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

Handbook

For section authors and the guideline working party

2014

Clinical Guidelines Network
Cancer Council Australia
This handbook 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  
Product 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

Suggested citation:  

Acknowledgments

The structure of this handbook is based on the 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 in 2006 by Phoebe Holt and Michael Frommer, Sydney Health Projects Group, School of Public Health, University of Sydney.
Tables
Table 1: Literature search summary table ..................................................................................12
Table 2: Body of evidence assessment matrix ..........................................................................18
Table 3: Example of responses to public comments posted online ....................................25

Figures
Figure 1: NHS example of using Boolean operators ............................................................... 9
Figure 2: Literature search record .........................................................................................11
Figure 3: Screenshot overall quality appraisal section ..........................................................14
Figure 4: Body of evidence table ..........................................................................................16
Figure 5: Screenshot user page – link to body of evidence ...................................................17
Figure 6: Screenshot clinical question appendices ...............................................................17
Figure 7: Screenshot completed evidence assessment matrix ...........................................20
Figure 8: Example of format of guideline content on the Wiki ...........................................21
Figure 9: Screenshot clinical question display including buttons .......................................24
Figure 10: User menu ...........................................................................................................24
Figure 11: Screenshot user page Project Officer – Literature search .....................................26
Figure 12: Screenshot screening new literature search results ...........................................27
Figure 13: Screenshot internal comments button ................................................................27
Introduction

Cancer Council Australia (CCA) 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 and 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 Flowchart 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 (WP)

2. Develop clinical questions
   Responsibility of working party

3. Develop a search strategy and search the literature
   Responsibility of CCA

4. Critically appraise and summarise the literature
   Responsibility of working party

5. Assess the body of evidence, formulate recommendations and write guideline content
   Responsibility of working party

6. Upload question content to CCA Cancer Guidelines Wiki and WP internal review
   Responsibility of CCA and working party

7. Public consultation and guideline dissemination
   Responsibility of CCA, invited colleges, societies and greater public

8. Incorporate new evidence and update chapters
   Joint responsibility of CCA and working party

---

Establishment of a small management committee to oversee guideline production. Multidisciplinary members are invited to join the greater working party to specify the purpose, scope, target audience and timelines for the guidelines.

Note: If existing guideline or specific clinical questions of an existing guideline are considered for adaptation, the ADAPTE protocol is followed (see Appendix 1).

The clinical questions are structured according to PICO. Limits and exclusion criteria to the questions are added.

Each clinical question is translated into a search strategy. Pilot studies are conducted and the working party is consulted to refine the search strategy. The planned strategy is carried out and the literature search is documented. Relevant articles are uploaded to the CCA Cancer Guidelines Wiki, and full text articles are obtained and sent to working party authors.

The evidence in each article is assessed according to NHMRC dimensions of evidence criteria using the CCA Cancer Guidelines Wiki critical appraisal form.

Guideline text, evidence summaries, recommendations and/or practice points are written in Microsoft Word. Grade recommendations are made using the body of evidence assessment matrix.

The CCA team uploads the guideline content to the CCA Cancer Guidelines Wiki. WP members internally review question content and recommendations in preparation for release of draft guidelines for public consultation process.

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. An online education program (Qstream) is developed to ensure guideline implementation and uptake.

PubMed and Embase evidence are continually screened by CCA as well as literature alerts submitted via the Cancer Guidelines Wiki by stakeholder and contributors. Relevant articles are uploaded to the CCA Cancer Guidelines Wiki for critical appraisal and incorporation into guideline content by WP authors. An 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

1.1 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 to develop a specific guideline.

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, including the person nominated to chair the guideline project. Other clinical experts and consumer representatives are invited to join the greater multidisciplinary guideline working party. The management committee is responsible for evaluating the conflict of interest statements of each working party member and putting appropriate management strategies in place where applicable (see Appendix 2).

1.2 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) or published studies, or will they also include grey literature? 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.
- Are there any existing high quality guidelines or particular clinical questions of an existing clinical practice guideline that could be assessed for adaption for the Australian healthcare context? (See Appendix 1 for a description of the guideline adaption process.)

2 Developing key clinical questions
Responsibility of the working party

2.1 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 uses the PICO approach to formulate 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

1
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(s)**
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

**2.2 Addition of limits and exclusion criteria to the clinical questions**

Adding limits to the search are essential for ensuring that articles retrieved by the literature search are relevant and help answer the 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)
- **Language**: usually limited to articles written in English
- **Date**: will there be a lower date limit?

The literature searcher and question authors also pre-specify what inclusion/exclusion criteria apply for each clinical question as part of the systematic review protocol (i.e. 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]). This determines which studies are selected for inclusion (see 3.4).
3 Developing the search strategy and completing the literature search
Responsibility of Cancer Council Australia

3.1 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, and 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’.²

Figure 1: NHS example of using Boolean operators

![Diagram of search strategy with Boolean operators and examples of search terms](source)


3.2 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 the topic
- the quality of the literature (e.g. if there is high level evidence available, or if the methodology filter needs to be expanded)
- the best search terms to use for this topic (e.g. indexing terms, text words or a combination)
- the databases 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 doing the systematic literature review 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.

3.3 Carrying out the planned search strategy and documenting the literature search

The CCA Project Officer’s systematic literature review 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.
Figure 2 illustrates how one database literature search is being recorded in the Wiki.

Figure 2: Literature search record

Clinical question

Clinical question: Is minimally invasive lobectomy as effective as open lobectomy for treatment of operable stage 1 NSCLC? Database

PubMed

Search Command


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
- Adjuvant treatment
- Irrelevant

Relevant Articles

- Citation ChOS. Do YW. Lee EB 2011
- Citation Derlinger CE, Fernandez F, Meyers BF, Pratt W, Zoled JB, Patterson GA et al. 2010
- Citation Scott WJ, Matteotti RS, Egleston BL, Oseni s, Flaherty JF 2010
- Citation Scott WJ, Allen MS, Darling G, Meyers B, Decker PA, Putnam JB et al. 2010
- Citation Handy JR, Asaph JW, Douville EC, Ott NY, Grunkemeier GL, Wu Y 2010
- Citation Yang X, Wang S, au J 2009
- Citation Flores RM, Ihekwazu UN, Rikk N, Dyco J, Bains MS, Downey R et al. 2011
3.4 Selecting, sorting and disseminating the literature search results

The Cancer Council Australia Project Officer screens the literature results for each 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 previously defined exclusion criteria
- obtains full text articles through subscriptions to University e-Libraries. Occasionally the full text article is unavailable for free and the article must be purchased
- produces a list of all included studies along with the full citation and whether or not a full text article is available
- sends the literature search results, full text articles and an individualised Wiki help document to the working party member responsible for that question
- organises webinar/teleconferences to teach working party authors how to navigate the Wiki platform and how to use the critical appraisal form to appraise the literature.

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

Table 1: Literature search summary table

<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, inoperable, irrelevant</td>
<td>[1]</td>
</tr>
<tr>
<td>Cochrane Library</td>
<td>2011-03-02</td>
<td>17</td>
<td>16</td>
<td>Surgery alone, unresectable NSCLC, palliative radiotherapy, small cell lung cancer, irrelevant</td>
<td>[1]</td>
</tr>
<tr>
<td>Embase</td>
<td>2011-03-02</td>
<td>217</td>
<td>215</td>
<td>Duplicate studies, nonsystematic reviews, inoperable NSCLC, postoperative radiotherapy, induction radiotherapy, surgery vs. radiotherapy, mixed stage study with less than 5 stage III patients, retrospective reviews, prognostic factors, phase III clinical</td>
<td>[2], [3]</td>
</tr>
</tbody>
</table>
4 Reviewing and summarising the evidence

Responsibility of the working party

4.1 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 (see Figure 3). 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 CCA of the missing article. The submitted article is screened by the Project Officer to ensure that it meets the inclusion criteria for the clinical question. If the criteria are fulfilled, the additional citation is added to the literature search documentation under ‘other sources’, and a critical appraisal is assigned and added to the body of evidence for that clinical question. The Project Officer can be contacted on (02) 8063 4142.
Figure 3: The CCA Cancer Guidelines Wiki critical appraisal form

Critically appraise an article: Critical appraisal: Bradley JD, Paulus R, Graham MV, Ettinger DS, Johnstone DW, Pilepich MV et al.  2005

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: (see appendix 5, 1 Critical appraisal help form: Study design glossary)

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: (see appendix 5)

Quality of the evidence:
The question mark icon in the critical appraisal form links to help material on how to rate the different components of the evidence using the CCA Cancer Guidelines Wiki critical appraisal form. All help forms, including a Glossary of Terms, can be found in Appendix 5.
4.2 Body of evidence

A ‘body of evidence’ table is automatically generated by the system, which includes all articles that the author has evaluated as relevant and included as part of the underlying body of evidence for the particular clinical question.

The table summarises and groups the critical appraisals of the literature. This table enables a person to see at a glance the evidence for a clinical question. It is sortable by various factors, such as size of the effect, and the quality, relevance and level of evidence, by clicking the arrow button next to each heading – see Figure 4 below.

The authors use this table when they assess the evidence, formulate recommendations and write the clinical question content. (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.)

Figure 4: Body of evidence table

<table>
<thead>
<tr>
<th>Critical appraisal</th>
<th>Article</th>
<th>Level of Evidence</th>
<th>Quality of Evidence</th>
<th>Size of Effect</th>
<th>Number of Patients</th>
<th>Relevance of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lester JF, MacBeth F, Toy E, Coles B J</td>
<td>Lester JF, MacBeth F, Toy E, Coles B J</td>
<td>I</td>
<td>High</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Oktawana G, Mackay JA, Evans WK, Ung VC, Lung Cancer Disease Site Group of Cancer Care Ontario’s Program in Evidence-based Care 2006</td>
<td>Oktawana G, Mackay JA, Evans WK, Ung VC, Lung Cancer Disease Site Group of Cancer Care Ontario’s Program in Evidence-based Care 2006</td>
<td>I</td>
<td>High</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

4.2.1 How to access this table

This table can be accessed by the author from his/her user page by clicking ‘body of evidence’ next to the relevant clinical question (see Figure 5).
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 question(s), 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 are 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

See page Figure 8: Example of format of guideline content on the wiki for the template that is completed by each working party author.

5.1 Step 1: Assessing the body of evidence

The first step is the assessment of the body of evidence in terms of evidence quantity, level and quality (risk of bias) by investigating the evidence table and identifying the high level evidence/high quality studies and grouping the
different findings of the studies in order to generate an evidence-based recommendation/recommendations for the clinical question (see Section 4.2.1 on how to access the body of evidence table).

Every author is asked to rate five key components (evidence base, consistency, clinical impact, generalizability and applicability) of the ‘body of evidence’.

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.

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.

For a detailed explanation in regards to each key component, see Appendix 4: Excerpt NHMRC Definitions of the components of the evidence statement. The components described above should be rated according to the matrix shown in Table 2.

Table 2: Body of evidence assessment matrix

<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>Evidence base**</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>
<tr>
<td>Generalisability</td>
<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>
<td>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</td>
<td>population/s studied in body of evidence differ to target population and hard to judge whether it is sensible to generalise to target population</td>
</tr>
<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>

---

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

**Note:** If a clinical question results in several evidence-based recommendations, a body of evidence assessment matrix has to be completed for each individual recommendation.

### 5.2 Step 2: Summarising the evidence by developing brief evidence summary statement(s)

The guideline authors synthesise 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 this 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,III-2</td>
<td>1,3,9</td>
</tr>
</tbody>
</table>

### 5.3 Step 3: Formulating the recommendation

Based on the evidence assessment and evidence summary statement, the author develops the wording for the recommendation/s. A 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. (For more information, see National Health and Medical Research Council “How to use the evidence: assessment and application of scientific evidence” Canberra: 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)
5.4 Step 4: Determining an overall recommendation grade

The question author then determines the overall recommendation grade by looking at the individual components determined in step 1 (see 5.1).

If most components are rated A then the overall recommendation grade will also be A**.

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

Figure 7: Screenshot completed evidence assessment matrix

If an overall recommendation 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, the clinician will have to determine 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.

CCA understands 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

CCA uses the NHMRC grading system for recommendations, as it is well recognised and based on an international standard. The recommendation grades are as follows:

<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 judgments and to 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.

5.5 Step 5: Development of content narrative

The next step for the authors is to develop the more detailed narrative to complete the content format (see example of format of Wiki guideline content below). The body of evidence is summarised in narrative form and the author is able to make qualifying statements about the evidence base.

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 when the references are uploaded to the CCA Cancer Guidelines Wiki.

Addition of practice points where applicable

Occasionally, clinicians may find that there is an important practice point that they wish to emphasise for which there is low level, low quality or no evidence. This is typically an aspect of treatment that is regarded as sound clinical practice, however, no good-quality evidence is available. These are called practice points. Practice points do not require an accompanying evidence summary, as it is not applicable. Consensus among guideline committee members is required for practice points. For example:

**Practice point:**

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

**Figure 8: Example of format of guideline content on the Wiki**

<table>
<thead>
<tr>
<th>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?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INTRODUCTION</strong></td>
</tr>
<tr>
<td>Background narrative text here</td>
</tr>
<tr>
<td><strong>Subheading – e.g. Radiation alone versus Combination Chemoradiotherapy</strong></td>
</tr>
<tr>
<td>Narrative text here and evidence summary and recommendation(s) and/or practice point(s) where relevant</td>
</tr>
<tr>
<td><strong>Subheading – e.g. Concurrent versus Sequential Therapy</strong></td>
</tr>
<tr>
<td>Narrative text here and evidence summary and recommendation(s) and/or practice point(s) where relevant</td>
</tr>
<tr>
<td><strong>EVIDENCE SUMMARY AND RECOMMENDATIONS</strong></td>
</tr>
<tr>
<td>Evidence summary</td>
</tr>
<tr>
<td>“Intervention x is as effective as...” or “There is insufficient data to support...”</td>
</tr>
</tbody>
</table>
**Recommendation**

*Patients with xxx should be offered...*

**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>Satisfactory</th>
<th>D: Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume of evidence</strong></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><strong>Consistency</strong></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><strong>Clinical impact</strong></td>
<td>very large</td>
<td>substantial</td>
<td>moderate</td>
<td>slight or restricted</td>
</tr>
<tr>
<td><strong>Generalisability</strong></td>
<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>
<td>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</td>
<td>population/s studied in body of evidence differ to target population and hard to judge whether it is sensible to generalise to target population</td>
</tr>
<tr>
<td><strong>Applicability</strong></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>

**Practice point:**

*Consider...*

**ISSUES REQUIRING FURTHER CLINICAL STUDY**

If gaps in the evidence are identified during the evidence review, please note areas for further research including a brief description. There are formatted as a research question.

**REFERENCES**

List of references used.

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 the draft content is made publicly available. This is done to ensure that there are no inconsistencies of recommendations between content sections.

Guideline working party members review the content directly on the Wiki by using the internal comments button underneath each comment section (see Figure 9).* This way all comments relating to a particular section are captured in one place.

Figure 9: Screenshot clinical question display including commenting buttons

* In order to be able to view the internal comment buttons and have comments automatically associated to a username, users need to be ‘logged in’.

Working party members will be automatically alerted of any new comments by a daily commenting digest sent via email. Alternatively, they can access all new comments directly on the Wiki in the user menu bar.

Figure 10: User Menu

A face-to-face meeting of the working party is organised to approve all content and recommendations of the final draft guidelines before they are released for public consultation.

7  Public consultation and guideline dissemination
Responsibility of Cancer Council Australia, invited colleges, societies and the greater public

7.1  Public consultation

Once the draft guidelines have been finalised, 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 consultation will take place online using the online commenting tool.

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. Afterwards, there is a short period to allow for consideration and incorporation of public comments, then the draft status is 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.

The table below gives some examples of how to respond to public comments.

Table 3: Examples of responses to stakeholder comments posted online

<table>
<thead>
<tr>
<th>Type of comment</th>
<th>Example of a response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliments about the guideline</td>
<td>Thank you for your comments. We appreciate your feedback about the guideline...</td>
</tr>
<tr>
<td>A specific change was recommended and has subsequently been made</td>
<td>Thank you for your comment; we have changed xx to xxx.</td>
</tr>
<tr>
<td>A specific change was recommended and has subsequently been partially made</td>
<td>Thank you for your comment; we have added a section on xxx discussed in other parts of the guideline. We have incorporated some of your suggestions into the text. Note: It is possible to include a link to the content history that illustrate the text differences. Please contact the CCA Guidelines Team for a demonstration of how to do this.</td>
</tr>
<tr>
<td>A specific change was recommended and has subsequently NOT been made</td>
<td>Thank you for your submission. Although the use of xxx is an option for xx management, there are a number of factors supporting the decision to consider x as the preferred method. Firstly, there is a larger evidence base for x. Secondly, ie. cost implications, timing, availability...</td>
</tr>
<tr>
<td>Asks for something that is outside the scope of the guideline</td>
<td>Thank you for bringing this to our attention. I agree that we should have a section on multidisciplinary care and meetings, but it was not within the scope of the present document, as it was not one of the topics we addressed. I think it might be best placed in a section on initial assessment/patient selection for treatment, which is yet to be written and is planned for the future. This will require a systematic review including a search strategy and literature appraisal.</td>
</tr>
</tbody>
</table>

A summary of comments can be generated interactively per question and/or for the full guideline in the appendix (see Figure 6).

7.2 Guideline dissemination and implementation

Cancer Council Australia’s clinical practice guidelines are made available online internationally via the CCA Cancer 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 continuing medical education (CME) events. A significant effort is made to introduce the guidelines to senior undergraduate medical students and to encourage the relevant learned colleges (surgeons, radiation oncologists, pathologists and general practitioners) to support the guidelines. Involvement in resident and registrar education activities helps foster guideline integration into hospital and community practice. Use of Cancer Council Australia guidelines as part of core curriculum in specialty exams will be encouraged.

CCA implemented a web analytics solution on the CCA Clinical Guidelines Wiki to track and analyse website traffic and the effectiveness of marketing efforts. Amongst other things, the web analytics solution enables CCA 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
• how the visitors navigated through a guideline/the website.

In addition, CCA is also inviting external reviewers to complete a brief feedback survey. This gives an in-depth insight into guideline usage and allows CCA to evaluate and refine dissemination measures.

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 and 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. This is being piloted through the development of QStream modules (previously known as Spaced Education), which were originally developed by Harvard Medical School. QStream programs have shown to improve knowledge acquisition in a number of randomised trials with medical practitioners. (Visit: [http://qstream.com/company/brain-science/](http://qstream.com/company/brain-science/))

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).

Once online guidelines are established, the original PubMed and Embase literature searches are automatically re-run on a monthly basis. The monthly alerts generate pending literature searches which appear on the Project Officer’s user page (see Figure 10), who screens them for relevance (see Figure 11). Any relevant articles are added to the main author’s user page as new pending appraisals.

Figure 11: Screenshot userpage Project Officer – Literature search

![Screenshot of Project Officer's Literature Search page]

- Guidelines Barrett’s (pending literature searches)
- Guidelines Colorectal cancer/Colonoscopy surveillance (pending literature searches)
- Guidelines Endometrial cancer/Treatment/Early stage (pending literature searches)
- Guidelines Lung cancer (pending literature searches)
- Guidelines Sarcoma (pending literature searches)
Working party authors receive notifications if there is any new literature awaiting appraisal for their clinical question. The question authors appraise the new evidence and check if it should result in a content update.

In addition, public comments are received on an ongoing basis. Members of the public can continue to comment and notify CCA of new evidence by attaching references of new research papers to their comments.

The CCA Project Officer will be notified of any new literature submissions by members of the public on his/her user page, check if they are relevant and confirm whether they meet the inclusion and exclusion criteria of the systematic review period. If the literature submitted is irrelevant, it will be excluded and the reasons are recorded. If the submission is relevant, a critical appraisal is assigned and the question author further appraises the literature submission and determines if the content needs to be updated. The person who submitted new evidence is able to track their submission by visiting the permalink to their submission.

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 content or when there is new literature awaiting critical appraisal and incorporation into their guideline section. Each question has a main author who nominates two co-authors to help 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 (see Figure 13). If required, the whole working party can be consulted.

An annual working party meeting is organised by CCA to revise all guideline content changes made by individual author teams since the last working party meeting.
9 References


Appendix 1: Guideline adaption process

Where there are existing guidelines or recommendations that answer particular clinical questions, these can be considered for adaption. In some cases, the evidence-base for an existing recommendation can be used to answer a particular clinical question and develop a recommendation. This eliminates the need for de novo recommendation development.

If an existing guideline(s), specific clinical questions of (an) existing guideline(s) or the evidence-base of an existing guideline(s) is considered for adaption, the ADAPTE process is followed.9

Firstly, relevant guidelines are identified by searching guideline portals, such as:
- National Guideline Clearinghouse
- the Guidelines International Network Guideline library
- websites of recognised guideline development groups – e.g. National Institute for Clinical Excellence (NICE), Scottish Intercollegiate Guideline Network (SIGN), the American Society of Clinical Oncology (ASCO), Cancer Ontario.

Guidelines identified as relevant are evaluated using the AGREE II instrument. To be considered for adaptation, an existing guideline must score at least 70% for each of the following AGREE II domains, rigour, clarity and editorial independence. The guideline working party then considers whether the identified recommendation(s) might be endorsed (the recommendations can be used without modification); adapted (the recommendations should be modified); or the evidence-base can be used, taking into account the currency of the guidelines, as well as the relevance of the recommendations in regards to the population of interest and Australian healthcare context.

If using an existing evidence base, the systematic review for the existing recommendation is updated to include the latest evidence. This means the literature search is updated, results of the search update are screened against inclusion/exclusion criteria, included studies are critically appraised and the evidence tables are updated with the new included studies. Recommendations are then developed as described under Section 5: Assessing the body of evidence, writing the clinical question content and formulating the evidence statements and recommendations.

Levels of evidence and recommendation grades are determined according to NHMRC criteria. The adaption process followed for a particular Cancer Council Australia guideline is described in the guideline development process section of the respective guideline. The technical report of the individual clinical question includes the detailed adaption report. This contains the reference to the evidence base of the existing guideline; results of the literature search update; critical appraisals of any additional studies that were included as a result of the literature update; and the evidence table for the results of any update. This is available in the Appendices section on the clinical question page.

Appendix 2: Conflict of interest management

Prior to commencement of a guideline project, working party nominees are asked to declare in writing any interests relevant to the guideline development. The small management committee is responsible for evaluating all statements.

All declarations are added to the register of interests for the guidelines. The register is made available to the working party throughout the development of the guideline, allowing members to take any potential conflicts of interest into consideration during discussions, decision making and formulation of recommendations. Members are asked to update their information throughout the guideline development if they became aware of any changes to their interests.

If working party members are identified as having a significant real or perceived conflict of interest, the
management committee will decide to put conflict of interest management strategies in place. For example, a conflicted member might be excluded from writing the guideline, or she/he may need to leave the meeting room whilst the particular topic is discussed. Alternatively, the member could remain present but not participate in the discussion or decision making where they are conflicted.

A summary of conflict of interest management strategies is published together with the register.

Once the guidelines enter the updating phase, guideline working party members are responsible to update their conflict of interest statements if a new interest arises. The members will receive a formal reminder to review their statements and ensure statements are up-to-date prior to the annual meetings, which are scheduled to review all content updates of a specific guideline.

Appendix 3: 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:


This publication includes the accompanying handbooks:


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

Appendix 4: Excerpt NHMRC Definitions of the components of the evidence statement

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 (i.e. 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, as it ensures that the results are likely to be replicable or if they are 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 a Level I \(^2\) statistic\(^2\), 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, and 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
- 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 (e.g. 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 (e.g. 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 healthcare 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 (e.g. availability of trained staff, clinic time, specialised equipment, tests or other resources) and cultural factors (e.g. attitudes to health issues, including those that may affect compliance with the recommendation).

\(^1\) 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](http://www.sign.ac.uk/guidelines/fulltext/50/annexd.html) Accessed 19.10.07)

\(^2\) Whereas most statistical tests of heterogeneity (e.g. Cochrane’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)
### Appendix 5: Critical Appraisal help forms

#### Critical appraisal help form: Designations of ‘levels of evidence’ according to type of clinical question

<table>
<thead>
<tr>
<th>Level</th>
<th>Intervention 1</th>
<th>Diagnostic accuracy 2</th>
<th>Prognosis</th>
<th>Aetiology 3</th>
<th>Screening Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>I 4</td>
<td>A systematic review of level II studies</td>
<td>A systematic review of level II studies</td>
<td>A systematic review of level II studies</td>
<td>A systematic review of level II studies</td>
<td>A systematic review of level II studies</td>
</tr>
<tr>
<td>II</td>
<td>A randomised controlled trial</td>
<td>A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among consecutive persons with a defined clinical presentation</td>
<td>A prospective cohort study</td>
<td>A prospective cohort study</td>
<td>A randomised controlled trial</td>
</tr>
<tr>
<td>III-1</td>
<td>A pseudorandomised controlled trial (i.e. alternate allocation or some other method)</td>
<td>A study of test accuracy with: an independent, blinded comparison with a valid reference standard, among non-consecutive persons with a defined clinical presentation</td>
<td>All or none</td>
<td>All or none</td>
<td>A pseudorandomised controlled trial (i.e. alternate allocation or some other method)</td>
</tr>
</tbody>
</table>
| III-2 | A comparative study with concurrent controls:  
  - Non-randomised, experimental trial  
  - Cohort study  
  - Case-control study  
  - Interrupted time series with a control group | A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence | Analysis of prognostic factors amongst persons in a single arm of a randomised controlled trial | A retrospective cohort study | A comparative study with concurrent controls:  
  - Non-randomised, experimental trial  
  - Cohort study  
  - Case-control study |
| III-3 | A comparative study without concurrent controls:  
  - Historical control study  
  - Two or more single arm study  
  - Interrupted time series without a parallel control group | Diagnostic case-control study | A retrospective cohort study | A case-control study | A comparative study without concurrent controls:  
  - Historical control study  
  - Two or more single arm study |
| IV    | Case series with either post-test or pre-test/post-test outcomes | Study of diagnostic yield (no reference standard) | Case series, or cohort study of persons at different stages of disease | A cross-sectional study or case series | Case series |

Excerpt pages from National health and Medical Research Council. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Canberra: NHMRC, 2009
Explanatory notes

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

2 The dimensions of evidence apply only to studies of diagnostic accuracy. To assess the effectiveness of a diagnostic test there also needs to be a consideration of the impact of the test on patient management and health outcomes (Medical Services Advisory Committee 2005, Sackett and Haynes 2002).

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 (i.e. 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 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 (Whiting et al 2003).

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 (Mulherin and Miller 2002).

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 unbiassed 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.

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 research questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms are rare and cannot feasibly be captured within randomised controlled trials; physical harms and psychological harms may need to be addressed by different study designs; harms from diagnostic testing include the likelihood of false positive and false negative results; harms from screening include the likelihood of false alarm and false reassurance results.

Note B: When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding research question e.g. level II intervention evidence; level IV diagnostic evidence; level III-2 prognostic evidence. Source: Hierarchies adapted and modified from: NHMRC 1999; Bandolier 1999; Lijmer et al. 1999; Phillips et al. 2001.
1 Critical appraisal help form: Study design glossary

Study design categories

Each study design is grouped into 10 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 category (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. If the clinical question addresses one or more risk factors, three quality appraisals tools specifically designed to evaluate the study designs nested case control, case control and cohort studies are available.

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. **Quasi-experimental (Pseudo-randomised trials, non-randomised trials)**
   This includes 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

4. **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.

   Included studies:
   - 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

5. 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
• Population based case-control study

6. 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
• All or none study
• Single arm phase I clinical trial
• Single arm phase II clinical trial
• Interrupted time series (without control)
• Predictive accuracy study of a single test

7. 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.

8. Health economic study
Health economics studies are defined here as full economic evaluation studies, partial economic evaluation studies, and single effectiveness studies that include more limited information relating to the description, measurement or valuation of resource use associated with interventions.

See Cochrane handbook for more detailed explanation:
http://handbook.cochrane.org/chapter_15/15_1_2_economics_and_economic_evaluation.htm

9. Diagnostic accuracy (cross-sectional) study
The outcomes from one or more diagnostic tests are compared with outcomes from a reference standard test in a group or groups of people.

Included studies:
• Diagnostic accuracy cohort type study
• Diagnostic accuracy case-control study
• Diagnostic accuracy population based case-control study
• Diagnostic yield study
10. Other
We have included this category for any study designs that do not fit any of the main categories.

11. Risk factor – Nested case control study
Adapted from the Newcastle-Ottawa tool developed by Expert Advisory Panel for Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer.

12. Risk factor – Case control study
Adapted from the Newcastle-Ottawa tool developed by Expert Advisory Panel for Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer.

13. Risk factor – Cohort study
Adapted from the Newcastle-Ottawa tool developed by Expert Advisory Panel for Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer.

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 small pox 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 known as ‘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 randomised 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 100-150 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. Cross-over 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.*
Diagnosis 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.

Health economic study
Health economics studies are defined here as full economic evaluation studies, partial economic evaluation studies, and single effectiveness studies that include more limited information relating to the description, measurement or valuation of resource use associated with interventions.

See Cochrane handbook for more detailed explanation: http://handbook.cochrane.org/chapter_15/15_1_2_economics_and_economic_evaluation.htm

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 versus B and B versus C to determine A versus C but without statistical adjustment for B).
Critical appraisal – risk of bias assessment


Once the risk of bias assessment is completed for the study designs 1-7, the system will generate a suggestion on what your overall risk of bias rating should be according to a specific algorithm.

You would then select your overall risk of bias rating for the study (as seen below).

If needed, the free text field allows you to enter additional comments in regards to your risk of bias rating.

The risk of bias assessment forms for each study category and the key for the overall risk of bias rating that are used in the online form are listed below for your information.

For the study designs ‘health economic study’ and ‘diagnostic accuracy study’, no overall risk of bias rating is generated. We are using the Drummond checklist to assess risk of bias for health economic studies and the QUADAS-2 tool to assess risk of bias for diagnostic accuracy studies, which does not involve assigning an overall risk of bias score.
Risk of bias 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 summarised?
   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. If there was heterogeneity, were sources of heterogeneity explored?
   2= Yes
   1= Some attempt was made
   0= No
   N/A No heterogeneity

Key to overall risk of bias rating

**Low risk of bias**: a review that received 2 for all relevant questions (Questions 1-3)

**Moderate risk of bias**: a review that received 1 and 2 for all relevant questions (Questions 1-3)

**High risk of bias**: a review that received 0 for any of the relevant questions (Questions 1-3)

Answers to question 4 and 5 serve as additional risk of bias information for systematic reviews that pooled data. They are not factored into the calculation of the overall risk of bias score.
Risk of bias assessment form: Randomised Controlled Trials Cochrane Tool

Domain: Sequence generation

Include description of the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups in the text box.

1. What was the risk of bias from the random sequence generation?

Ratings:
- **Low:** Investigators describe a random component in the sequence generation such as a random number table, computerized random number generator, coin tossing or throwing dice
- **High:** Investigators describe a non-random component in the sequence generation such as using odd or even date of birth, hospital or clinic record number or preference of the participant
- **Unclear:** Insufficient information to permit judgement

Domain: Allocation concealment

Include description of the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment in the text box.

2. What was the risk of bias from the allocation concealment?

Ratings:
- **Low:** Participants and investigators enrolling participants could not foresee assignment because one of the following (or equivalent) was used: central allocation, sequentially numbered drug containers of identical appearance or sequentially numbered opaque sealed envelopes
- **High:** Participants or investigators enrolling participants could possibly foresee assignments, such as allocation based on: Envelopes without safeguards (unsealed or non-opaque), alternation or rotation, date of birth, open random allocation schedule or case record number
- **Unclear:** Insufficient information to permit judgement

Domain: Blinding

Include description of all measures used, if any, to blind outcome assessors from knowledge of which intervention a participant received in the text box. Provide any information relating to whether the intended blinding was effective.

3. What was the risk of bias from the blinding of participants and personnel and outcome assessors?

Ratings:
- **Low:** No blinding, but review authors judge that the outcomes are not likely to be influenced by lack of blinding, blinding of participants/personnel ensured and unlikely broken, participants or some personnel not blinded but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias
- **High:** No/incomplete blinding and outcome is likely to be influenced by lack of blinding, blinding attempted but likely the blinding could have been broken, participants or some personnel not blinded and the non-blinding likely to introduce bias
- **Unclear:** Insufficient information or study did not address this outcome

Domain: Incomplete outcome data

Include description of the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis in the text box. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomized participants), reasons for attrition/exclusions where reported, and any re-inclusions in analyses performed by the review authors.
4. What was the risk of bias from incomplete outcome data?

Ratings:
- **Low**: No missing outcome data, reasons for missing outcome data unlikely to be related to true outcome, missing outcome data balanced across intervention groups
- **High**: Reasons for missing outcome data likely to be related to true outcome with either imbalance in numbers of reasons for missing data across intervention groups, ‘as-treated’ analysis done with substantial departure of the intervention received from that assigned at randomisation, inappropriate application of simple imputation
- **Unclear**: Insufficient reporting to permit judgement or study did not address this outcome

**Domain: Selective outcome reporting**

Include in the text box how the possibility of selective outcome reporting was examined by the review authors and what was found.

5. What was the risk of bias from selective outcome reporting?

Ratings:
- **Low**: Study protocol is available and all pre-specified outcomes have been reported, study protocol is not available but it is clear that the published reports include all expected outcomes
- **High**: Not all pre-specified primary outcomes have been reported, one or primary outcomes are reported using measurements, analysis method or subsets of data that were not pre-specified, reported outcomes were not pre-specified, study report fails to include results for a key outcome that would be expected
- **Unclear**: Insufficient information to permit judgement (majority of studies will fall into this category)

**Domain: Other sources of bias**

Describe any other sources of bias in the text box.

6. What was the risk of bias from other sources?

Ratings:
- **Low**: Study appears to be free of other sources of bias
- **High**: There is at least one important risk of bias, such as potential sources of bias related to the specific study design used, stopped early due to data-dependent process, extreme baseline imbalance, claims to be fraudulent or some other problem
- **Unclear**: Insufficient information to assess whether an important risk of bias exists or insufficient rationale or evidence that an identified problem will introduce bias

**Key to overall risk of bias rating** (note that it is recommended to refer to the individual risk of bias rating for each domain compared to the overall score as it is more specific)

- **Low risk of bias**: a review that received ‘low risk of bias’ for all domains
- **High risk of bias**: a review that received ‘high risk of bias’ for at least one domain
- **Unclear risk of bias**: a review that received ‘unclear risk of bias’ for at least one domain
Risk of bias assessment form: Quasi-experimental studies (Pseudo-randomised)

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 technology used?
   2= Yes
   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)

2. Comparability of groups on demographic characteristics and clinical features
   2= Comparable
   1= Not comparable but adjusted analysis used
   0= Not comparable and not adjusted for differences

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

Key to overall risk of bias rating

Low risk of bias: a review that received 2 for all risk of bias criteria

Moderate risk of bias: Received 2 and 1 for all risk of bias criteria

High risk of bias: Received 0 for all risk of bias criteria or 1 and 0 for all risk of bias criteria or received 0 for any of the risk of bias criteria
### Risk of bias assessment form: 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 risk of bias rating

Low risk of bias: a review that received 2 for all risk of bias criteria

Moderate risk of bias: Received 2 and 1 for all risk of bias criteria

High risk of bias: Received 0 for all risk of bias criteria or 1 and 0 for all risk of bias criteria or received 0 for any of the risk of bias criteria
### Risk of bias assessment form: Case-control study

0. **Subject selection**
   
   (a) **Cases**
   
   2= Representative of eligible patients (Equals a population-based cancer registry)
   
   1= Selected group (Equals a multi-institutional study)
   
   0= Highly selected or not described (Equals a single institution study)

(b) **Adequacy of case definition**

   2= Independent validation of outcome (blind to exposure status)
   
   1= Taken from medical records, self report without independent validation
   
   0= Highly selected, inappropriate or not described

(c) **Controls**

   2= From same population as cases and same exclusion criteria used
   
   1= Selected group (e.g. hospital controls)
   
   0= Highly selected, inappropriate or not described

1. **Comparability of groups on demographic characteristics and important potential confounders**

   2= Comparable (or matched)
   
   1= Not comparable but adjusted analysis used
   
   0= Not comparable and not adjusted for differences

2. **Ascertainment of exposure/treatment**

   2= Blinded to case/control status
   
   0= No or not described

3. **Follow-up complete and all patients included in the analysis?**

   2= Complete response
   
   1= Non-response unlikely to introduce bias (>80% in both cases and controls)
   
   0= Low response rate (<80%), non-responders not described, differential response in cases/controls, or no details provided

### Key to overall risk of bias rating

**Low risk of bias:** a review that received 2 for all risk of bias criteria

**Moderate risk of bias:** Received 2 and 1 for all risk of bias criteria

**High risk of bias:** Received 0 for all risk of bias criteria or 1 and 0 for all risk of bias criteria or received 0 for any of the risk of bias criteria
Risk of bias assessment form: Case Series

1. Subject Selection
   2= Representative of eligible patients (Equals a population-based cancer registry)
   1= Selected group  (Equals a multi-institutional study)
   0= Highly selected or not described (Equals a single institution study)

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 risk of bias rating

Low risk of bias: a review that received 2 for three main criteria

Moderate risk of bias: Received 2 and 1 for all three criteria

High risk of bias: Received 0 for all three criteria or 0 and 1 for all three criteria or received 0 for any of the three criteria
Risk of bias assessment form: Cross-sectional studies

1. Subject Selection
   2= Representative of eligible patients (Equals a population-based cancer registry)
   1= Selected group (Equals a multi-institutional study)
   0= Highly selected or not described (Equals a single institution study)

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 risk of bias rating

Low risk of bias: a review that received 2 for three main criteria

Moderate risk of bias: Received 2 and 1 for all three criteria

High risk of bias: Received 0 for all three criteria or 0 and 1 for all three criteria or received 0 for any of the three criteria
## Risk of bias assessment form: Health economic
### Drummond checklist (Drummond 1996)

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Not clear</th>
<th>Not appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study design</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The research question is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The economic importance of the research question is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The viewpoint(s) of the analysis are clearly stated and justified.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The rationale for choosing alternative programmes or interventions compared is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The alternatives being compared are clearly described.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The form of economic evaluation used is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. The choice of form of economic evaluation is justified in relation to the questions addressed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. The source(s) of effectiveness estimates used are stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Details of the design and results of effectiveness study are given (if based on a single study).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. The primary outcome measure(s) for the economic evaluation are clearly stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Methods to value benefits are stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Details of the subjects from whom valuations were obtained were given.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Productivity changes (if included) are reported separately.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. The relevance of productivity changes to the study question is discussed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Quantities of resource use are reported separately from their unit costs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Methods for the estimation of quantities and unit costs are described.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Currency and price data are recorded.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Details of currency of price adjustments for inflation or currency conversion are given.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Details of any model used are given.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. The choice of model used and the key parameters on which it is based are justified.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analysis and interpretation of results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Time horizon of costs and benefits is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. The discount rate(s) is stated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. The choice of discount rate(s) is justified.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. An explanation is given if costs and benefits are not discounted.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Details of statistical tests and confidence intervals are given for stochastic data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. The approach to sensitivity analysis is given.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>28.</td>
<td>The choice of variables for sensitivity analysis is justified.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>The ranges over which the variables are varied are justified.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.</td>
<td>Relevant alternatives are compared.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>Incremental analysis is reported.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>Major outcomes are presented in a disaggregated as well as aggregated form.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.</td>
<td>The answer to the study question is given.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.</td>
<td>Conclusions follow from the data reported.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>Conclusions are accompanied by the appropriate caveats.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Source:**


See also Cochrance Handbook for more information about this quality appraisal tool http://handbook.cochrane.org/chapter_15/15_5_addressing_risk_of_bias.htm
**Quadas-2 appraisal tool**

The QUADAS-2 appraisal tool is designed to assess the risk of bias of primary diagnostic accuracy studies. It consists of four key domains covering patient selection, index test, reference standard, and flow of patients through the study and timing of the index test(s) and reference standard (“flow and timing”). Each section asks to complete information fields and questions to support the risk of bias judgment of each key domain.

### 1. Patient selection

(see also **1.2.1 DOMAIN 1: PATIENT SELECTION** in Quadas-2 background document):

**Intended use of test:**

**Prior tests and any referral filters:**

Describe prior tests and any referral filters

**Presentation:**

Describe condition that defined entry into study

**Setting:**

Describe setting eg tertiary, hospital, specialist clinic or primary care

**Was a diagnostic case-control design avoided?**

Yes/No/Unclear

**Either consecutive or random sample?**

Yes/No/Unclear

**Did study avoid inappropriate exclusions?**

Yes/No/Unclear

Please enter reasons in text field below:

If comparing more than one index test was the design fully paired (both tests performed on same patient) or paired randomised?

Yes/No/Unclear/Not applicable (only 1 index test)
If a paired randomised design was used, was allocation to groups concealed and was the generation of allocation sequence adequate?

Yes/No/Unclear/Not applicable (did not use paired randomised design)

**Could the selection of participants have introduced bias?**

**RISK: Low/High/Unclear**

---

2. Index test 1

(see also 1.2.2 DOMA IN 2: INDEX TEST in Quadas-2 background document):

Describe index test and how it was conducted and interpreted:

Were the index test results interpreted without knowledge of the results of the reference standard?\(^1\)

Yes/No/Unclear/Not applicable (no reference standard)

If a threshold was used, was it pre-specified?

Yes/No/Unclear/Not applicable (no threshold used)

If 2 tests are being compared have they been assessed independently/ blind to each other?\(^1\)

Yes/No/Unclear/Not applicable (only 1 index test)

**Could the conduct or interpretation of the index test have introduced bias?**

**RISK: Low/High/Unclear**
3. Index test 2 – if comparing 2 index tests

Describe index test and how it was conducted and interpreted:

Were the index test results interpreted without knowledge of the results of the reference standard?¹
Yes/No/Unclear/Not applicable (no reference standard used/ only 1 index test)

If a threshold was used, was it pre-specified?
Yes/No/Unclear/Not applicable – (no threshold used/ only 1 index test)

Could the conduct or interpretation of the index test have introduced bias?
RISK: Low/High/Unclear

4. Reference Standard
(see also 1.2.3 DOMAIN 3: REFERENCE STANDARD in Quadas-2 background document):

Reference standard

Describe the reference standard and how it was conducted and interpreted:

Is the reference standard likely to correctly classify the target condition?²
Yes/No/Unclear/Not applicable (no reference standard used)
*Note: If the ref standard is not 100% accurate the reviewer/working group will need to pre-specify, which % if any would be acceptable eg 99%, 98%?

Were the reference standard results interpreted without knowledge of the results of the index test/s?
Yes/No/Unclear/Not applicable (no reference standard used)

Was the reference test standard independent of the index test (ie the index test did not form part of the reference standard)?
Yes/No/Unclear/Not applicable (no reference standard used)

Could the reference standard, its conduct, or its interpretation have introduced bias?
RISK: Low/High/Unclear/Not applicable
5. Flow and Timing

<table>
<thead>
<tr>
<th>Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Describe the time interval and any interventions between index test(s) and reference standard</th>
</tr>
</thead>
</table>

| If a predictive test ie the reference standard is a later event that the test aims to predict, were any subsequent interventions between test and later event blind to test result? | Yes/No/Unclear/Not applicable (not a predictive test) |

| Was there an appropriate interval between index test(s) and reference standard (If appropriate appraisers/working party will need to predefine what is an appropriate interval)? | Yes/No/Unclear/Not applicable (no reference test) |

| Did all participants or a random sample of participants receive a reference standard test? | Yes/No (If “No” appraisers/working party will need to predefine maximum acceptable proportion not verified)/Unclear/Not applicable (no reference test) |

| Did all patients receive the same reference standard irrespective of index test result? | Yes/No/Unclear/Not applicable- no reference test |

| Were all test results including unclear results reported? | Yes/No/Unclear |

| Were all patients included in the analysis? | Yes/No/Unclear |

| **Could the patient flow have introduced bias?** |

| **RISK: Low/High/Unclear** |

---

This is relevant to assess the level of evidence.
Risk of bias assessment form: Prevalence and Incidence studies

1. Source population of subjects
   2. Population based
      1. Selected group (multiple institutions)
      0. Highly selected (single institution) or not described

2. Subject (sample) selection
   2. Representative of eligible patients (consecutive or random sample e.g. states all patients recruited in given time frame)
      1. Selected group (debatable whether group is representative e.g. consecutive sample but extensive exclusion criteria)
      0. Highly selected (selection at clinician’s discretion regardless of whether sample consecutive), unclear or not described

3. Measurement of condition
   2. Objective and reliable (based on registry or other equivalent source)
      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 risk of bias rating

**Low risk of bias**: scores 2 for all items

**Moderate risk of bias**: scores 1 or 2 for all items

**High risk of bias**: scores 0 for at least one item
Risk of bias assessment tools for risk factor questions

NESTED CASE-CONTROL STUDIES RISK OF BIAS ASSESSMENT TOOL (RISK FACTORS)

(Adapted from the Newcastle-Ottawa tool for QA of clinical cohort studies and case-control studies for use in the Prostate Cancer Foundation of Australia and Cancer Council Australia auspiced - Development of Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer)

Bias in selection of participants into nested case-control study

Sources of cases and controls

1. Drawn from the same population* (low risk)
2. Drawn from different populations but unlikely to introduce bias (moderate risk)
3. Drawn from different populations and likely to introduce bias OR insufficient information to tell (high risk)

* This will usually be the case when a case-control study is nested in a single cohort containing exposed and unexposed people and cases accrue during follow-up of the whole cohort.

Selection of cases and controls

1. Cases and controls are randomly selected from all available cases and controls; controls matched to cases by risk set* (either at selection or during analysis) (low risk)
2. Only one of the two criteria in 1 is met (moderate risk)
3. Neither criterion in 1 is met OR insufficient information to tell (high risk)

*Risk set defined by sex, age group, date of entry into cohort and date of case-defining event

Bias due to error in outcome measurement

Definition of cases (outcome)

1. Outcome precisely specified and with pathological or other objective confirmation (low risk)
2. Outcome precisely specified but without known pathological or other objective confirmation OR outcome precisely specified, self-reported and cases blind to hypotheses related to outcome (moderate risk)
3. Outcome imprecisely specified OR outcome self-reported and cases not blind to hypotheses related to outcome OR insufficient information to tell (high risk)

Definition of controls

1. Objective evidence of no past history of outcome of interest (low risk)
2. Self-report of no past history of outcome of interest OR insufficient information to tell (moderate risk)
Was outcome of interest likely to have been absent at the time to which the exposure refers?

1. Yes (low risk)
2. No but outcome unlikely to affect exposure measurement (moderate risk)
3. No and outcome likely to affect exposure measurement OR insufficient information to tell (high risk)

Was follow-up long enough for outcome to occur as a consequence of the measured exposure? (Requires prior specification of a sufficient follow-up period)

1. Yes (low risk)
2. No OR insufficient information to tell (high risk)

Bias due to error in exposure measurement

Measurement of error

1. Objective measurements from pre-existing records or baseline physical or biological assessment or structured interview, each blind to case or control status (low risk)
2. Objective measurements from pre-existing records or baseline physical or biological assessment not blind to case or control status OR structured interview blind to case or control status (moderate risk)
3. Structured interview not blind to case or control status OR self-administered questionnaire OR insufficient information to tell (high risk)

Was the same method used to measure exposure in cases and controls?

1. Yes (low risk)
2. No OR insufficient information to tell (high risk)

Bias due to non-participation

Participation rate in cohort

1. Participation rate in exposed cohort is ≤10 percentage points different from non-exposed cohort OR exposed and non-exposed are from the same cohort (low risk)
2. Participation rate in exposed cohort is >10 percentage points but <20 percentage points different from non-exposed cohort (moderate risk)
3. Participation rate in exposed cohort ≥20 percentage points different from non-exposed cohort OR insufficient information to tell (high risk)

Participation (response) rate for cases

1. ≥70% participation rate (≥80% response rate) (low risk)
2. ≥50 to <70% participation rate (≥60 to <80% response rate) (moderate risk)
3. <50% participation rate (<60% response rate) OR insufficient information to tell (high risk)
4. Not applicable – new data not being collected from participants

Footnote: Existing at or before baseline, where baseline is the time at which a participant is recorded to have entered the cohort or, if obtained after baseline, a time before onset of symptoms of the outcome or any likely effect of the developing outcome on the exposure.
Participation (response) rate for cases

1. ≥70% participation rate (≥80% response rate) (low risk)
2. ≥50 to <70% participation rate (≥60 to <80% response rate) (moderate risk)
3. <50% participation rate (<60% response rate) OR insufficient information to tell (high risk)
4. Not applicable – new data not being collected from participants

Participation (response) rate for controls

1. ≥60% participation rate (≥70% response rate) (low risk)
2. ≥40 to <60% participation rate (≥50 to <70% response rate) (moderate risk)
3. <40% participation rate (<50% response rate) OR insufficient information to tell (high risk)
4. Not applicable – new data not being collected from participants

Difference in participation rate (response rate) between cases and controls

1. Participation or response rate in cases ≤10 percentage points different from controls (low risk)
2. Participation or response rate in cases is >10 to ≤15 percentage points different from controls (moderate risk)
3. Participation or response rate in cases is >15 percentage points different from controls OR insufficient information to tell (high risk)
4. Not applicable - new data not being collected from participants

Bias due to missing data

Completeness of follow-up of cohort

1. Active or passive follow-up of participants with methods for ascertainment of outcome and death clearly described AND with methods for ascertainment of emigration from population-at-risk clearly described or censoring at date of last follow-up OR there is a plausible estimate of >90% follow-up (low risk)
2. Active or passive follow-up with methods for ascertainment of outcome, death and emigration from population-at-risk not clearly described OR there is a plausible estimate of 70 – 90% follow-up (moderate risk)
3. Active or passive follow-up with methods for ascertainment of one or more of outcome, death or emigration not described OR there was probably <70% follow-up OR insufficient information to tell (high risk)

Accuracy of dates of outcome or censoring

1. Dates of outcome or censoring ascertained to within one year (low risk)
2. One or more of dates of outcome or censoring not ascertained to within one year OR insufficient information to tell (moderate risk)

Difference in follow-up between exposed and non-exposed members of cohort

1. Follow-up methods are the same and likely to achieve the same completeness of follow-up in exposed and unexposed participants (low risk)
2. Completeness of follow-up in exposed and unexposed participants is unlikely to be the same but difference between the two is, or would be likely to be, small (<10%) (moderate risk)
3. Completeness of follow-up in exposed and unexposed participants is very unlikely to be the same and difference between the two is, or is likely to be, large (≥10%) OR insufficient information to tell (high risk)
**Difference in missing data for exposure between cases and controls**

1. Difference in missing data for exposure <10 percentage points (low risk)
2. Difference in missing data for exposure ≥10 to <20 percentage points (moderate risk)
3. Difference in missing data for exposure ≥20 percentage points OR insufficient information to tell (high risk)

**Bias due to confounding**

**Comparability of cases and controls with respect to potentially important confounding variables (Requires prior specification of potentially important confounders)**

1. Age and other potentially important confounders measured and controlled by design or in analysis (low risk)
2. Age and some but not all other potentially important confounders controlled by design or in analysis (moderate risk)
3. No potentially important confounders or only age controlled by design or in analysis OR insufficient information to tell (high risk)

**Analysis bias**

**Analysis appropriate to design**

1. When controls are frequency matched to cases, matching variables are controlled in the analysis OR when controls are individually matched to cases, a conditional analysis is used or matching variables are controlled in the analysis (low risk)
2. None of the above OR insufficient information to tell (high risk)

**Covariates are appropriately included in statistical analysis models**

1. Variables measuring the same underlying concept or lying in the same causal pathway ARE NOT included together as covariates in statistical analysis models (low risk)
2. Variables measuring the same underlying concept or lying in the same causal pathway ARE included together as covariates in statistical analysis models OR insufficient information to tell (high risk)

**Key to overall risk of bias rating**

- **High risk of bias** – high risk of bias in any domain
- **Moderate risk of bias** – moderate or low risk of bias in all domains, no high risk domains
- **Low risk of bias** – all domains low risk of bias, no moderate or high risk domains
## COHORT STUDIES RISK OF BIAS ASSESSMENT TOOL (RISK FACTORS)

(Adapted from the Newcastle-Ottawa tool for QA of clinical cohort studies for use in the Prostate Cancer Foundation of Australia and Cancer Council Australia auspiced - Development of Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer)

### Bias in selection of participants into study

**Selection of the exposed and non-exposed cohorts**

1. Drawn from the same population (low risk)
2. Drawn from different populations but unlikely to introduce bias (moderate risk)
3. Drawn from different populations and likely to introduce bias OR insufficient information to tell (high risk)

### Bias due to error in exposure measurement

**Measurement of exposure**

1. Objective measurements from pre-existing records or baseline\(^1\) physical or biological assessment blind to outcome status (low risk)
2. Objective measurements from pre-existing records or baseline\(^1\) physical or biological assessment not blind to outcome status, OR structured interview (moderate risk)
3. Self-administered questionnaire OR insufficient information to tell (high risk)

### Bias due to error in outcome measurement

**Measurement of outcome**

1. Outcome measurement unlikely to be influenced by exposure (low risk)
2. Objective outcome measurement possibly influenced by exposure (moderate risk)
3. Objective outcome measurement probably influenced by exposure OR self-reported outcome OR insufficient information to tell (high risk)

**Was outcome of interest absent at the time to which the exposure refers?**

1. Yes (low risk)
2. No but outcome unlikely to affect exposure measurement (moderate risk)
3. No and outcome likely to affect exposure measurement OR insufficient information to tell (high risk)

**Was follow-up long enough for outcome to occur as a consequence of measured exposure?** (Requires prior specification of a sufficient follow-up period)

1. Yes (low risk)
2. No OR insufficient information to tell (high risk)

\(^1\) Existing at or before baseline, where baseline is the time at which a participant is recorded to have entered the cohort or, if obtained after baseline, before onset of symptoms of the outcome or any likely effect of the developing outcome on the exposure
Bias due to non-participation

Participation rate

1. Participation rate in exposed cohort is ≤10 percentage points different from non-exposed cohort OR exposed and non-exposed are from the same cohort (low risk)
2. Participation rate in exposed cohort is >10 percentage points but <20 percentage points different from non-exposed cohort (moderate risk)
3. Participation rate in exposed cohort ≥20 percentage points different from non-exposed cohort OR insufficient information to tell (high risk)

Bias due to missing data

Completeness of follow-up

1. Active or passive follow-up of participants with methods for ascertainment of outcome and death clearly described AND with methods for ascertainment of emigration from population-at-risk clearly described or censoring at date of last follow-up OR there is a plausible estimate of >90% follow-up (low risk)
2. Active or passive follow-up with methods for ascertainment of outcome, death and emigration from population-at-risk not clearly described OR there is a plausible estimate of 70 – 90% follow-up (moderate risk)
3. Active or passive follow-up with methods for ascertainment of one or more of outcome, death or emigration not described OR there was probably <70% follow-up OR insufficient information to tell (high risk)

Accuracy of dates of outcome or censoring

1. Dates of outcome or censoring ascertained to within one year (low risk)
2. One or more of dates of outcome or censoring not ascertained to within one year OR insufficient information to tell (moderate risk)

Difference in follow-up between exposed and non-exposed

1. Follow-up methods are the same and likely to achieve the same completeness of follow-up in exposed and unexposed participants (low risk)
2. Completeness of follow-up in exposed and unexposed participants is unlikely to be the same but difference between the two is, or would be likely to be, small (<10%) (moderate risk)
3. Completeness of follow-up in exposed and unexposed participants is very unlikely to be the same and difference between the two is, or is likely to be, large (≥10%) OR insufficient information to tell (high risk)

Difference in missing data for exposure between those with or without the outcome

1. Difference in missing data for exposure <10 percentage points (low risk)
2. Difference in missing data for exposure ≥10 to <20 percentage points (moderate risk)
3. Difference in missing data for exposure ≥20 percentage points (high risk) OR insufficient information to tell (high risk)
Bias due to confounding

Comparability of exposed and non-exposed cohorts with respect to potentially important confounding variables (Requires prior specification of potentially important confounders)

1. Age and other potentially important confounders measured and controlled by design or in analysis (low risk)
2. Age and some but not all other potentially important confounders controlled by design or in analysis (moderate risk)
3. No potentially important confounders or only age controlled by design or in analysis OR insufficient information to tell (high risk)

Analysis bias

Covariates are appropriately included in statistical analysis models

1. Variables measuring the same underlying concept or lying in the same causal pathway ARE NOT included together as covariates in statistical analysis models (low risk)
2. Variables measuring the same underlying concept or lying in the same causal pathway ARE included together as covariates in statistical analysis models OR insufficient information to tell (high risk)

Key to overall risk of bias rating

High risk of bias – high risk of bias in any domain
Moderate risk of bias – moderate or low risk of bias in all domains, no high risk domains
Low risk of bias – all domains low risk of bias, no moderate or high risk domains
CASE-CONTROL STUDIES RISK OF BIAS ASSESSMENT TOOL (RISK FACTORS)

(Adapted from the Newcastle-Ottawa tool for QA of clinical case-control studies for use in the Prostate Cancer Foundation of Australia and Cancer Council Australia auspiced – Development of Clinical Practice Guidelines for PSA Testing and Early Management of Test-Detected Prostate Cancer)

Bias in selection of participants into study

Selection of cases and controls

1. Drawn from the same population (low risk)
2. Drawn from different populations but unlikely to introduce bias (moderate risk)
3. Drawn from different populations and likely to introduce bias OR insufficient information to tell (high risk)

Bias due to error in outcome measurement

Definition of cases (outcome)

1. Outcome precisely specified and with pathological or other objective confirmation (low risk)
2. Outcome precisely specified but without known pathological or other objective confirmation OR outcome precisely specified, self-reported and cases blind to hypotheses related to outcome (moderate risk)
3. Outcome imprecisely specified OR outcome self-reported and cases not blind to hypotheses related to outcome OR insufficient information to tell (high risk)

Definition of controls

1. Objective evidence of no past history of outcome of interest (low risk)
2. Self-report of no past history of outcome of interest OR insufficient information to tell (moderate risk)

Bias due to error in exposure measurement

Measurement of exposure

1. Objective measurements from pre-existing\(^1\) records or biological assessment blind to case or control status (low risk)
2. Objective measurements from pre-existing\(^1\) records or biological assessment not blind to case or control status OR structured interview blind to case or control status (moderate risk)
3. Structured interview not blind to case or control status OR self-administered questionnaire OR insufficient information to tell (high risk)

Temporality of exposure

1. Exposure precedes onset of disease in cases and is from a corresponding calendar time in controls (low risk)
2. Exposure precedes onset of disease in cases but exposure in controls is not from a calendar time corresponding to exposure in cases (moderate risk)
3. Exposure does not precede onset of disease in cases OR insufficient information to tell (high risk)

\(^1\) Existing before onset of symptoms of the outcome or any likely effect of the developing outcome on the exposure
Was the same method used to measure exposure in cases and controls?

1. Yes (low risk)
2. No OR insufficient information to tell (high risk)

Bias due to non-participation

Participation (response) rate for cases

1. ≥70% participation rate (≥80% response rate) (low risk)
2. ≥50 to <70% participation rate (≥60 to <80% response rate) (moderate risk)
3. <50% participation rate (<60% response rate) OR insufficient information to tell (high risk)
4. Not applicable – new data not being collected from participants

Participation (response) rate for controls

1. ≥60% participation rate (≥70% response rate) (low risk)
2. ≥40 to <60% participation rate (≥50 to <70% response rate) (moderate risk)
3. <40% participation rate (<50% response rate) OR insufficient information to tell (high risk)
4. Not applicable – new data not being collected from participants

Difference in participation rate (response rate) between cases and controls

1. Participation (response) rate in cases ≤10 percentage points different from controls (low risk)
2. Participation (response) rate in cases is >10 to ≤15 percentage points different from controls (moderate risk)
3. Participation (response) rate in cases is >15 percentage points different from controls OR insufficient information to tell (high risk)

Bias due to missing data

Difference in missing data for exposure between cases and controls

1. Difference in missing data for exposure < 10 percentage points (low risk)
2. Difference in missing data for exposure ≥10 to <20 percentage points (moderate risk)
3. Difference in missing data for exposure ≥20 percentage points OR insufficient information to tell (high risk)

Bias due to confounding

Comparability of cases and controls with respect to potentially important confounding variables
(Requires prior specification of potentially important confounders)

1. Age and other potentially important confounders measured and controlled by design or in analysis (low risk)
2. Age and some but not all other potentially important confounders controlled by design or in analysis (moderate risk)
3. No potentially important confounders or only age controlled by design or in analysis OR insufficient information to tell (high risk)
Analysis bias

Analysis appropriate to design

1. When controls are frequency matched to cases, matching variables are controlled in the analysis OR when controls are individually matched to cases, a conditional analysis is used or matching variables are controlled in the analysis (low risk)
2. None of the above OR insufficient information to tell (high risk)

Covariates are appropriately included in statistical analysis models

1. Variables measuring the same underlying concept or lying in the same causal pathway ARE NOT included together as covariates in statistical analysis models (low risk)
2. Variables measuring the same underlying concept or lying in the same causal pathway ARE included together as covariates in statistical analysis models OR insufficient information to tell (high risk)

Key to overall risk of bias rating

High risk of bias – high risk of bias in any domain
Moderate risk of bias – moderate or low risk of bias in all domains, no high risk domains
Low risk of bias – all domains low risk of bias, no moderate or high risk domains
**Risk of bias assessment form: Randomised Controlled Trials (used prior to November 2015)**

1. Was the trial double-blinded?

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>I am reasonably certain that the trial was double-blinded (eg identical placebo, active placebo, double-dummy, no revealing side-effects).</td>
</tr>
<tr>
<td>1</td>
<td>Trial was double-blinded but may have limitations (eg method of blinding inappropriate, ta injection with no double-dummy, different treatment schedules, side-effects may unblind) or single-blinded (eg outcomes assessed blind, objective outcomes, no revealing side-effects).</td>
</tr>
<tr>
<td>0</td>
<td>Outcomes not blinded, substantial side-effects, or not reported.</td>
</tr>
</tbody>
</table>

2. Concealment of treatment allocation schedule

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Adequately concealed (eg central randomisation, numbered or coded bottles, drugs prepared by pharmacy).</td>
</tr>
<tr>
<td>1</td>
<td>Inadequately concealed (eg numbered envelopes, sealed envelopes, alternation, medical record number, date of birth).</td>
</tr>
<tr>
<td>0</td>
<td>No concealment or unclear (eg no approach described, open randomisation lists, person doing recruitment to toss a coin).</td>
</tr>
</tbody>
</table>

3. Inclusion of all randomised participants in analysis (ie intention-to-treat anal.)

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>No exclusions or survival analysis used with all subjects included (note: follow-up may not be complete but balanced between the comparison groups).</td>
</tr>
<tr>
<td>1</td>
<td>Exclusions not likely to cause bias (some incomplete follow-up but balanced between comparison groups + survival analysis not used).</td>
</tr>
<tr>
<td>0</td>
<td>Too many exclusions, differential loss in comparison groups, or not reported.</td>
</tr>
</tbody>
</table>

**The field below is not considered when calculating the evidence risk of bias rating**

4. Generation of allocation sequences

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adequate (eg random number table, computer random number generator, coin tossing, card shuffling).</td>
</tr>
<tr>
<td>0</td>
<td>Inadequate or not reported.</td>
</tr>
</tbody>
</table>

**Key to overall risk of bias rating**

- **Low risk of bias**: 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))
- **Moderate risk of bias**: Received 2 and 1 for all three main criteria
- **High risk of bias**: Received 0 for all three criteria or 0 and 1 for all three criteria or received 0 for any of the three criteria

Answer for question 4 is considered as additional information and not considered when calculating the overall risk of bias score. Risk of bias assessment questions 1 to 3 for randomised control trials are evidence-based categories (Schulz et al (1995); Jadad et al (1996). Generation of allocation sequences has been shown not to influence outcomes.
References


Further resources


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.

Relevance of the Evidence for Medical Tests

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>Direct evidence of an effect on patient-relevant clinical outcomes beyond the outcomes achieved using an existing test; such as survival, quality of life and morbidity(^1).</td>
</tr>
<tr>
<td>2</td>
<td>Evidence of the clinical performance of the test on test accuracy or intermediate outcomes(^2) beyond that achieved using an existing test, where evidence is available to estimate patient-relevant outcomes in the same population and clinical pathway(^3).</td>
</tr>
<tr>
<td>3</td>
<td>Evidence of the clinical performance of the test on test accuracy or intermediate outcomes beyond that achieved using an existing test, where assumptions required to estimate patient-relevant outcomes from existing evidence for a different definition of the target condition, population or pathway are judged to be reasonable(^4).</td>
</tr>
<tr>
<td>4</td>
<td>Evidence of the clinical performance of the test on test accuracy or intermediate outcome, but where there is insufficient evidence to estimate patient-relevant outcomes(^5).</td>
</tr>
<tr>
<td>5</td>
<td>Evidence confined to analytic performance of the test(^6).</td>
</tr>
</tbody>
</table>

\(^1\) Example: The ARTISTIC randomised controlled trial comparing human papillomavirus (HPV) testing as an addition to cytology versus cytology alone for cervical screening, reported on differences in high grade pre-cancer and cancer detection rates and psychological distress to provide direct evidence about the clinical effectiveness (benefits and harms) of this HPV screening strategy (1).

\(^2\) Clinical performance includes test accuracy for diagnosis, prognosis, or prediction of treatment response or harm. Intermediate outcomes include change in patient management.

\(^3\) This includes studies that provide evidence of test accuracy for a target condition, and there is other evidence available to estimate the clinical consequences of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Example: Accuracy study of computed tomography (CT) colonography versus Barium enema for the detection of colorectal polyps \(\geq 10\text{mm}\) in a symptomatic population, where the benefits of early excision of polyps \(\geq 10\text{mm}\) has been established by observational evidence about risk of progression to cancer (2).

\(^4\) This includes studies that provide evidence of test accuracy for a target condition, but the evidence available to estimate the clinical consequences of TP, TN, FP, and FN are based on a different classification of disease or test population.

Example: Accuracy study of CT colonography versus barium enema for the detection of colorectal polyps of all sizes, including polyps \(\leq 5\text{mm}\), but the clinical benefits for removing polyps is based on evidence from studies of polyps \(> 5\text{mm}\).
This includes studies that provide evidence of test accuracy for a target condition, but that there is insufficient evidence to estimate the clinical consequences of TP and FN.

**Example:** Accuracy study of magnetic resonance imaging (MRI) versus mammography for the staging early breast cancer where evidence is available to demonstrate improved MRI sensitivity for the detection of multifocal cancer and higher rates of mastectomy but the benefits of mastectomy versus breast cancer surgery in women with multifocal disease detected by MRI alone has not yet been demonstrated in clinical studies (3).

This includes studies that provide evidence of analytic performance (including reliability and validity of the test), without evidence of the test accuracy for a target condition.

**Reference List**


Size of the effect for interventions

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 (i.e. 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

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

Size of effect for medical tests

For studies evaluating medical tests, please apply the pre-specified minimum clinical performance threshold for this test question 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 (i.e. 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

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