the body of evidence, writing the clinical question content and formulating evidence statements and recommendations

Responsibility of the working party

Once the working party members have critically appraised the relevant literature for their assigned clinical questions, the next step for them is to review the evidence and develop the content for the clinical question.

Each clinical question is typically structured in the following way

1) Sections that summarise and review the literature written as narratives. The content is grouped under sub-headings.
2) Evidence summaries including levels of evidence and links to the key references
3) Recommendation(s) including the overall recommendation grade
4) Practice points for key practice points, for which there is no evidence

(insert/link to question template)


Step 1: Assessing the body of evidence

The first step is the assessment of the body of evidence. Every author is asked to rate five key components (evidence base, consistency, clinical impact, generalizability and applicability) of the ‘body of evidence’ for each recommendation.

1. The evidence base*, in terms of the number of studies, level of evidence and quality of studies (risk of bias).
2. The consistency* of the study results.
3. The potential clinical impact* of the proposed recommendation.
4. The generalisability* of the body of evidence to the target population for the guideline.
5. The applicability* of the body of evidence to the Australian healthcare context

(*For further explanation, see NHMRC Definitions of the components of the evidence statement on page 44.)

The first two components give a picture of the internal validity of the study data in support of efficacy (for an intervention), accuracy (for a diagnostic test), or strength of association (for a prognosis or aetiological question). The third component addresses the likely clinical impact of the proposed recommendation. The last two components consider external factors that may influence the effectiveness of the proposed recommendation in practice, in terms of the generalisability of study results to the intended target population for the Guideline and setting of the proposed recommendation, and applicability to the Australian (or other local) health care system.
The components described above should be rated according to the matrix shown in Table Body of evidence assessment matrix.

<table>
<thead>
<tr>
<th>Recommendation Component</th>
<th><strong>Recommended Grade</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>A</strong></td>
</tr>
<tr>
<td>Evidence base**</td>
<td></td>
</tr>
<tr>
<td>one or more level I studies with a low risk of bias or several level II studies with a low risk of bias</td>
<td>one or two level II studies with a low risk of bias or a systematic review/several level III studies with a low risk of bias</td>
</tr>
<tr>
<td>Consistency**</td>
<td></td>
</tr>
<tr>
<td>all studies consistent</td>
<td>most studies consistent and inconsistency may be explained</td>
</tr>
<tr>
<td>Clinical impact</td>
<td>very large</td>
</tr>
<tr>
<td>Generalisability</td>
<td></td>
</tr>
<tr>
<td>population/s studied in body of evidence are the same as the target population for the guideline</td>
<td>population/s studied in the body of evidence are similar to the target population for the guideline</td>
</tr>
<tr>
<td>Applicability</td>
<td></td>
</tr>
<tr>
<td>directly applicable to Australian healthcare context</td>
<td>applicable to Australian healthcare context with few caveats</td>
</tr>
</tbody>
</table>

1 Level of evidence determined from level of evidence criteria
2 If there is only one study, rank this component as ‘not applicable’
3 For example 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. Consequently guideline developers find that the body of evidence supporting a recommendation rarely consists of entirely one rating for all the important components (outlined above). For example, a body of evidence may contain a large number of studies with a low risk of bias and consistent findings, but which are not directly applicable to the target population or Australian healthcare context and have only a limited 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 NHMRC evidence grading system is designed to allow for this mixture of components, while still reflecting the overall body of evidence supporting a guideline recommendation.

Step 2: Writing evidence summary statement(s)

The guideline authors synthesize the evidence relating to each evidence component by developing brief evidence statements. The levels of evidence are indicated and the relevant references are included. The number of articles to cite depends on the available evidence, but ideally will be no more than five articles, as CCA is interested in the highest level evidence available. Therefore, if an author has a mixture of level I, II, III and IV studies for an evidence summary statement, the author is encouraged to discard the level III and IV articles and only include the highest evidence in the evidence summary. For example:
<table>
<thead>
<tr>
<th>Evidence summary</th>
<th>Level</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PORT has an adverse effect on survival in patients with pN0-N1 non-small-cell lung cancer.</td>
<td>I,II-2</td>
<td>1,3,9</td>
</tr>
</tbody>
</table>

**Step 3: Formulating the recommendation**

Based on the evidence assessment and evidence summary statement, the author develops the wording for the recommendation. The recommendation needs to address the specific clinical question and is ideally written as an action statement. The wording of the recommendation should reflect the strength of the body of evidence. Words such as ‘must’ or ‘should’ are used when the evidence underpinning the recommendation is strong, and words such as ‘might’ or ‘could’ are used when the evidence base is weaker. (NHMRC 2000)

The resulting recommendation from the evidence summary example above was:

**Recommendation**

Post-operative radiation therapy is not recommended for patients with pN0-N1 non-small-cell lung cancer because of a potential detrimental effect on survival.

Grade C (see step 3 and 4)

**Step 4: Determining an overall recommendation grade**

The question author then determines the overall recommendation grade (link to grade explanation) by looking at the individual components that he had determined in step 1. **If most components are rated A then the overall recommendation grade will also be A**.

** For a recommendation to be graded A or B, the volume and consistency of evidence must also be graded either A or B!

If a recommendation overall grade is on the cusp, it is not possible to assign two grades. For example, if the components of a recommendation are rated B, A, B, C, A then the clinician will have to choose whether to grade the recommendation as A or B. In this case most clinicians err on the side of caution and allocate the lower grade.

We understand that applying evidence in real clinical situations is not usually straightforward and thus the body of evidence supporting a recommendation rarely consists of the same grade for all five components. Consequently 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.

**Recommendation grade explanation**

Cancer Council Australia uses the NHMRC grading system for recommendations, as it is well recognised and based on an international standard. The recommendation grades are as follows:
### Grade of Recommendation

<table>
<thead>
<tr>
<th>Grade of Recommendation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Body of evidence can be trusted to guide practice</td>
</tr>
<tr>
<td>B</td>
<td>Body of evidence can be trusted to guide practice in most situations</td>
</tr>
<tr>
<td>C</td>
<td>Body of evidence provides some support for recommendation(s) but care should be taken in its application</td>
</tr>
<tr>
<td>D</td>
<td>Body of evidence is weak and recommendation must be applied with caution</td>
</tr>
<tr>
<td>PP (practice point)</td>
<td>Where no good-quality evidence is available but there is consensus among Guideline committee members, consensus-based recommendations are given, these are called practice points</td>
</tr>
</tbody>
</table>

Recommendation grades 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.

### E. Practice point

On occasion clinicians may find that there is an important practical point that they wish to emphasise for which there is no evidence. This is typically an aspect of treatment that is regarded as sound clinical practice that no-one is likely to question. These evidence-free recommendations are called practice points. As there is no evidence supporting a practice point, an accompanying evidence summary is not required.

For example:

**Practice point:**

It is advisable to review unexpected pathology results with the reporting pathologist.

### D. Referencing style

Cancer Council Australia guidelines use the New England Journal of Medicine referencing style. However, when the authors write the clinical question content in Word we ask them to reference using the Harvard format (author, year) as this is much easier for us when we upload references to the CCA Cancer Guidelines Wiki.
Clinical question: What is the recommended treatment approach for the definitive management of patients with good performance status and inoperable stage III non-small-cell lung cancer?

Introduction/background narrative text here

Subheading e.g. Radiation alone versus Combination Chemoradiotherapy

Narrative text here and evidence summary and recommendation(s) and/or practice point(s) where relevant

Practice point:

It is advisable to...

Subheading e.g. Concurrent versus Sequential Therapy

Narrative text here and evidence summary and recommendation(s) and/or practice point(s) where relevant

<table>
<thead>
<tr>
<th>Evidence summary</th>
<th>Level</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, the evidence supports...</td>
<td>I,II</td>
<td>4,6</td>
</tr>
</tbody>
</table>

Recommendation

Patients are best managed...

Grade A

Each recommendation requires an accompanying evidence matrix. Please place an X in the relevant box.

<table>
<thead>
<tr>
<th>Recommendation Component</th>
<th>A Excellent</th>
<th>B Good</th>
<th>C Satisfactory</th>
<th>D Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume of evidence</td>
<td>one or more level I studies with a low risk of bias or several level II studies with a low risk of bias</td>
<td>one or two level II studies with a low risk of bias or a systematic review/several level III studies with a low risk of bias</td>
<td>one or two level III studies with a low risk of bias, or level I or II studies with a moderate risk of bias</td>
<td>level IV studies, or level I to III studies/systematic reviews with a high risk of bias</td>
</tr>
<tr>
<td>Consistency</td>
<td>all studies consistent</td>
<td>most studies consistent and inconsistency may be explained</td>
<td>some inconsistency reflecting genuine uncertainty around clinical question</td>
<td>evidence is inconsistent</td>
</tr>
<tr>
<td>Clinical impact</td>
<td>very large</td>
<td>substantial</td>
<td>moderate</td>
<td>slight or restricted</td>
</tr>
</tbody>
</table>
### Generalisability

<table>
<thead>
<tr>
<th></th>
<th>population/s studied in body of evidence are the same as the target population for the guideline</th>
<th>population/s studied in the body of evidence are similar to the target population for the guideline</th>
<th>population/s studied in body of evidence differ to target population for guideline but it is clinically sensible to apply this evidence to target population</th>
<th>population/s studied in body of evidence differ to target population and hard to judge whether it is sensible to generalise to target population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicability</td>
<td>directly applicable to Australian healthcare context</td>
<td>applicable to Australian healthcare context with few caveats</td>
<td>probably applicable to Australian healthcare context with some caveats</td>
<td>not applicable to Australian healthcare context</td>
</tr>
</tbody>
</table>

### References


### 6 Uploading guideline content to Cancer Council Australia Cancer Guidelines Wiki

**Responsibility of Cancer Council Australia**

When the authors have developed their final draft content, they return it to Cancer Council Australia and it is uploaded to the CCA Cancer Guidelines Wiki. The content is internally reviewed by the working party for a brief period before making the draft content publicly available, in order to ensure that there are no inconsistencies of recommendations between content sections.

### 7 Public consultation and guideline dissemination

**Responsibility of Cancer Council Australia, invited colleges, societies and the greater public**

**A Public consultation**

Once the draft guidelines have been developed and uploaded, they will undergo an initial public consultation period of 30 days. Relevant colleges, societies, organisations and individuals will be notified by email and encouraged to review the guidelines and provide comments and suggestions for improvement. Public commenting has been enabled on the CCA Cancer Guidelines Wiki and public consultation will take place online.

The initial deadline of 30 days is to encourage relevant parties to comment within a predetermined timeframe and to allow the guidelines to be published online while in draft format. After the 30 days and a short time frame to allow for consideration and incorporation of public comments the draft status will be removed.

The guideline working party and section authors are required to respond to questions or comments raised in both the initial and ongoing public consultation online. This may involve changes to the guideline content. A summary of comments can be generated interactively per question and/or for the full guideline. CCA will continue to allow ongoing public commenting and content updating by authors to ensure the guidelines are kept as current as possible.
B Guideline dissemination and implementation

Council Australia’s clinical practice guidelines are made available online internationally via the CCA Clinical Guidelines Wiki. Initial and ongoing public consultation will be advertised and awareness of the guidelines will be achieved via media coverage through multiple outlets. The URL link to the guidelines will be distributed directly to relevant professional and other interested groups via email, print and social media campaigns as well as through meetings, national conferences and other CME events. A significant effort will be made to have the guidelines introduced to senior undergraduate medical students and to encourage the relevant learned colleges (surgeons, radiation oncologists, pathologists and general practitioners) to support the guidelines and to foster their integration into hospital and community practice through resident and registrar education activities.

A web analytics solution has been implemented on our CCA Clinical Guidelines Wiki to track and analyse website traffic and the effectiveness of marketing efforts. Amongst other things, the web analytics solution enables us to track the number of unique visitors separated by user group, how users accessed the site (a direct URL entry obtained through an email or other means, referral from another website, such as Google or other search engines, Cancer Council Australia, Clinical Oncological Society of Australia, Facebook, LinkedIn, www.cancerquestions.com.au), how much average time different visitor groups spent on a particular guideline page, which countries the visitors are from and how the visitors navigated through a guideline/the website. In addition, we are also inviting external reviewers to complete a brief feedback survey. This allows Cancer Council Australia to get an in-depth insight into guideline usage and allows us to evaluate and refine dissemination measures.

Use of Cancer Council Australia guidelines as part of core curriculum in specialty exams will be encouraged. It is recognised that a planned approach is necessary to overcome specific barriers to implementation in particular settings and to identify appropriate incentives to encourage uptake of guideline recommendations. Implementation of the guidelines will require a combination of effective strategies and may include further CME initiatives, the development and promotion of computer-assisted decision aids, electronic decision-support systems and the creation of audit and other clinical tools. For example, Cancer Council Australia is aiming to develop online learning modules to reinforce content knowledge for participants, thus support guideline implementation and uptake. Cancer Council Australia is piloting the development of QStream modules (previously known under Spaced Education), originally developed by Harvard Medical School. QStream programs have shown to improve knowledge acquisition in a number of randomised trials with medical practitioners.1

By allowing guideline stakeholders to comment on guidelines content and submit new evidence on an ongoing basis, Cancer Council Australia is encouraging its stakeholders to engage with the actual guideline content on a long-term basis, which will ensure wide stakeholder consultation and up-to-date guideline content.

8 Incorporating new evidence and updating guideline content
Joint responsibility of Cancer Council Australia and the working party

Cancer Council Australia aims to produce concise, clinically relevant and up-to-date electronic guidelines. Keeping the guidelines current is a collaborative effort between the working party, Cancer Council Australia and the public through submitting comments and suggesting new evidence. The first phase is to develop online guidelines. (see section 1, p6).

Once online guidelines are established, the original PubMed and Embase literature searches are automatically re-run on a regular basis. The new results are emailed through to the CCA Project Officer

1 For more information please visit http://www.spaceded.com/info/research
who screens them for relevance, records the literature search update results and uploads selected articles to the CCA Cancer Guidelines Wiki to be critically appraised and updated by the working party authors in the guideline content. Working party authors receive notifications if there is any new literature awaiting appraisal for their clinical question. The question authors eventually appraise the new evidence and check if this results in a content update.

Members of the public can suggest new evidence by using the “Submit new evidence” feature on every clinical question page. The CCA Project Officer will be notified of public literature submissions, check if they are relevant and meet the inclusion and exclusion criteria. If the literature submitted is not relevant, it will be excluded. If the submission is relevant, a critical appraisal will be assigned and the question author will appraise the literature submission further and determine if the content needs to be updated. The public submitting new evidence are able to track their submission by visiting the permalink to their submission (link to further information).

Authorship of a guideline question is an ongoing commitment. Working party authors will be notified by Cancer Council Australia staff (or they may choose to be notified automatically via an email alert from the CCA Cancer Guidelines Wiki) when someone has commented on their chapter or when there is new literature awaiting critical appraisal and incorporation into their guideline section. Each question has a main author and the main author nominates two co-authors to help him revise and update the guideline content. It is the responsibility of the question author to address submitted comments, appraise new relevant papers and make the required updates to the guideline content. If it is not straightforward how comments and new evidence should be integrated, the main author can use an internal access-restricted online commenting tool and consult with the two nominated co-authors. If required, the whole working party can be consulted.

An annual working party meeting is organised by CCA to revise all changes that were made by individual working party authors to guideline content.
Appendix 1: NHMRC publications which provide further information on guideline development

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


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


References


NHMRC Definitions of the components of the evidence statement

1. Evidence base

The evidence base is assessed in terms of the quantity, level and quality (risk of bias) of the included studies:

- Quantity of evidence reflects the number of the studies that have been included as the evidence base for each guideline (and listed in the evidence summary table or text). The quantity assessment also takes into account the number of patients in relation to the frequency of the outcomes measured (ie the statistical power of the studies). Small, underpowered studies that are otherwise sound may be included in the evidence base if their findings are generally similar — but at least some of the studies cited as evidence must be large enough to detect the size and direction of any effect. Alternatively, the results of the studies could be considered in a meta-analysis to increase the power and statistical precision of the effect estimate.

- Level of evidence reflects the best study types for the specific type of question (see Part B, Table 3). The most appropriate study design to answer each type of clinical question (intervention, diagnostic accuracy, aetiology or prognosis) is level II evidence. Level I studies are systematic reviews of the appropriate level II studies in each case. Study designs that are progressively less robust for answering each type of question are shown at levels III and IV. Systematic reviews of level III and IV studies are ascribed the same level of evidence as the studies included in the review to address each outcome. For example, a systematic review of cohort studies and case series for an intervention question would be given a Level III-2 ranking in the hierarchy, even if the quality of the systematic review was exceptional. The levels of evidence hierarchy is specifically concerned with the risk of bias in the presented results that is related to study design (see Explanatory note 4 to Table 3), whereas the quality of the evidence is assessed separately.

- Quality of evidence reflects how well the studies were conducted in order to eliminate bias, including how the subjects were selected, allocated to groups, managed and followed up and how the study outcomes were measured (see Part B, Dimensions of evidence, and Table 4 for further information).

2. Consistency

The consistency component of the ‘body of evidence’ assesses whether the findings are consistent across the included studies (including across a range of study populations and study designs). It is important to determine whether study results are consistent to ensure that the results are likely to be replicable or only likely to occur under certain conditions. Ideally, for a meta-analysis of randomised studies, there should be a statistical analysis of heterogeneity showing little statistical difference (consistent or homogenous) between the studies. However, given that statistical tests for heterogeneity are underpowered, presentation of an I² statistic, as well as an appraisal of the likely reasons for the differences in results across studies, would be useful. Heterogeneity in the results of studies may be due to differences in the study design, the quality of the studies (risk of bias), the population studied, the definition of the outcome being assessed, as well as many other factors. Non-randomised studies may have larger estimates of effect as a result of the greater bias in such studies; however, such studies may also be important for confirming or questioning results from randomised trials in larger populations that may be more representative of the target population for the proposed guideline.

3. Clinical impact

Clinical impact is a measure of the potential benefit from application of the guideline to a population. Factors that need to be taken into account when estimating clinical impact include:
• the relevance of the evidence to the clinical question, the statistical precision and size of the effect (including clinical importance) of the results in the evidence-base, and the relevance of the effect to the patients, compared with other management options (or none)
• the duration of therapy required to achieve the effect, and
• the balance of risks and benefits (taking into account the size of the patient population concerned).

4. Generalisability

This component covers how well the subjects and settings of the included studies will match those of the Guideline recommendations, specifically the patient population being targeted by the Guideline and the clinical setting where the recommendation will be implemented. Population issues that might influence the relative importance of recommendations include gender, age or ethnicity, baseline risk, or the level of care (eg community or hospital). This is particularly important for evidence from randomised controlled trials (RCTs), as the setting and entry requirements for such trials are generally narrowly based and, therefore, may not be representative of all the patients to whom the recommendation may be applied in practice. Confirmation of RCT evidence by broader-based population studies may be helpful in this regard (see ‘2. Consistency’). Basically, an assessment of generalisability is about determining whether the available body of evidence is answering the clinical question that was asked. In the case of studies of diagnostic accuracy, a number of additional criteria also need to be taken into account, including the stage of the disease (eg early versus advanced), the duration of illness and the prevalence of the disease in the study population as compared to the target population for the guideline.

5. Applicability

This component addresses whether the evidence base is relevant to the Australian health care system generally, or to more local settings for specific recommendations (such as rural areas or cities).

Factors that may reduce the direct application of study findings to the Australian or more local settings include organisational factors (eg availability of trained staff, clinic time, specialised equipment, tests or other resources) and cultural factors (eg attitudes to health issues, including those that may affect compliance with the recommendation).

References

1. Adapted from the Scottish Intercollegiate Guidelines Network (SIGN) guide to using their Considered Judgement Form (available from http://www.sign.ac.uk/guidelines/fulltext/50/annexd.html Accessed 19.10.07)
2. whereas most statistical tests of heterogeneity (eg Cochran’s Q) assess whether heterogeneity exists between studies, I² is a statistic that quantifies how much heterogeneity exists between the studies (see Higgins & Thompson, 2002)