

The guidelines manual: appendices B–I

Tailored service improvement support

Published: 30 November 2012

nice.org.uk

Contents

Appendix B: Methodology checklist: systematic reviews and meta-analyses	4
Checklist	4
Notes on use of Methodology checklist: systematic reviews and meta-analyses	5
Appendix C: Methodology checklist: randomised controlled trials.....	8
Checklist	8
Notes on use of Methodology checklist: randomised controlled trials.....	11
Appendix D: Methodology checklist: cohort studies.....	19
Checklist	19
Notes on use of Methodology checklist: cohort studies	22
Appendix E: Methodology checklist: case-control studies.....	29
Checklist	29
Notes on use of the Methodology checklist: case-control studies.....	33
Appendix F: Methodology checklist: the QUADAS-2 tool for studies of diagnostic test accuracy... 37	
Phase 1: State the review question.....	37
Phase 2: Draw a flow diagram for the primary study.....	38
Phase 3: Risk of bias and applicability judgements.....	38
Notes on use of Methodology checklist: the QUADAS-2 tool for studies of diagnostic test accuracy	41
Appendix G: Methodology checklist: economic evaluations.....	42
Checklist.....	42
Notes on use of Methodology checklist: economic evaluations	44
Supporting references	57
Appendix H: Methodology checklist: qualitative studies	59
Checklist	59
Notes on use of Methodology checklist: qualitative studies	64
Further reading	68
Appendix I: Methodology checklist: prognostic studies.....	69
Checklist.....	69

Notes on use of Methodology checklist: prognostic studies 69

Appendix B: Methodology checklist: systematic reviews and meta-analyses

Checklist

Study identification <i>Include author, title, reference, year of publication</i>			
Guideline topic:	Review question no:		
Checklist completed by:			
SCREENING QUESTIONS			
In a well-conducted, relevant systematic review:	<i>Circle or highlight one option for each question</i>		
The review addresses an appropriate and clearly focused question that is relevant to the guideline review question	Yes	No	Unclear
The review collects the type of studies you consider relevant to the guideline review question	Yes	No	Unclear
The literature search is sufficiently rigorous to identify all the relevant studies	Yes	No	Unclear
Study quality is assessed and reported	Yes	No	Unclear
An adequate description of the methodology used is included, and the methods used are appropriate to the question	Yes	No	Unclear

If the review does not meet some or all of these criteria, it may still be useful as a source of references, but should not be relied upon on its own to address a review question.

If you have insufficient information on the design or quality of individual studies, you should use the checklists for studies on interventions (see appendices [C](#), [D](#) and [E](#)) to appraise each study. Each study should appear as a separate entry in the evidence table (see [appendix J](#)); the review should not appear in the evidence table.

If you plan to use the review as a whole, you will need to complete a row in an evidence table for the systematic review and input the results into an evidence profile as appropriate.

Notes on use of Methodology checklist: systematic reviews and meta-analyses

A systematic review uses explicit and systematic methods to identify, appraise and summarise the literature according to predetermined criteria. If the methods and criteria used to do this are not described or are not sufficiently detailed, it is not possible to make a thorough evaluation of the quality of the review.

The terms 'systematic review' and 'meta-analysis' are often used interchangeably. The term 'meta-analysis' is often used incorrectly to describe a systematic review that has used quantitative methods to summarise the results. However, it should be noted that meta-analysis refers only to the statistical techniques used to combine studies; thus not all meta-analyses are systematic reviews.

This checklist is intended for use with systematic reviews of questions about interventions and questions about diagnosis. It can potentially be used for any other types of question, although it has been designed primarily for the first two.

The aim of this checklist is to consider the suitability of the systematic review to answer a guideline review question. This assessment has two aspects: firstly, whether the question addressed by the review (in terms of the populations, interventions, comparisons and outcomes considered) is appropriate to answer the review question addressed by the guideline, and secondly, whether the methodology used for the review is sufficiently robust to permit a valid conclusion.

For each question in this section, you should indicate whether or not it has been addressed in the review. Choose 'unclear' if this aspect of the review process was ignored, or is not described in the report.

The review addresses an appropriate and clearly focused question that is relevant to the guideline review question

If the question addressed by the systematic review is not clearly stated, it will be difficult to determine whether the review is adequate to answer the guideline review question. If the question is not clear, the systematic review is unlikely to be a good one because it is difficult to be systematic in addressing an unclear question. The review report should give a clear description of the population considered, the interventions, exposures or tests evaluated, comparators, and outcomes evaluated. Inclusion and exclusion criteria should be clearly described. Outcomes considered should be clearly described within the methodology, including a precise definition and acceptable methods of measuring. The appropriateness of the question addressed in the systematic review for answering

the guideline review question can be determined by comparing these components. If the review does not consider all of the outcomes that are judged to be important to your guideline review question, you may still be able to use the outcome data but may need to review the individual studies to obtain other outcome data.

The review collects the type of studies you consider relevant to the guideline review question

You should be clear about the characteristics of studies that you consider will adequately address your guideline review question. These may relate to minimum design or quality characteristics (for example, randomised trials only). Systematic reviews should report the types of studies they sought, including any inclusion/exclusion criteria used. You can use this information to quickly assess the review's suitability for your purpose.

The literature search is sufficiently rigorous to identify all the relevant studies

Systematic and rigorous searches can help to minimise publication biases and identify as many relevant data as possible. Exact search terms depend on the review question, but there are core databases that should have been searched for every question. As a minimum, a well-conducted review should look at Embase and MEDLINE. For questions about interventions in particular, the Cochrane Database of Systematic Reviews (CDSR), the Database of Abstracts of Reviews of Effects (DARE) and the Cochrane Central Register of Controlled Trials (CENTRAL) should also be searched. The dates on which the searches were carried out should be given in the review. Good-quality reviews will also attempt to identify relevant studies by handsearching of key journals and examining reference lists of retrieved studies for further references.

If the methods used to locate studies are not clearly reported, it will be difficult to determine whether the review is likely to have missed important relevant studies. Ideally, the search strategy used should be reported in sufficient detail that the process could be replicated.

Any restrictions applied to the search (such as language or year of publication) should also be reported. You should consider how these might have influenced the findings of the review.

Advice from the information specialist (and/or other members of the Guideline Development Group) working on the guideline may be useful to determine whether any important search terms have been omitted.

If the search described in the review is judged to be inadequate to identify all relevant studies, it may be possible to expand the search by including additional databases or extra search terms

within the search strategy, or by updating the search to identify more recently published studies. Any additional studies identified by this expanded search should be appraised for quality using the appropriate NICE checklist (see [appendices C–I](#)). They should appear individually in separate rows in an evidence table.

Study quality is assessed and reported

The inclusion of poor-quality studies within a review can result in biased estimates of effect. A well-conducted systematic review should have used clear criteria to assess whether individual studies had been appropriately designed and conducted, before deciding whether to include or exclude them. These criteria should be clearly described and should be reported for each study included. The quality appraisal checklists in [appendices C–I](#), as appropriate for the type of question and study design, can be used as a guide to the types of quality criteria that should be considered.

If there is no indication of such a quality assessment, the review is unlikely to be reliable enough to be used in formulating guideline recommendations. It may be necessary to obtain and quality appraise the individual studies as part of your review.

An adequate description of the methodology used is included, and the methods used are appropriate to the question

In common with primary research, the approach used to analyse the data should be described and justified where appropriate. This may include the choice of statistical test used to analyse the outcome data, meta-analytical techniques and approaches to dealing with heterogeneity, including the specification of any subgroup analyses and sensitivity analyses.

Appendix C: Methodology checklist: randomised controlled trials

Checklist

Study identification Include author, title, reference, year of publication					
Guideline topic:		Review question no:			
Checklist completed by:					
		Circle or highlight one option for each question			
A. Selection bias (systematic differences between the comparison groups)					
<u>A1</u>	An appropriate method of randomisation was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups)	Yes	No	Unclear	N/A
<u>A2</u>	There was adequate concealment of allocation (such that investigators, clinicians and participants cannot influence enrolment or treatment allocation)	Yes	No	Unclear	N/A
<u>A3</u>	The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was selection bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	

Likely direction of effect:					
<ul style="list-style-type: none"> • • • • • 					
B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)					
<u>B1</u>	The comparison groups received the same care apart from the intervention(s) studied	Yes	No	Unclear	N/A
<u>B2</u>	Participants receiving care were kept 'blind' to treatment allocation	Yes	No	Unclear	N/A
<u>B3</u>	Individuals administering care were kept 'blind' to treatment allocation	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was performance bias present? If so, what is the likely direction of its effect?					
<ul style="list-style-type: none"> • • • • 					
Low risk of bias	Unclear/unknown risk	High risk of bias			
Likely direction of effect:					
<ul style="list-style-type: none"> • • • • 					
C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants)					
<u>C1</u>	All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)	Yes	No	Unclear	N/A

<u>C2</u>	a. How many participants did not complete treatment in each group?				
	b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Yes	No	Unclear	N/A
<u>C3</u>	a. For how many participants in each group were no outcome data available? .				
	b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available).	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was attrition bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	
Likely direction of effect:					
D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)					
<u>D1</u>	The study had an appropriate length of follow-up	Yes	No	Unclear	N/A
<u>D2</u>	The study used a precise definition of outcome	Yes	No	Unclear	N/A

<u>D3</u>	A valid and reliable method was used to determine the outcome	Yes	No	Unclear	N/A
<u>D4</u>	Investigators were kept 'blind' to participants' exposure to the intervention	Yes	No	Unclear	N/A
<u>D5</u>	Investigators were kept 'blind' to other important confounding and prognostic factors	Yes	No	Unclear	N/A
<p>Based on your answers to the above, in your opinion was detection bias present? If so, what is the likely direction of its effect?</p> <ul style="list-style-type: none"> • • • • 					
Low risk of bias		Unclear/unknown risk		High risk of bias	
<p>Likely direction of effect:</p> <ul style="list-style-type: none"> • • • • 					

Notes on use of Methodology checklist: randomised controlled trials

The studies covered by this checklist are designed to answer questions about the relative effects of interventions such as drugs, psychological therapies, operations or placebos. Such studies can include comparisons of 'test and treat strategies' involving a diagnostic test and subsequent management. The checklist does not cover comparisons of diagnostic test accuracy or questions about prognosis.

Some of the items on this checklist may need to be filled in individually for different outcomes reported by the study. It is therefore important that the systematic reviewer has a clear idea of what the important outcomes are before appraising a study. You are likely to need input from the Guideline Development Group in defining the important outcomes.

Checklist items are worded so that a 'yes' response always indicates that the study has been designed/conducted in such a way as to minimise the risk of bias for that item. An 'unclear' response to a question may arise when the item is not reported or not clearly reported. 'N/A' should be used when a randomised controlled trial cannot give an answer of 'yes' no matter how well it has been done.

This checklist is designed to assess the internal validity of the study; that is, whether the study provides an unbiased estimate of what it claims to show. Internal validity implies that the differences observed between groups of participants allocated to different interventions may (apart from the possibility of random error) be attributed to the intervention under investigation. Biases are characteristics that are likely to make estimates of effect differ systematically from the truth.

Recording the presence and direction of bias

The checklist contains four sections (A–D), each of which addresses a potential source of bias relating to internal validity. At the end of each section you are asked to give your opinion on whether bias is present and to estimate the likely direction of this bias – that is, whether you think it will have increased or decreased the effect size reported by the study. It will not always be possible to determine the direction of bias, but thinking this through can help greatly in interpreting results.

A: Selection bias

Selection bias may be introduced into a study when there are systematic differences between the participants in the different treatment groups. As a result, the differences in the outcome observed may be explained by pre-existing differences between the groups rather than because of the treatment itself. For example, if the people in one group are in poorer health, then they are more likely to have a bad outcome than those in the other group, regardless of the effect of the treatment. The treatment groups should be similar at the start of the study – the only difference between the groups should be the intervention received.

Randomisation

There are two aspects to randomisation:

- generation of the random allocation sequence that results in groups that differ only randomly

- allocation concealment, so that both the participant and the investigator are unaware of which group the next participant will be allocated to when entering the study.

A1. An appropriate method of randomisation was used to allocate participants to treatment groups

If an appropriate method of randomisation has been used, each participant should have an equal chance of ending up in any of the treatment groups. Examples of random allocation sequences include random numbers generated by computer, tables of random numbers, and drawing of lots or envelopes. The allocation sequence should not be related to outcome or prognosis, or be predictable, such as date of birth or admission date.

There are some more complicated ways of allocating people to treatment groups that minimise the differences between groups, such as block randomisation and minimisation. Although these are not truly random, they are usually considered to be adequate for the purpose. If a study does not report the method of randomisation used, this should be scored as 'unclear'.

A2. There was adequate concealment of allocation

If investigators are aware of the allocation group for the next participant being enrolled in the study, there is potential for participants to be enrolled in an order that results in imbalances in important characteristics. For example, a clinician might feel that participants who are more unwell are likely to do better on a new, experimental, treatment and be tempted to enrol such participants when they know they will be allocated to that group. This would result in the participants in the intervention group being, on average, more unwell. Concealment of treatment group may not always be feasible (as in, for example, a comparison of a surgical with a medical intervention), but concealment of allocation up until the point of enrolment in the study should always be possible.

The information presented within the paper should provide some assurance that allocations were not known until at least the point of enrolment. Centralised allocation, computerised allocation systems and the use of coded identical containers are all regarded as adequate methods of concealment. Sealed envelopes can be considered as adequate concealment if the envelopes are serially numbered and opaque, and allocation is performed by a third party. Poor methods of allocation concealment include alternation, or the use of case record numbers, date of birth or day of the week.

If the method of allocation concealment used is regarded as poor, or relatively easy to subvert, the study must be given a lower quality rating. If a study does not report any concealment approach, this should be scored as 'unclear'.

A3. The groups were comparable at baseline, including all major confounding and prognostic factors

Studies may report the distributions of potential prognostic and confounding factors in the comparison groups, or important differences in the distribution of these factors may be noted.

Formal tests comparing the groups are problematic – failure to detect a difference does not mean that a difference does not exist, and multiple comparisons of factors may falsely detect some differences that are not real.

Clinical input may be required to determine whether all likely confounders have been considered. Confounding factors may differ according to outcome, so you will need to consider potential confounding factors for all of the outcomes that are of interest to your review.

B: Performance bias

Performance bias refers to systematic differences between the comparison groups in the care provided to the participants, other than the intervention under investigation.

This may consist of additional treatment, advice or counselling, rather than a physical intervention, or even simply a belief about the effects of an intervention. If performance bias is present, it can be difficult to attribute any observed effect to the experimental treatment rather than to the other factors.

B1. The comparison groups received the same care apart from the intervention(s) studied

There should be no differences between the treatment groups apart from the intervention received. If some participants received additional treatment (known as 'co-intervention'), this treatment is a potential confounding factor that may compromise the results.

Blinding

Blinding (also known as masking) refers to the process of withholding information about treatment allocation or exposure status from those involved in the study who could potentially be influenced by this information. This can include participants, investigators, those administering care and those involved in data collection and analysis. If people are aware of the treatment allocation or exposure status ('unblinded'), this can bias the results of studies, either intentionally or unintentionally, through the use of other effective co-interventions, decisions about withdrawal, differential

reporting of symptoms or influencing concordance with treatment. Blinding of those assessing outcomes is covered in section D on detection bias.

Blinding of participants and carers is not always possible, particularly in studies of non-drug interventions, and so performance bias may be a particular issue in these studies. It is important to think about the likely size and direction of bias caused by failure to blind.

The terms 'single blind', 'double blind' and even 'triple blind' are sometimes used in studies. Unfortunately, they are not always used consistently. Commonly, when a study is described as 'single blind', only the participants are blind to their group allocation. When both participants and investigators are blind to group allocation, the study is often described as 'double blind'. It is preferable to record exactly who was blinded, if reported, to avoid misunderstanding.

B2. Participants receiving care were kept 'blind' to treatment allocation

The knowledge of assignment to a particular treatment group may affect outcomes, such as a study participant's reporting of symptoms, self-use of other known interventions or even dropping out of the study.

B3. Individuals administering care were kept 'blind' to treatment allocation

If individuals who are administering the intervention and/or other care to the participant are aware of treatment allocation, they may treat participants receiving one treatment differently from those receiving the comparison treatment; for example, by offering additional co-interventions.

C: Attrition bias

Attrition refers to the loss of participants during the course of a study. Attrition bias occurs when there are systematic differences between the comparison groups with respect to participants lost, or differences between participants lost to the study and those who remain. Attrition can occur at any point after participants have been allocated to their treatment groups. As such, it includes participants who are excluded after allocation (and may indicate a violation of eligibility criteria), those who do not complete treatment (whether or not they continue measurement) and those who do not complete outcome measurement (regardless of whether or not treatment was completed). Consideration should be given to why participants dropped out, as well as how many. Participants who dropped out of a study may differ in some significant way from those who remained as part of the study throughout. Drop-out rates and reasons for dropping out should be similar across all treatment groups. The proportion of participants excluded after allocation should be stated in the

study report, and the possibility of attrition bias considered within the analysis; however, these are not always reported.

C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)

If the comparison groups are followed up for different lengths of time, then more events are likely to occur in the group followed up for longer, distorting the comparison. This may be overcome by adjusting the denominator to take the time into account; for example by using person-years.

C2a. How many participants did not complete treatment in each group?

A very high number of participants dropping out of a study should give concern. The drop-out rate may be expected to be higher in studies conducted over a longer period of time. The drop-out rate includes people who did not even start treatment; that is, they were excluded from the study after allocation to treatment groups.

C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)

If there are systematic differences between groups in terms of those who did not complete treatment, consider both why participants dropped out and whether any systematic differences in those who dropped out may be related to the outcome under study, such as potential confounders. Systematic differences between groups in terms of those who dropped out may also result in treatment groups that are no longer comparable with respect to potential confounding factors.

C3a. For how many participants in each group were no outcome data available?

A very high number of participants for whom no outcome data were available should give concern.

C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)

If there are systematic differences between groups in terms of those for whom no outcome data were available, consider both why the outcome data were not available and whether there are any systematic differences between participants for whom outcome data were and were not available.

D: Detection bias (this section should be completed individually for each important relevant outcome)

The way outcomes are assessed needs to be standardised for the comparison groups; failure to 'blind' people who are assessing outcomes can also lead to bias, particularly with subjective outcomes. Most studies report results for more than one outcome, and it is possible that detection bias may be present in a study for some, but not all, outcomes. It is therefore recommended that this section is completed individually for each important outcome that is relevant to the guideline review question under study. To avoid biasing your review, you should identify the relevant outcomes **before** considering the results of the study. Clinical input may be required to identify the most important outcomes for a review.

D1. The study had an appropriate length of follow-up

The follow-up of participants after treatment should be of an adequate length to identify the outcome of interest. This is particularly important when different outcomes of interest occur early and late after an intervention. For example, after surgical interventions there is usually an early harm because of side effects, with benefits apparent later on. A study that is too short will give an unbalanced assessment of the intervention.

For events occurring later, a short study will give an imprecise estimate of the effect, which may or may not also be biased. For example, a late-occurring side effect will not be detected in the treatment arm if the study is too short.

D2. The study used a precise definition of outcome

D3. A valid and reliable method was used to determine the outcome

The outcome under study should be well defined. It should be clear how the investigators determined whether participants experienced, or did not experience, the outcome. The same methods for defining and measuring outcomes should be used for all participants in the study. Often there may be more than one way of measuring an outcome (for example, physical or laboratory tests, questionnaire, reporting of symptoms). The method of measurement should be valid (that is, it measures what it claims to measure) and reliable (that is, it measures something consistently).

D4. Investigators were kept 'blind' to participants' exposure to the intervention

D5. Investigators were kept 'blind' to other important confounding and prognostic factors

In this context the 'investigators' are the individuals who are involved in making the decision about whether a participant has experienced the outcome under study. This can include those responsible for taking physical measurements and recording symptoms, even if they are not ultimately responsible for determining the outcome. Investigators can introduce bias through differences in measurement and recording of outcomes, and making biased assessments of a participant's outcome based on the collected data. The degree to which lack of blinding can introduce bias will vary depending on the method of measuring an outcome, but will be greater for more subjective outcomes, such as reporting of pain.

Physical separation of the assessment from the participant (for example, sending samples off to a laboratory) can often be considered as blind if it can be assumed that the laboratory staff are unaware of the treatment assignment.

Appendix D: Methodology checklist: cohort studies

Checklist

Study identification					
Include author, title, reference, year of publication					
Guideline topic:			Review question no:		
Checklist completed by:					
					Circle or highlight one option for each question:
A. Selection bias (systematic differences between the comparison groups)					
<u>A1</u>	The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome[s] under study)	Yes	No	Unclear	N/A
<u>A2</u>	Attempts were made within the design or analysis to balance the comparison groups for potential confounders	Yes	No	Unclear	N/A
<u>A3</u>	The groups were comparable at baseline, including all major confounding and prognostic factors	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was selection bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	
Likely direction of effect:					
.					
.					
.					
.					
B. Performance bias (systematic differences between groups in the care provided, apart from the intervention under investigation)					

<u>B1</u>	The comparison groups received the same care apart from the intervention(s) studied	Yes	No	Unclear	N/A
<u>B2</u>	Participants receiving care were kept 'blind' to treatment allocation	Yes	No	Unclear	N/A
<u>B3</u>	Individuals administering care were kept 'blind' to treatment allocation	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was performance bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	
Likely direction of effect:					
<ul style="list-style-type: none"> 					
C. Attrition bias (systematic differences between the comparison groups with respect to loss of participants)					
<u>C1</u>	All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)	Yes	No	Unclear	N/A
<u>C2</u>	a. How many participants did not complete treatment in each group?				
	b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)	Yes	No	Unclear	N/A
<u>C3</u>	a. For how many participants in each group were no outcome data available?				
	b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)	Yes	No	Unclear	N/A

Based on your answers to the above, in your opinion was attrition bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	
Likely direction of effect:					
.					
.					
.					
.					
D. Detection bias (bias in how outcomes are ascertained, diagnosed or verified)					
<u>D1</u>	The study had an appropriate length of follow-up	Yes	No	Unclear	N/A
<u>D2</u>	The study used a precise definition of outcome	Yes	No	Unclear	N/A
<u>D3</u>	A valid and reliable method was used to determine the outcome	Yes	No	Unclear	N/A
<u>D4</u>	Investigators were kept 'blind' to participants' exposure to the intervention	Yes	No	Unclear	N/A
<u>D5</u>	Investigators were kept 'blind' to other important confounding and prognostic factors	Yes	No	Unclear	N/A
Based on your answers to the above, in your opinion was detection bias present? If so, what is the likely direction of its effect?					
Low risk of bias		Unclear/unknown risk		High risk of bias	
Likely direction of effect:					
.					
.					
.					
.					

Notes on use of Methodology checklist: cohort studies

Cohort studies are designed to answer questions about the relative effects of interventions, such as drugs, psychological therapies, operations or placebos. Such studies can include comparisons of 'test and treat strategies' involving a diagnostic test and subsequent management. This checklist does not cover comparisons of diagnostic test accuracy or questions about prognosis.

Some of the items on this checklist may need to be filled in individually for different outcomes reported by the study. It is therefore important that the systematic reviewer has a clear idea of what the important outcomes are **before** appraising a study. You are likely to need input from the Guideline Development Group in defining the important outcomes.

Checklist items are worded so that a 'yes' response always indicates that the study has been designed/conducted in such a way as to minimise the risk of bias for that item. An 'unclear' response to a question may arise when the item is not reported or is not reported clearly. 'N/A' should be used when a cohort study cannot give an answer of 'yes' no matter how well it has been done.

This checklist is designed to assess the internal validity of the study; that is, whether the study provides an unbiased estimate of what it claims to show. Internal validity implies that the differences observed between groups of participants allocated to different interventions may (apart from the possibility of random error) be attributed to the intervention under investigation. Biases are characteristics that are likely to make estimates of effect differ systematically from the truth.

Recording the presence and direction of bias

This checklist contains four sections (A–D), each of which addresses a potential source of bias relating to internal validity. At the end of each section you are asked to give your opinion on whether bias is present, and to estimate the likely direction of this bias – whether you think it will have increased or decreased the effect size reported by the study. It will not always be possible to determine the direction of bias, but thinking this through can help greatly in interpreting results.

A: Selection bias

Selection bias can be introduced into a study when there are systematic differences between the participants in the different treatment groups. As a result, the differences in the outcome observed may be explained by pre-existing differences between the groups rather than because of the treatment itself. For example, if the people in one group are in poorer health, then they are more

likely to have a bad outcome than those in the other group, regardless of the effect of the treatment. The treatment groups should be similar at the start of the study – the only difference between the groups should be in terms of the intervention received.

The main difference between randomised trials and non-randomised studies is the potential susceptibility of the latter to selection bias. Randomisation should ensure that, apart from the intervention received, the treatment groups differ only because of random variation. However, care needs to be taken in the design and analysis of non-randomised studies to take account of potential confounding factors. There are two main ways of accounting for potential confounding factors in non-randomised studies. Firstly, participants can be allocated to treatment groups to ensure that the groups are equal with respect to the known confounders. Secondly, statistical techniques can be used within the analysis to take into account known differences between groups. Neither of these approaches is able to address unknown or unmeasurable confounding factors, and it is important to remember that measurement of known confounders is subject to error. It can rarely, if ever, be assumed that all important factors relevant to prognosis and responsiveness to treatment are known. Hence, considerable judgement is needed to assess the internal validity of non-randomised studies; clinical input may be needed to identify potential confounding factors that should be taken into consideration.

A1. The method of allocation to treatment groups was unrelated to potential confounding factors

In non-randomised studies, there will usually be a reason why participants are allocated to the treatment groups (often as a result of clinician and/or patient choice). If this reason is linked to the outcome under study, this can result in confounding by indication (where the decision to treat is influenced by some factor that is related in turn to the treatment outcome). For example, if the participants who are the most ill are selected for the treatment, then the treatment group may experience worse outcomes because of this difference between the groups at baseline. It will not always be possible to determine from the report of a study which factors influenced the allocation of participants to treatment groups.

A2. Attempts were made within the design or analysis to balance the comparison groups for potential confounders

This represents an attempt when designing the study to ensure that the groups are similar in terms of known confounding or prognostic factors, in order to optimise comparability between the treatment groups. For example, in a matched design, the controls are deliberately chosen to be equivalent to the treatment group for any potential confounding variables, such as age and sex.

An alternative approach is to use statistical techniques to adjust for known confounding factors in the analysis.

A3. The groups were comparable at baseline, including all major confounding and prognostic factors

Studies may report the distributions of potential prognostic and confounding factors in the comparison groups, or important differences in these factors may be noted.

Formal tests comparing the groups are problematic – failure to detect a difference does not mean that a difference does not exist, and multiple comparisons of factors may falsely detect some differences that are not real.

Clinical input may be needed to determine whether all likely confounders have been considered. Confounding factors may differ according to outcome, so you will need to consider potential confounding factors for each of the outcomes that are of interest to your review.

B: Performance bias

Performance bias refers to systematic differences in the care provided to the participants in the comparison groups, other than the intervention under investigation.

This may consist of additional treatment, advice or counselling, rather than a physical intervention, or even simply a belief about the effects of an intervention. If performance bias is present, it can be difficult to attribute any observed effect to the experimental treatment rather than to the other factors.

Performance bias can be more difficult to determine in non-randomised studies than in randomised studies, because the latter are likely to have been better planned and executed according to strict treatment protocols that specify standardised interventions and care. It may be particularly difficult to determine performance bias for retrospective studies, where there is usually no control over standardisation.

B1. The comparison groups received the same care apart from the intervention(s) studied

There should be no differences between the treatment groups apart from the intervention received. If some participants received additional treatment (known as 'co-intervention'), this treatment is a potential confounding factor that may compromise the results.

Blinding

Blinding (also known as masking) refers to the process of withholding information about treatment allocation or exposure status from those involved in the study who could potentially be influenced by this information. This can include participants, investigators, those administering care and those involved in data collection and analysis. If people are aware of the treatment allocation or exposure status ('unblinded'), this can bias the results of studies, either intentionally or unintentionally, through the use of other effective co-interventions, decisions about withdrawal, differential reporting of symptoms or influencing concordance with treatment. Blinding of those assessing outcomes is covered in section D on detection bias.

Blinding of participants and carers is not always possible, particularly in studies of non-drug interventions, and so performance bias may be a particular issue in these studies. It is important to think about the likely size and direction of bias caused by failure to blind.

The terms 'single blind', 'double blind' and even 'triple blind' are sometimes used in studies. Unfortunately, they are not always used consistently. Commonly, when a study is described as 'single blind', only the participants are blind to their group allocation. When both participants and investigators are blind to group allocation the study is often described as 'double blind'. It is preferable to record exactly who was blinded, if reported, to avoid misunderstanding.

B2. Participants receiving care were kept 'blind' to treatment allocation

The knowledge of assignment to a particular treatment group may affect outcomes such as a study participant's reporting of symptoms, self-use of other known interventions or even dropping out of the study.

B3. Individuals administering care were kept 'blind' to treatment allocation

If individuals who are administering the intervention and/or other care to the participant are aware of treatment allocation, they may treat participants receiving one treatment differently from those receiving the comparison treatment; for example, by offering additional co-interventions.

C: Attrition bias

Attrition refers to the loss of participants during the course of a study. Attrition bias occurs when there are systematic differences between the comparison groups with respect to participants lost, or differences between the participants lost to the study and those who remain. Attrition can occur at any point after participants have been allocated to their treatment groups. As such, it includes

participants who are excluded after allocation (and may indicate a violation of eligibility criteria), those who do not complete treatment (whether or not they continue measurement) and those who do not complete outcome measurement (regardless of whether or not treatment was completed). Consideration should be given to why participants dropped out, as well as how many. Participants who dropped out of a study may differ in some significant way from those who remained as part of the study throughout. Drop-out rates and reasons for dropping out should be similar across all treatment groups. The proportion of participants excluded after allocation should be stated in the study report and the possibility of attrition bias considered within the analysis; however, these are not always reported.

C1. All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up)

If the comparison groups are followed up for different lengths of time, then more events are likely to occur in the group followed up for longer, distorting the comparison. This may be overcome by adjusting the denominator to take the time into account; for example by using person-years.

C2a. How many participants did not complete treatment in each group?

A very high number of participants dropping out of a study should give concern. The drop-out rate may be expected to be higher in studies conducted over a longer period of time. The drop-out rate includes people who did not even start treatment; that is, they were excluded from the study after allocation to treatment groups.

C2b. The groups were comparable for treatment completion (that is, there were no important or systematic differences between groups in terms of those who did not complete treatment)

If there are systematic differences between groups in terms of those who did not complete treatment, consider both why participants dropped out and whether any systematic differences in those who dropped out may be related to the outcome under study, such as potential confounders. Systematic differences between groups in terms of those who dropped out may also result in treatment groups that are no longer comparable with respect to potential confounding factors.

C3a. For how many participants in each group were no outcome data available?

A very high number of participants for whom no outcome data were available should give concern.

C3b. The groups were comparable with respect to the availability of outcome data (that is, there were no important or systematic differences between groups in terms of those for whom outcome data were not available)

If there are systematic differences between groups in terms of those for whom no outcome data were available, consider both why the outcome data were not available and whether there are any systematic differences between participants for whom outcome data were and were not available.

D: Detection bias (this section should be completed individually for each important relevant outcome)

The way outcomes are assessed needs to be standardised for the comparison groups; failure to 'blind' people who are assessing the outcomes can also lead to bias, particularly with subjective outcomes. Most studies report results for more than one outcome, and it is possible that detection bias may be present for some, but not all, outcomes. It is therefore recommended that this section is completed individually for each important outcome that is relevant to the guideline review question under study. To avoid biasing your review, you should identify the relevant outcomes **before** considering the results of the study. Clinical input may be required to identify the most important outcomes for a review.

D1. The study had an appropriate length of follow-up

The follow-up of participants after treatment should be of an adequate length to identify the outcome of interest. This is particularly important when different outcomes of interest occur early and late after an intervention. For example, after surgical interventions there is usually early harm because of side effects, with benefits apparent later on. A study that is too short will give an unbalanced assessment of the intervention.

For events occurring later, a short study will give an imprecise estimate of the effect, which may or may not also be biased. For example, a late-occurring side effect will not be detected in the treatment arm if the study is too short.

D2. The study used a precise definition of outcome

D3. A valid and reliable method was used to determine the outcome

The outcome under study should be well defined and it should be clear how the investigators determined whether participants experienced, or did not experience, the outcome. The same methods for defining and measuring outcomes should be used for all participants in the study.

Often there may be more than one way of measuring an outcome (for example, physical or laboratory tests, questionnaire, reporting of symptoms). The method of measurement should be valid (that is, it measures what it claims to measure) and reliable (that is, it measures something consistently).

D4. Investigators were kept 'blind' to participants' exposure to the intervention

D5. Investigators were kept 'blind' to other important confounding and prognostic factors

In this context the 'investigators' are the individuals who are involved in making the decision about whether a participant has experienced the outcome under study. This can include those responsible for taking physical measurements and recording symptoms, even if they are not ultimately responsible for determining the outcome. Investigators can introduce bias through differences in measurement and recording of outcomes, and making biased assessments of a participant's outcome based on the collected data. The degree to which lack of blinding can introduce bias will vary depending on the method of measuring an outcome, but will be greater for more subjective outcomes, such as reporting of pain.

Physical separation of the assessment from the participant (for example, sending samples off to a laboratory) can often be considered as blind if it can be assumed that the laboratory staff are unaware of the treatment assignment.

Appendix E: Methodology checklist: case-control studies

Checklist

Study identification Include author, title, reference, year of publication			
Guideline topic:		Review question no:	
Checklist completed by:			
Section 1: Internal validity			
		<i>Circle or highlight one option for each question</i>	
<u>1.1</u>	The study addresses an appropriate and clearly focused question.	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<i>Selection of participants</i>			
<u>1.2</u>	The cases and controls are taken from comparable populations	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<u>1.3</u>	The same exclusion criteria are used for both cases and controls	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<u>1.4</u>	What was the participation rate for each group (cases and controls)?	Cases: Controls:	

<u>1.5</u>	Participants and non-participants are compared to establish their similarities or differences	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<u>1.6</u>	Cases are clearly defined and differentiated from controls	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<u>1.7</u>	It is clearly established that controls are not cases	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<i>Assessment</i>			
<u>1.8</u>	Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<u>1.9</u>	Exposure status is measured in a standard, valid and reliable way	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<i>Confounding factors</i>			

1.10	The main potential confounders are identified and taken into account in the design and analysis	Well covered Adequately addressed Poorly addressed	Not addressed Not reported Not applicable
<i>Statistical analysis</i>			
1.11	Have confidence intervals been provided?		

Section 2: Description of the study

(This information is required for evidence tables to facilitate cross-study comparisons. Please complete all sections for which information is available.)

Please print clearly

2.1	How many people participated in the study?	<p><i>List the numbers of cases and controls separately.</i></p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>
2.2	What are the main characteristics of the study population?	<p><i>Include all characteristics used to identify both cases and controls - for example, age, sex, social class, disease status.</i></p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>
2.3	What environmental or prognostic factor is being investigated?	<p>.</p> <p>.</p> <p>.</p> <p>.</p>

2.4	What comparisons are made?	<p><i>Normally only one factor will be compared, but in some cases the extent of exposure may be stratified – for example, non-smokers vs light, moderate or heavy smokers. Note all comparisons here.</i></p> <ul style="list-style-type: none"> • • • •
2.5	For how long are participants followed up?	<p><i>This is the length of time over which participant histories are tracked in the study.</i></p> <ul style="list-style-type: none"> • • • •
2.6	What outcome measure(s) is/are used?	<p><i>List all outcomes that are used to assess the impact of the chosen environmental or prognostic factor.</i></p> <ul style="list-style-type: none"> • • • • •
2.7	What size of effect is identified?	<p><i>Effect size should be expressed as an odds ratio. If any other measures are included, note them as well. Include p-values and any confidence intervals that are provided.</i></p> <ul style="list-style-type: none"> • • • •

2.8	How was the study funded?	<p><i>List all sources of funding quoted in the article, whether government, voluntary sector or industry.</i></p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>
2.9	Does this study help to answer your guideline review question?	<p><i>Summarise the main conclusions of the study and indicate how it relates to the review question.</i></p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>

Notes on use of the Methodology checklist: case-control studies

Case-control studies are designed to answer questions of the type 'What are the factors that caused this event?'. They involve comparison of individuals who have an outcome with other individuals from the same population who do not have the outcome. These studies start after the outcome of an event, and can be used to assess multiple causes of a single event. They are generally used to assess the causes of a new problem but they may also be useful for the evaluation of population-based interventions such as screening.

The questions in section 1 are aimed at establishing the internal validity of the study under review – that is, making sure that it has been carried out carefully, and that any link between events and outcomes is clearly established. Each question covers an aspect of methodology that has been shown to make a significant difference to the conclusions of a study.

Case-control studies need to be designed very carefully – the complexity of their design is often not appreciated by investigators, and so many poor-quality studies are conducted. The questions in this checklist are designed to identify the main features that should be present in a well-designed study. There are few criteria that should, alone and unsupported, lead to rejection of a study. However, a study that fails to address or report on more than one or two of the questions in the checklist should almost certainly be rejected.

For each question in this section you should choose one of the following categories to indicate how well it has been addressed in the study:

- well covered
- adequately addressed
- poorly addressed
- not addressed (not mentioned, or this aspect of study design was ignored)
- not reported (mentioned, but with insufficient detail to allow assessment to be made)
- not applicable.

Question

1.1 The study addresses an appropriate and clearly focused question

Unless a clear and well-defined question is specified, it will be difficult to assess how well the study has met its objectives or how relevant it is to the question you are trying to answer.

Selection of participants

1.2 The cases and controls are taken from comparable populations

Study participants may be selected from the target population (all individuals to which the results of the study could be applied), from the source population (a defined subset of the target population from which participants are selected) or from a pool of eligible people (a clearly defined and counted group selected from the source population). A study that does not include clear definitions of the source population should be rejected.

1.3 The same exclusion criteria are used for both cases and controls

All selection and exclusion criteria should be applied equally to cases and controls. Failure to do so may introduce a significant degree of bias into the results of the study.

1.4 What was the participation rate for each group (cases and controls)?

Differences between the eligible population and the study participants are important because they may influence the validity of the study. A participation rate can be calculated by dividing the number of study participants by the number of people who are eligible to participate. It is more useful if it is calculated separately for cases and controls. If the participation rate is low, or there is a large difference in rate between cases and controls, the study results may be invalid because of

differences between participants and non-participants. In these circumstances the study should be downgraded, and rejected if the differences are very large.

1.5 Participants and non-participants are compared to establish their similarities or differences

Even if participation rates are comparable and acceptable, it is still possible that the participants selected to act as cases or controls may differ from other members of the source population in some significant way. A well-conducted case-control study will look at samples of those not participating among the source population to ensure that the participants are a truly representative sample.

1.6 Cases are clearly defined and differentiated from controls

The method of selection of cases is of critical importance to the validity of the study. Investigators have to be certain that cases are truly cases, but must balance this with the need to ensure that the cases admitted into the study are representative of the eligible population. The issues involved in case selection are complex, and should ideally be evaluated by someone with a good understanding of the design of case-control studies. If there is no information on how cases were selected it is probably safest to reject the study as a source of evidence.

1.7 It is clearly established that controls are not cases

Just as it is important to be sure that cases are true cases, it is important to be sure that controls do not have the outcome under investigation. Controls should be chosen so that information on exposure status can be obtained or assessed in a similar way to that used for the selection of cases. If the methods of control selection are not described, the study should be rejected. If different methods of selection are used for cases and controls, the study should be evaluated by someone with a good understanding of the design of case-control studies.

Assessment

1.8 Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment

If there is a possibility that case ascertainment was influenced by knowledge of exposure status, assessment of any association is likely to be biased. A well-conducted study should take this into account in the design of the study.

1.9 Exposure status is measured in a standard, valid and reliable way

The inclusion of evidence from other sources or previous studies that demonstrate the validity and reliability of the assessment methods, or the fact that the measurement method is a recognised procedure, should increase confidence in study quality.

Confounding factors

1.10 The main potential confounders are identified and taken into account in the design and analysis

Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors is one of the principal reasons why observational studies are not more highly rated as a source of evidence. The report of the study should indicate which potential confounders have been considered, and how they have been assessed or accounted for in the analysis. Clinical judgement should be used to consider whether all likely confounders have been taken into account. If the measures used to address the potential effects of confounders are considered inadequate, the study should be downgraded or rejected, depending on how serious the risk of confounding is considered to be. A study that does not address the possibility of confounding should be rejected.

Statistical analysis

1.11 Have confidence intervals been provided?

Confidence intervals are the preferred method for indicating the precision of statistical results, and can be used to differentiate between an inconclusive study and a study that shows no effect. Studies that report a single value with no assessment of precision should be treated with caution.

Section 2 of the checklist asks you to summarise key points about the study that will be added to an evidence table (see appendix J) in the next stage of the process.

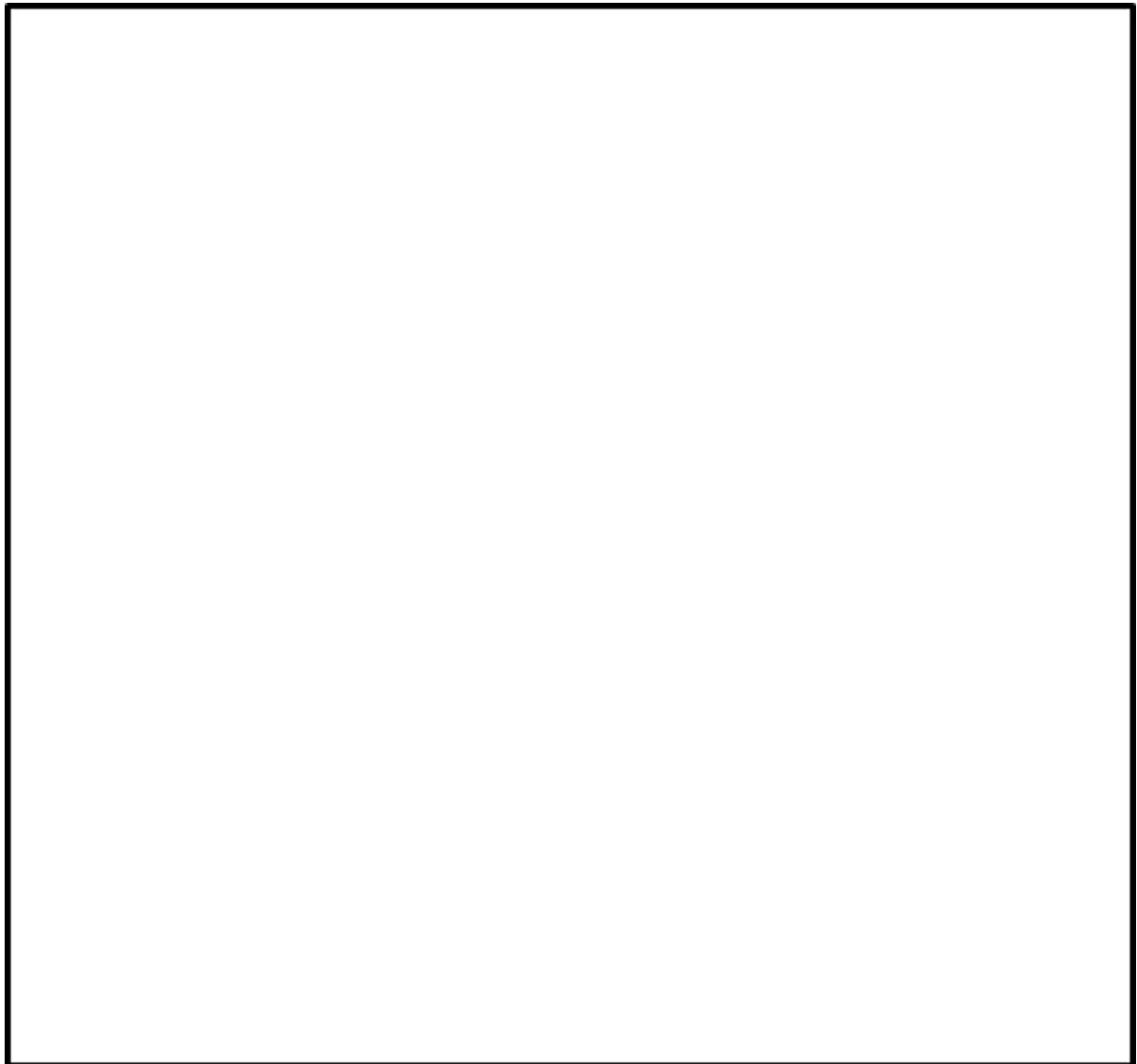
Appendix F: Methodology checklist: the QUADAS-2 tool for studies of diagnostic test accuracy

The following checklists are taken directly from the 'QUADAS-2' publication (Whiting PF Rutjes AWS, Westwood ME et al. and the QUADAS-2 group [2011] QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 155: 529–36) and from the [QUADAS website](#).

Phase 1: State the review question

Patients (setting, intended use of index test, presentation, prior testing): .
Index test(s): .
Reference standard and target condition: .

Phase 2: Draw a flow diagram for the primary study



Phase 3: Risk of bias and applicability judgements

QUADAS-2 is structured so that four key domains are each rated in terms of the risk of bias and the concern regarding applicability to the review question (as stated in Phase 1). Each key domain has a set of signalling questions to help reach the judgements regarding bias and applicability.

Domain 1: Patient selection

A. Risk of bias

Describe methods of patient selection:	
<ul style="list-style-type: none"> · · · · 	
Was a consecutive or random sample of patients enrolled?	Yes / No / Unclear
Was a case-control design avoided?	Yes / No / Unclear
Did the study avoid inappropriate exclusions?	Yes / No / Unclear
<p>Could the selection of patients have introduced bias?</p> <p>Risk: Low / High / Unclear</p>	
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting):	
<ul style="list-style-type: none"> · · · 	
<p>Is there concern that the included patients do not match the review question?</p> <p>Concern: Low / High / Unclear</p>	

Domain 2: Index test(s)

A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	
<ul style="list-style-type: none"> · · · · 	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes / No / Unclear
If a threshold was used, was it pre-specified?	Yes / No / Unclear

<p>Could the conduct or interpretation of the index test have introduced bias? Risk: Low / High / Unclear</p>
<p>B. Concerns regarding applicability</p>
<p>Is there concern that the index test, its conduct, or interpretation differ from the review question? Concern: Low / High / Unclear</p>

Domain 3: Reference standard

<p>A. Risk of bias</p>	
<p>Describe the reference standard and how it was conducted and interpreted:</p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>	
<p>Is the reference standard likely to correctly classify the target condition?</p>	<p>Yes / No / Unclear</p>
<p>Were the reference standard results interpreted without knowledge of the results of the index test?</p>	<p>Yes / No / Unclear</p>
<p>Could the reference standard, its conduct, or its interpretation have introduced bias? Risk: Low / High / Unclear</p>	
<p>B. Concerns regarding applicability</p>	
<p>Is there concern that the target condition as defined by the reference standard does not match the review question? Concern: Low / High / Unclear</p>	

Domain 4: Flow and timing

<p>A. Risk of bias</p>

<p>Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):</p> <p>.</p> <p>.</p> <p>.</p> <p>Describe the time interval and any interventions between index test(s) and reference standard:</p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>	
Was there an appropriate interval between index test(s) and reference standard?	Yes / No / Unclear
Did all patients receive a reference standard?	Yes / No / Unclear
Did patients receive the same reference standard?	Yes / No / Unclear
Were all patients included in the analysis?	Yes / No / Unclear
<p>Could the patient flow have introduced bias?</p> <p>Risk: Low / High / Unclear</p>	

Notes on use of Methodology checklist: the QUADAS-2 tool for studies of diagnostic test accuracy

For the accompanying notes on how to use the QUADAS-2 tool, please see:

- Whiting PF, Rutjes AWS, Westwood ME et al. and the QUADAS-2 group (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 155: 529–36
- The [QUADAS website](#)

Appendix G: Methodology checklist: economic evaluations

This checklist is designed to determine whether an economic evaluation provides evidence that is useful to inform the decision-making of the Guideline Development Group (GDG) (see [chapter 7](#)). It is not intended to judge the quality of the study per se or the quality of reporting.

Checklist

Study identification <i>Include author, title, reference, year of publication</i>		
Guideline topic:		Question no:
Checklist completed by:		
Section 1: Applicability (relevance to specific guideline review question(s) and the NICE reference case^[a]) <i>This checklist should be used first to filter out irrelevant studies.</i>	Yes/ Partly/ No /Unclear /NA	Comments
<u>1.1</u> Is the study population appropriate for the guideline?		
<u>1.2</u> Are the interventions and services appropriate for the guideline?		
<u>1.3</u> Is the healthcare system in which the study was conducted sufficiently similar to the current UK NHS context?		
<u>1.4</u> Are costs measured from the NHS and personal social services (PSS) perspective?		
<u>1.5</u> Are non-direct health effects on individuals excluded?		
<u>1.6</u> Are both costs and health effects discounted at an annual rate of 3.5%?		
<u>1.7</u> Is the value of health effects expressed in terms of quality-adjusted life years (QALYs)?		

<p><u>1.8</u> Are changes in health-related quality of life (HRQoL) reported directly from patients and/or carers?</p>		
<p><u>1.9</u> Is the valuation of changes in HRQoL (utilities) obtained from a representative sample of the general public?</p>		
<p><u>1.10</u> Overall judgement: Directly applicable/Partially applicable/Not applicable There is no need to use section 2 of the checklist if the study is considered 'not applicable'.</p>		
<p>Other comments:</p> <p>.</p> <p>.</p> <p>.</p> <p>.</p>		
<p>Section 2: Study limitations (the level of methodological quality) <i>This checklist should be used once it has been decided that the study is sufficiently applicable to the context of the clinical guideline^[6].</i></p>	<p>Yes/ Partly /No/ Unclear/ NA</p>	<p>Comments</p>
<p><u>2.1</u> Does the model structure adequately reflect the nature of the health condition under evaluation?</p>		
<p><u>2.2</u> Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?</p>		
<p><u>2.3</u> Are all important and relevant health outcomes included?</p>		
<p><u>2.4</u> Are the estimates of baseline health outcomes from the best available source?</p>		
<p><u>2.5</u> Are the estimates of relative treatment effects from the best available source?</p>		
<p><u>2.6</u> Are all important and relevant costs included?</p>		
<p><u>2.7</u> Are the estimates of resource use from the best available source?</p>		

<u>2.8</u> Are the unit costs of resources from the best available source?			
<u>2.9</u> Is an appropriate incremental analysis presented or can it be calculated from the data?			
<u>2.10</u> Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?			
<u>2.11</u> Is there no potential conflict of interest?			
<u>2.12</u> Overall assessment: Minor limitations/Potentially serious limitations/Very serious limitations			
Other comments:			
<p>^[a] As detailed in chapter 5 of NICE's Guide to the methods of technology appraisal. The guide notes that there may be important barriers to applying reference-case methods, and in these cases the reasons for not applying reference-case methods should be clearly specified and justified, and the likely implications should, as far as possible, be quantified.</p> <p>^[b] The items and notes in this checklist have been developed from guidance in NICE's Guide to the methods of technology appraisal; Evers S, Goossens M, de Vet H et al. (2005) Criteria list for assessment of methodological quality of economic evaluations – CHEC. <i>International Journal of Technology Assessment in Health Care</i> 21:240–5; and Philips Z, Ginnelly L, Sculpher M et al. (2004) Review of guidelines for good practice in decision-analytic modelling in health technology assessment. <i>Health Technology Assessment</i> 8.</p>			

Notes on use of Methodology checklist: economic evaluations

For all questions:

- answer 'yes' if the study fully meets the criterion
- answer 'partly' if the study largely meets the criterion but differs in some important respect
- answer 'no' if the study deviates substantively from the criterion

- answer 'unclear' if the report provides insufficient information to judge whether the study complies with the criterion
- answer 'NA (not applicable)' if the criterion is not relevant in a particular instance.

For 'partly' or 'no' responses, use the comments column to explain how the study deviates from the criterion.

Section 1: applicability

1.1 Is the study population appropriate for the guideline?

The study population should be defined as precisely as possible and should be in line with that specified in the guideline scope and any related review protocols.

This includes consideration of appropriate subgroups that require special attention. For many interventions, the capacity to benefit will differ for participants with differing characteristics. This should be explored separately for each relevant subgroup as part of the base-case analysis by the provision of estimates of clinical and cost effectiveness. The characteristics of participants in each subgroup should be clearly defined and, ideally, should be identified on the basis of an a priori expectation of differential clinical or cost effectiveness as a result of biologically plausible known mechanisms, social characteristics or other clearly justified factors.

Answer 'yes' if the study population is fully in line with that in the guideline question(s) and if the study differentiates appropriately between important subgroups. Answer 'partly' if the study population is similar to that in the guideline question(s) but: (i) it differs in some important respects; or (ii) the study fails to differentiate between important subgroups. Answer 'no' if the study population is substantively different from that in the guideline question(s).

1.2 Are the interventions and services appropriate for the guideline?

All relevant alternatives should be included, as specified in the guideline scope and any related review protocols. These should include routine and best practice in the NHS, existing NICE guidance and other feasible options.

Answer 'yes' if the analysis includes all options considered relevant for the guideline, even if it also includes other options that are not relevant. Answer 'partly' if the analysis omits one or more relevant options but still contains comparisons likely to be useful for the guideline. Answer 'no' if the analysis does not contain any relevant comparisons.

1.3 Is the healthcare system in which the study was conducted sufficiently similar to the current UK NHS context?

This relates to the overall structure of the healthcare system within which the interventions were delivered. For example, an intervention might be delivered on an inpatient basis in one country whereas in the UK it would be provided in the community. This might significantly influence the use of healthcare resources and costs, thus limiting the applicability of the results to a UK setting. In addition, old UK studies may be severely limited in terms of their relevance to current NHS practice.

Answer 'yes' if the study was conducted within the UK and is sufficiently recent to reflect current NHS practice. For non-UK or older UK studies, answer 'partly' if differences in the healthcare setting are unlikely to substantively change the cost-effectiveness estimates. Answer 'no' if the healthcare setting is so different that the results are unlikely to be applicable in the current NHS.

1.4 Are costs measured from the NHS and personal social services (PSS) perspective?

The decision-making perspective of an economic evaluation determines the range of costs that should be included in the analysis. NICE works in a specific context; in particular, it does not set the budget for the NHS. The objective of NICE is to offer guidance that represents an efficient use of available NHS and PSS resources. For these reasons, the perspective on costs used in the NICE reference case is that of the NHS and PSS. Productivity costs and costs borne by patients and carers that are not reimbursed by the NHS or PSS are not included in the reference case. The reference case also excludes costs to other government bodies, although these may sometimes be presented in additional analyses alongside the reference case.

Answer 'yes' if the study only includes costs for resource items that would be paid for by the NHS and PSS. Also answer 'yes' if other costs have been included in the study, but the results are presented in such a way that the cost effectiveness can be calculated from an NHS and PSS perspective. Answer 'partly' if the study has taken a wider perspective but the other non-NHS/PSS costs are small in relation to the total expected costs and are unlikely to change the cost-effectiveness results. Answer 'no' if non-NHS/PSS costs are significant and are likely to change the cost-effectiveness results.

Some interventions may have a substantial impact on non-health outcomes or costs to other government bodies (for example, treatments to reduce illicit drug misuse may have the effect of reducing drug-related crime). In such situations, if the economic study includes non-health costs in such a way that they cannot be separated out from NHS/PSS costs, answer 'no' but consider

retaining the study for critical appraisal. If studies containing non-reference-case costs are retained, use the comments column to note why.

1.5 Are non-direct health effects on individuals excluded?

In the NICE reference case, the perspective on outcomes should be all direct health effects, whether for patients or, when relevant, other people (principally carers). This is consistent with an objective of maximising health gain from available healthcare resources. Some features of healthcare delivery that are often referred to as 'process characteristics' may ultimately have health consequences; for example, the mode of treatment delivery may have health consequences through its impact on concordance with treatment. Any significant characteristics of healthcare technologies that have a value to people that is independent of any direct effect on health should be noted. These characteristics include the convenience with which healthcare is provided and the level of information available for patients.

This question should be viewed in terms of what is excluded in relation to the NICE reference case; that is, non-health effects.

Answer 'yes' if the measure of health outcome used in the analysis excludes non-health effects (or if such effects can be excluded from the results). Answer 'partly' if the analysis includes some non-health effects but these are small and unlikely to change the cost-effectiveness results. Answer 'no' if the analysis includes significant non-health effects that are likely to change the cost-effectiveness results.

1.6 Are both costs and health effects discounted at an annual rate of 3.5%?

The need to discount to a present value is widely accepted in economic evaluation, although the specific rate varies across jurisdictions and over time. NICE considers it appropriate to discount costs and health effects at the same rate. The annual rate of 3.5%, based on the recommendations of the UK Treasury for the discounting of costs, applies to both costs and health effects.

Answer 'yes' if both costs and health effects (for example, quality-adjusted life years [QALYs]) are discounted at 3.5% per year. Answer 'partly' if costs and health effects are discounted at a rate similar to 3.5% (for example, costs and effects are both discounted at 3% per year). Answer 'no' if costs and/or health effects are not discounted, or if they are discounted at a rate (or rates) different from 3.5% (for example, 5% for both costs and effects, or 6% for costs and 1.5% for effects). Note in the comments column what discount rates have been used. If all costs and health effects accrue within a short time (roughly a year), answer 'NA'.

1.7 Is the value of health effects expressed in terms of quality-adjusted life years (QALYs)?

The QALY is a measure of a person's length of life weighted by a valuation of their health-related quality of life (HRQoL) over that period.

Given its widespread use, the QALY is considered by NICE to be the most appropriate generic measure of health benefit that reflects both mortality and effects on HRQoL. It is recognised that alternative measures exist (such as the healthy-year equivalent), but few economic evaluations have used these methods and their strengths and weaknesses are not fully established.

NICE's position is that an additional QALY should be given the same weight regardless of the other characteristics of the patients receiving the health benefit.

Answer 'yes' if the effectiveness of the intervention is measured using QALYs; answer 'no' if not. There may be circumstances when a QALY cannot be obtained or where the assumptions underlying QALYs are considered inappropriate. In such situations answer 'no', but consider retaining the study for appraisal. Similarly, answer 'no' but retain the study for appraisal if it does not include QALYs but it is still thought to be useful for GDG decision-making: for example, if the clinical evidence indicates that an intervention might be dominant, and estimates of the relative costs of the interventions from a cost-minimisation study are likely to be useful. When economic evaluations not using QALYs are retained for full critical appraisal, use the comments column to note why.

1.8 Are changes in health-related quality of life (HRQoL) reported directly from patients and/or carers?

In the NICE reference case, information on changes in HRQoL as a result of treatment should be reported directly by patients (and directly by carers when the impact of treatment on the carer's health is also important). When it is not possible to obtain information on changes in patients' HRQoL directly from them, data should be obtained from carers (not from healthcare professionals).

For consistency, the EQ-5D is NICE's preferred measure of HRQoL in adults. However, when EQ-5D data are not available or are inappropriate for the condition or the effects of treatment, other multi-attribute utility questionnaires (for example, SF6D, QWB or HUI) or mapping methods from disease-specific questionnaires may be used to estimate QALYs. For studies not reporting QALYs, a variety of generic or disease-specific methods may be used to measure HRQoL.

Answer 'yes' if changes in patients' HRQoL are estimated by the patients themselves. Answer 'partly' if estimates of patients' HRQoL are provided by carers. Answer 'no' if estimates come from healthcare professionals or researchers. Note in the comments column how HRQoL was measured (EQ-5D, QWB, HUI and so on). Answer 'NA' if the cost-effectiveness study does not include estimates of HRQoL (for example, studies reporting 'cost per life year gained' or cost-minimisation studies).

1.9 Is the valuation of changes in HRQoL (utilities) obtained from a representative sample of the general public?

The NICE reference case specifies that the valuation of changes in HRQoL (utilities) reported by patients should be based on public preferences elicited using a choice-based method (such as the time trade-off or standard gamble) in a representative sample of the UK population.

Answer 'yes' if HRQoL valuations were obtained using the EQ-5D UK tariff. Answer 'partly' if the valuation methods were comparable to those used for the EQ-5D. Answer 'no' if other valuation methods were used. Answer 'NA' if the study does not apply valuations to HRQoL (for studies not reporting QALYs). In the comments column note the valuation method used (such as time trade-off or standard gamble) and the source of the preferences (such as patients or healthcare professionals).

1.10 Overall judgement

Classify the applicability of the economic evaluation to the clinical guideline, the current NHS situation and the context for NICE guidance as one of the following:

- **Directly applicable** – the study meets all applicability criteria, or fails to meet one or more applicability criteria but this is unlikely to change the conclusions about cost effectiveness.
- **Partially applicable** – the study fails to meet one or more applicability criteria, and this could change the conclusions about cost effectiveness.
- **Not applicable** – the study fails to meet one or more applicability criteria, and this is likely to change the conclusions about cost effectiveness. Such studies would usually be excluded from further consideration and there is no need to continue with the rest of the checklist.

Section 2: study limitations

2.1 Does the model structure adequately reflect the nature of the health condition under evaluation?

This relates to the choice of model and its structural elements (including cycle length in discrete time models, if appropriate). Model type and its structural aspects should be consistent with a coherent theory of the health condition under evaluation. The selection of treatment pathways, whether health states or branches in a decision tree, should be based on the underlying biological processes of the health issue under study and the potential impact (benefits and adverse consequences) of the intervention(s) of interest.

Answer 'yes' if the model design and assumptions appropriately reflect the health condition and intervention(s) of interest. Answer 'partly' if there are aspects of the model design or assumptions that do not fully reflect the health condition or intervention(s) but these are unlikely to change the cost-effectiveness results. Answer 'no' if the model omits some important aspect of the health condition or intervention(s) and this is likely to change the cost-effectiveness results. Answer 'NA' for economic evaluations based on data from a clinical study which do not extrapolate treatment outcomes or costs beyond the study context or follow-up period.

2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?

The time horizon is the period of analysis of the study: the length of follow-up for participants in a trial-based evaluation, or the period of time over which the costs and outcomes for a cohort are tracked in a modelling study. This time horizon should always be the same for costs and outcomes, and should be long enough to include all relevant costs and outcomes relating to the intervention. A time horizon shorter than lifetime could be justified if there is no differential mortality effect between options, and the differences in costs and HRQoL relate to a relatively short period (for example, in the case of an acute infection).

Answer 'yes' if the time horizon is sufficient to include all relevant costs and outcomes. Answer 'partly' if the time horizon may omit some relevant costs and outcomes but these are unlikely to change the cost-effectiveness results. Answer 'no' if the time horizon omits important costs and outcomes and this is likely to change the cost-effectiveness results.

2.3 Are all important and relevant health outcomes included?

All relevant health outcomes should include direct health effects relating to harms from the intervention (adverse effects) as well as any potential benefits.

Answer 'yes' if the analysis includes all relevant and important harms and benefits. Answer 'partly' if the analysis omits some harms or benefits but these would be unlikely to change the cost-effectiveness results. Answer 'no' if the analysis omits important harms and/or benefits that would be likely to change the cost-effectiveness results.

2.4 Are the estimates of baseline health outcomes from the best available source?

The estimate of the overall net treatment effect of an intervention is determined by the baseline risk of a particular condition or event and/or the relative effects of the intervention compared with the relevant comparator treatment. The overall net treatment effect may also be determined by other features of the people comprising the population of interest.

The process of assembling evidence for economic evaluations should be systematic – evidence must be identified, quality assessed and, when appropriate, pooled, using explicit criteria and justifiable and reproducible methods. These principles apply to all categories of evidence that are used to estimate clinical and cost effectiveness, evidence for which will typically be drawn from a number of different sources.

The sources and methods for eliciting baseline probabilities should be described clearly. These data can be based on 'natural history' (patient outcomes in the absence of treatment or with routine care), sourced from cohort studies. Baseline probabilities may also be derived from the control arms of experimental studies. Sometimes it may be necessary to rely on expert opinion for particular parameters.

Answer 'yes' if the estimates of baseline health outcomes reflect the best available evidence as identified from a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates are not derived from a systematic review but are likely to reflect outcomes for the relevant group of patients in routine NHS practice (for example, if they are derived from a large UK-relevant cohort study). Answer 'no' if the estimates are unlikely to reflect outcomes for the relevant group in routine NHS practice.

2.5 Are the estimates of relative treatment effects from the best available source?

The objective of the analysis of clinical effectiveness is to produce an unbiased estimate of the mean clinical effectiveness of the interventions being compared.

The NICE reference case indicates that evidence on outcomes should be obtained from a systematic review, defined as the systematic location, inclusion, appraisal and synthesis of evidence to obtain a reliable and valid overview of the data relating to a clearly formulated question.

Synthesis of outcome data through meta-analysis is appropriate provided that there are sufficient relevant and valid data obtained using comparable measures of outcome.

Head-to-head randomised controlled trials (RCTs) provide the most valid evidence of relative treatment effect. However, such evidence may not always be available. Therefore, data from non-randomised studies may be required to supplement RCT data. Any potential bias arising from the design of the studies used in the assessment should be explored and documented.

Data from head-to-head RCTs should be presented in the base-case analysis, if available. When head-to-head RCTs exist, evidence from indirect or mixed treatment comparison analyses may be presented if it is considered to add information that is not available from the head-to-head comparison. This indirect or mixed treatment comparison must be fully described and presented as additional to the base-case analysis. (A 'mixed treatment comparison' estimates effect sizes using both head-to-head and indirect comparisons.)

If data from head-to-head RCTs are not available, indirect treatment comparison methods should be used. (An 'indirect treatment comparison' is a synthesis of data from a network of trials that compare the interventions of interest with other comparators.)

When multiple interventions are being assessed that have not been compared within a single RCT, data from a series of pairwise head-to-head RCTs should be presented. Consideration should also be given to presenting a combined analysis using a mixed treatment comparison framework if it is considered to add information that is not available from the head-to-head comparison.

Only indirect or mixed treatment comparison methods that preserve randomisation should be used. The principles of good practice for standard meta-analyses should also be followed in mixed and indirect treatment comparisons.

The methods and assumptions that are used to extrapolate short-term results to final outcomes should be clearly presented and there should be documentation of the reasoning underpinning the choice of survival function.

Evidence for the evaluation of diagnostic technologies should normally incorporate evidence on diagnostic accuracy. It is also important to incorporate the predicted changes in health outcomes and costs resulting from treatment decisions based on the test result. The general principles guiding the assessment of the clinical and cost effectiveness of diagnostic interventions should be the same as for other technologies. However, particular consideration of the methods of analysis may be required, particularly in relation to evidence synthesis. Evidence for the effectiveness of diagnostic technologies should include the costs and outcomes for people whose test results lead to an incorrect diagnosis, as well as for those who are diagnosed correctly.

As for other technologies, RCTs have the potential to capture the pathway of care involving diagnostic technologies, but their feasibility and availability may be limited. Other study designs should be assessed on the basis of their fitness for purpose, taking into consideration the aim of the study (for example, to evaluate outcomes, or to evaluate sensitivity and specificity) and the purpose of the diagnostic technology.

Answer 'yes' if the estimates of treatment effect appropriately reflect all relevant studies of the best available quality, as identified through a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates of treatment effect are not derived from a systematic review but are similar in magnitude to the best available estimates (for example, if the economic evaluation is based on a single large study with treatment effects similar to pooled estimates from all relevant studies). Answer 'no' if the estimates of treatment effect are likely to differ substantively from the best available estimates.

2.6 Are all important and relevant costs included?

Costs related to the condition of interest and incurred in additional years of life gained as a result of treatment should be included in the base-case analysis. This should include the costs of handling non-adherence to treatment and treating side effects. Costs that are considered to be unrelated to the condition or intervention of interest should be excluded. If introduction of the intervention requires additional infrastructure to be put in place, consideration should be given to including such costs in the analysis.

Answer 'yes' if all important and relevant resource use and costs are included given the perspective and the research question in the economic study under consideration. Answer 'partly' if some

relevant resource items are omitted but these are unlikely to affect the cost-effectiveness results. Answer 'no' if important resource items are omitted and these are likely to affect the cost-effectiveness results.

2.7 Are the estimates of resource use from the best available source?

It is important to quantify the effect of the interventions on resource use in terms of physical units (for example, days in hospital or visits to a GP) and valuing those effects in monetary terms using appropriate prices and unit costs. Evidence on resource use should be identified systematically. When expert opinion is used as a source of information, any formal methods used to elicit these data should be clearly reported.

Answer 'yes' if the estimates of resource use appropriately reflect all relevant evidence sources of the best available quality, as identified through a recent well-conducted systematic review of the literature. Answer 'partly' if the estimates of resource use are not derived from a systematic review but are similar in magnitude to the best available estimates. Answer 'no' if the estimates of resource use are likely to differ substantively from the best available estimates.

2.8 Are the unit costs of resources from the best available source?

Resources should be valued using the prices relevant to the NHS and PSS. Given the perspective of the NICE reference case, it is appropriate for the financial costs relevant to the NHS/PSS to be used as the basis of costing, although these may not always reflect the full social opportunity cost of a given resource. A first point of reference in identifying costs and prices should be any current official listing published by the Department of Health and/or the Welsh Government.

When the acquisition price paid for a resource differs from the public list price (for example, pharmaceuticals and medical devices sold at reduced prices to NHS institutions), the public list price should be used in the base-case analysis. Sensitivity analysis should assess the implications of variations from this price. Analyses based on price reductions for the NHS will only be considered when the reduced prices are transparent and can be consistently available across the NHS, and if the period for which the specified price is available is guaranteed.

National data based on healthcare resource groups (HRGs) such as the Payment by Results tariff can be used when they are appropriate and available. However, data based on HRGs may not be appropriate in all circumstances (for example, when the definition of the HRG is broad, or the mean cost probably does not reflect resource use in relation to the intervention(s) under consideration). In such cases, other sources of evidence, such as micro-costing studies, may be more appropriate. When cost data are taken from the literature, the methods used to identify the sources should be

defined. When several alternative sources are available, a justification for the costs chosen should be provided and discrepancies between the sources explained. When appropriate, sensitivity analysis should have been undertaken to assess the implications for results of using alternative data sources.

Answer 'yes' if resources are valued using up-to-date prices relevant to the NHS and PSS. Answer 'partly' if the valuations of some resource items differ from current NHS/PSS unit costs but this is unlikely to change the cost-effectiveness results. Answer 'no' if the valuations of some resource items differ substantively from current NHS/PSS unit costs and this is likely to change the cost-effectiveness results.

2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?

An appropriate incremental analysis is one that compares the expected costs and health outcomes of one intervention with the expected costs and health outcomes of the next-best non-dominated alternative.

Standard decision rules should be followed when combining costs and effects, and should reflect any situation where there is dominance or extended dominance. When there is a trade-off between costs and effects, the results should be presented as an incremental cost-effectiveness ratio (ICER): the ratio of the difference in mean costs to the difference in mean outcomes of a technology compared with the next best alternative. In addition to ICERs, expected net monetary or health benefits can be presented using values placed on a QALY gained of £20,000 and £30,000.

For cost-consequence analyses, appropriate incremental analysis can only be done by selecting one of the consequences as the primary measure of effectiveness.

Answer 'yes' if appropriate incremental results are presented, or if data are presented that allow the reader to calculate the incremental results. Answer 'no' if: (i) simple ratios of costs to effects are presented for each alternative compared with a standard intervention; or (ii) if options subject to simple or extended dominance are not excluded from the incremental analyses.

2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?

There are a number of potential selection biases and uncertainties in any evaluation (trial- or model-based) and these should be identified and quantified where possible. There are three types of bias or uncertainty to consider:

- Structural uncertainty – for example in relation to the categorisation of different states of health and the representation of different pathways of care. These structural assumptions should be clearly documented and the evidence and rationale to support them provided. The impact of structural uncertainty on estimates of cost effectiveness should be explored by separate analyses of a representative range of plausible scenarios.
- Source of values to inform parameter estimates – the implications of different estimates of key parameters (such as estimates of relative effectiveness) must be reflected in sensitivity analyses (for example, through the inclusion of alternative scenarios). Inputs must be fully justified, and uncertainty explored by sensitivity analysis using alternative input values.
- Parameter precision – uncertainty around the mean health and cost inputs in the model. Distributions should be assigned to characterise the uncertainty associated with the (precision of) mean parameter values. Probabilistic sensitivity analysis is preferred, as this enables the uncertainty associated with parameters to be simultaneously reflected in the results of the model. In non-linear decision models – when there is not a straight-line relationship between inputs and outputs of a model (such as Markov models) – probabilistic methods provide the best estimates of mean costs and outcomes. Simple decision trees are usually linear. The mean value, distribution around the mean, and the source and rationale for the supporting evidence should be clearly described for each parameter included in the model. Evidence about the extent of correlation between individual parameters should be considered carefully and reflected in the probabilistic analysis. Assumptions made about the correlations should be clearly presented.

Answer 'yes' if an extensive sensitivity analysis was undertaken that explored all key uncertainties in the economic evaluation. Answer 'partly' if the sensitivity analysis failed to explore some important uncertainties in the economic evaluation. Answer 'no' if the sensitivity analysis was very limited and omitted consideration of a number of important uncertainties, or if the range of values or distributions around parameters considered in the sensitivity analysis were not reported.

2.11 Is there no potential conflict of interest?

The British Medical Journal (BMJ) defines competing interests for its authors as follows: "A competing interest exists when professional judgment concerning a primary interest (such as patients' welfare or the validity of research) may be influenced by a secondary interest (such as financial gain or personal rivalry). It may arise for the authors of a BMJ article when they have a financial interest that may influence, probably without their knowing, their interpretation of their results or those of others."

Whenever a potential financial conflict of interest is possible, this should be declared.

Answer 'yes' if the authors declare that they have no financial conflicts of interest. Answer 'no' if clear financial conflicts of interest are declared or apparent (for example, from the stated affiliation of the authors). Answer 'unclear' if the article does not indicate whether or not there are financial conflicts of interest.

2.12 Overall assessment

The overall methodological study quality of the economic evaluation should be classified as one of the following:

- **Minor limitations** – the study meets all quality criteria, or fails to meet one or more quality criteria but this is unlikely to change the conclusions about cost effectiveness.
- **Potentially serious limitations** – the study fails to meet one or more quality criteria, and this could change the conclusions about cost effectiveness.
- **Very serious limitations** – the study fails to meet one or more quality criteria, and this is highly likely to change the conclusions about cost effectiveness. Such studies should usually be excluded from further consideration.

Supporting references

National Institute for Health and Clinical Excellence (2008) [Social value judgements: principles for the development of NICE guidance \(second edition\)](#). London: National Institute for Health and Clinical Excellence

Philips Z, Ginnelly L, Sculpher M et al. (2004) [Review of guidelines for good practice in decision-analytic modelling in health technology assessment](#). Health Technology Assessment 8 (36)

Evers, S, Goossens M, de Vet H et al. (2005) [Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria](#). International Journal of Technology Assessment in Health Care 21: 240–5

Six workshops were held to enable NICE to explore and capture different perspectives on specific questions as part of the 2007 review of the 'Guide to the methods of technology appraisal'. Documents listed below include briefing papers that were produced to facilitate discussion at each of the workshops and working party meetings:

- [costs](#)
- [diagnostic technologies](#)
- [evidence synthesis \(indirect and mixed treatment comparisons\)](#)
- [identifying subgroups and exploring heterogeneity](#)
- [threshold](#)
- [exploring uncertainty](#)
- [health-related utility measurement](#).

These documents are available from the [NICE website](#).

Appendix H: Methodology checklist: qualitative studies

This checklist is based on checklists from:

Spencer L, Ritchie J, Lewis J et al. (2003) Quality in qualitative evaluation: a framework for assessing research evidence. London: Government Chief Social Researcher's Office

Public Health Resource Unit England (2006) Critical Appraisal Skills Programme (CASP) – making sense of evidence: 10 questions to help you make sense of qualitative research

National Training and Research Appraisal Group (NTRAG); contact: info@ntrag.co.uk

British Sociological Association (BSA)

Checklist

Study identification Include author, title, reference, year of publication	
Guidance topic:	Key research question/aim: .
Checklist completed by:	

	<i>Circle or highlight one option for each question</i>	
Section 1: theoretical approach		
<p>1.1 Is a qualitative approach appropriate?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> Does the research question seek to understand processes or structures, or illuminate subjective experiences or meanings (in healthcare this would apply to how care is organised and patient experiences of care)? Or could a quantitative approach better have addressed the research question? 	<p>Appropriate</p> <p>Inappropriate</p> <p>Not sure</p>	<p>Comments:</p>

<p>1.2 Is the study clear in what it seeks to do?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Is the purpose of the study discussed – aims/objectives/ research question(s)? • Are the values/assumptions/theory underpinning the purpose of the study discussed? 	<p>Clear Unclear Mixed</p>	<p>Comments:</p>
<p>Section 2: study design</p>		
<p>2.1 How defensible/rigorous is the research design/ methodology?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Are there clear accounts of the rationale/justification for the sampling, data collection and data analysis techniques used? 	<p>Defensible Not defensible Not sure</p>	<p>Comments:</p>
<p>Section 3: data collection</p>		
<p>3.1 How well was the data collection carried out?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Are the data collection methods clearly described? • Were the data collected appropriate to address the research question? 	<p>Appropriate Inappropriate Not sure/ inadequately reported</p>	<p>Comments:</p>
<p>Section 4: validity</p>		
<p>4.1 Is the context clearly described?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Are the characteristics of the participants and settings clearly defined? • Were observations made in a variety of circumstances and from a range of respondents? • Was context bias considered (that is, did the authors consider the influence of the setting where the study took place)? 	<p>Clear Unclear Not sure</p>	<p>Comments:</p>

<p>4.2 Were the methods reliable?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Were data collected by more than one method? • Were other studies considered with discussion about similar/different results? 	<p>Reliable</p> <p>Unreliable</p> <p>Not sure</p>	<p>Comments:</p>
<p>Section 5: analysis</p>		
<p>5.1 Are the data 'rich'?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • How well are the contexts of the data described? • Has the diversity of perspective and content been explored? • Has the detail of the data that were collected been demonstrated? • Are responses compared and contrasted across groups/sites? 	<p>Rich</p> <p>Poor</p> <p>Not sure/not reported</p>	<p>Comments:</p>
<p>5.2 Is the analysis reliable?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Did more than one researcher theme and code transcripts/data? • If so, how were differences resolved? • Were negative/discrepant results addressed or ignored? • Is it clear how the themes and concepts were derived from the data? 	<p>Reliable</p> <p>Unreliable</p> <p>Not sure/not reported</p>	<p>Comments:</p>

<p>5.3 Are the findings convincing?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Are the findings clearly presented? • Are the findings internally coherent (that is, are the results credible in relation to the study question)? • Are extracts from the original data included (for example, direct quotes from participants)? • Are the data appropriately referenced so that the sources of the extracts can be identified? • Is the reporting clear and coherent? 	<p>Convincing</p> <p>Not convincing</p> <p>Not sure</p>	<p>Comments:</p>
<p>5.4 Are the conclusions adequate?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • How clear are the links between data, interpretation and conclusions? • Are the conclusions plausible and coherent? • Have alternative explanations been explored and discounted? • Are the implications of the research clearly defined? • Is there adequate discussion of any limitations encountered? 	<p>Adequate</p> <p>Inadequate</p> <p>Not sure</p>	<p>Comments:</p>
<p>Section 6: ethics</p>		
<p>6.1 Was the study approved by an ethics committee?</p>	<p>Yes</p> <p>No</p> <p>Not sure/not reported/not applicable</p>	<p>Comments:</p>

<p>6.2 Is the role of the researcher clearly described?</p> <p><i>For example:</i></p> <ul style="list-style-type: none"> • Has the relationship between the researcher and the participants been adequately described? • Is how the research was explained and presented to the participants described? 	<p>Clear</p> <p>Not clear</p> <p>Not sure/not reported</p>	<p>Comments:</p>
---	--	------------------

After completion of quality appraisal using the checklist, the included studies can be presented in a 'Quality of the included studies' table, which summarises the quality of each study under the main criteria of population, methods and analysis, and also the relevance of the study to the population being considered in the guideline.

Table 1 is an example table from the full guideline for [Management of stable angina](#) (NICE clinical guideline 126) for the review question 'What are the information needs of people with stable angina?'. In order to answer this question, three qualitative studies were included and assessed for quality.

Table 1. Quality of the included studies (from [Management of stable angina](#) [NICE clinical guideline 126])

Study	Population	Methods	Analysis	Relevance to guideline population
Pier 2008	Well reported	Well reported	Well reported and credible	Australia. Patients with myocardial infarction, coronary artery bypass graft, angioplasty or angina from GP practices.
Weech 2003	Poorly reported	Poorly reported	Poorly reported	UK population. Patients suffering from angina who had been hospitalised in the coronary care ward.
McGillion 2004	Well reported	Well reported	Well reported and credible	Canada. Chronic stable angina patients living at home.

Notes on use of Methodology checklist: qualitative studies

The studies covered by this checklist are those that collect and analyse qualitative data – usually (but not exclusively) textual (written), spoken or observational data. Qualitative data are occasionally collected using structured questionnaires (for example, as thematically organised free-text comments) but such research needs to be scrutinised carefully because it may not meet acceptable quality criteria for consideration as a qualitative study.

There is considerable debate over which quality criteria should be used to assess qualitative studies. Quality in qualitative research can be assessed using the same broad concepts of validity (or trustworthiness) used for quantitative research, but these need to be put in a different contextual framework to take into account the aims of qualitative research. This checklist is based on the broadly accepted principles that characterise qualitative research and that may affect its validity; it is concerned with adequate reporting of key factors that affect the quality of qualitative research studies. The questions in the checklist are framed to encompass the variety of ways in which qualitative research is conducted. Care must be taken to apply the checklist in a way that matches the research methodology.

The following notes provide suggestions for completing the checklist. A list of publications on qualitative research is provided at the end of these notes for further reading on this topic.

Note that the sub-questions given as examples under each question in the checklist are intended to highlight some of the key issues to be considered for that question – they are not intended to be exhaustive. Please add any additional considerations in the comments box.

Section 1: theoretical approach

This section deals with the underlying theory and principles applied to the research.

1.1 Is a qualitative approach appropriate?

A qualitative approach can be judged to be appropriate when the research sets out to investigate phenomena that are not easy to quantify or measure accurately, or where such measurement would be arbitrary and inexact.

Qualitative research in a health setting commonly measures:

- personal experiences (for example, of a condition, treatment or situation)

- processes (for example, action research, practitioner or patient views on the acceptability of using new technology)
- personal values and beliefs (for example, about death, birth, disability)
- interactions and relationships (for example, the quality of the GP–patient relationship, the openness of a psychotherapeutic relationship)
- service evaluations (for example, what was good or bad about patients' experiences of a smoking cessation group).

If clear numerical measures could reasonably have been put in place, then consider whether a quantitative approach may have been more appropriate.

1.2 Is the study clear in what it seeks to do?

The design of qualitative research tends to be 'theory generative' rather than 'theory testing'; it is therefore unlikely that a research question will be found in the form of a hypothesis or null hypothesis in the way that you would expect in traditional quantitative research. Nevertheless, what the study is investigating should still be set out early and clearly. The research question should be set in context, with a summary of the background literature and the study's underpinning values and assumptions.

Section 2: study design

This section considers the robustness of the design of the research project.

2.1 How defensible/rigorous is the research design/methodology?

There are a large number of qualitative methodologies, and a tendency in healthcare studies to 'mix' aspects of different methodologies or to use a generic qualitative method. From a qualitative perspective, this should not compromise the quality of the study if the research design captures appropriate data and has an appropriate plan of analysis for the subject under investigation.

Sampling in qualitative research can be purposive. Qualitative research is not experimental and does not purport to be generalisable, and therefore does not require a large or random sample. People are usually 'chosen' for qualitative research based on being key informers. The choice of sample and sampling method should be described, ideally including any shortcomings of the sample.

Section 3: data collection

3.1 *How well was the data collection carried out?*

Assess whether the methods of data collection are described with details of the following:

- how the data were collected
- how the data were recorded and transcribed (if verbal data)
- how the data were stored
- what records were kept of the data collection.

Were these appropriate methods of data collection to use, given the aims of the research?

Section 4: validity

Assessing the validity of qualitative research is very different from assessing that of quantitative research. Qualitative research is much more focused on demonstrating the causes of bias rather than eliminating them. The report should include sections discussing the reflexive position of the researcher (their 'role' in the research), the context in which the research was conducted and the reliability of the actual data.

4.1 *Is the context clearly described?*

It is important when gauging the validity of qualitative data to consider whether the data are plausible and realistic. To make an accurate assessment of this, it is important to describe the context of the research in terms of the physical context (for example, youth club, GP surgery, gang headquarters) and who else was there (for example, participants are likely to position themselves very differently, and thus to respond very differently, in a discussion with parents present compared with a discussion with peers present). The participants should be described in enough detail to allow some insight into their life and situation and any potential context bias considered by the authors (that is, interpretation of the influence of the setting).

4.2 *Were the methods reliable?*

It is important that the method used to collect the data is appropriate for the research question and that the data generated map well to the aims of the study. Ideally, more than one method should have been used to collect data.

Section 5: analysis

Qualitative data analysis is very different from quantitative analysis. This does not mean that it should not be systematic and rigorous; however, systematisation and rigour require different methods of assessment.

5.1 Are the data 'rich'?

Qualitative researchers use the adjective 'rich' to describe data that are in-depth, convincing, compelling and detailed enough that they can provide some insight into the research participants' experience. It is also important to know the 'context' of the data – where they came from, what prompted them, what they pertain to, and so on.

5.2 Is the analysis reliable?

The analysis of data can be made more reliable by the researchers putting checks in place. Sections of data should be coded by another researcher or, as a minimum, a second researcher should check the coding for consistency. Participants may also verify the transcripts of their interview (or other data collection, if appropriate). Negative or discrepant results should be highlighted and discussed.

5.3 Are the findings convincing?

The results of the research should be convincing or credible. Findings should be presented clearly and organised logically and the authors should consider and explain any contradictions. Extracts from original data should be included where possible to give a fuller sense of the findings. These data should be appropriately referenced – although you would expect data to be anonymised, they still need to be referenced in relevant ways (for example, if sex differences were important, then you would expect extracts to be marked male or female).

5.4 Are the conclusions adequate?

This section is self explanatory.

Section 6: ethics

6.1 Was the study approved by an ethics committee?

All qualitative research involves ethical considerations, and these should be considered within any research report. Ideally there should be a full discussion of ethics, although this is rare because of space constraints in peer-reviewed journals. Any qualitative research should be approved by a

research ethics committee, and this should be stated in the report so that it is clear that every care was taken to protect research participants.

6.2 Is the role of the researcher clearly described?

The researcher should have considered their role in the research; for example, as a reader, interviewer or observer. This is often referred to as 'reflexivity'. The 'status' of the researcher can profoundly affect the data. For example, a middle-aged woman and an 18-year-old man are likely to get different responses to questions about sexual activity when interviewing a group of teenage boys. It is important to consider age, gender, ethnicity and 'insider' status (such as whether the interviewer or researcher is part of the group being researched or has the same condition or illness). The researcher can also profoundly influence the data by use of questions, opinions, judgements and so on, so it is important to know what the researcher's position is in this regard, and how the researcher introduced and talked about the research with the participants.

Further reading

Barbour RS (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal* 322: 1115–7

Daly J, Willis K, Small R et al. (2007) A hierarchy of evidence for assessing qualitative health research. *Journal of Clinical Epidemiology* 60: 43–9

Mays N, Pope C (2000) Assessing quality in qualitative research. *British Medical Journal* 320: 50–2

Miller G, Dingwall R, editors (1997) *Context and method in qualitative research*. London: Sage

Appendix I: Methodology checklist: prognostic studies

The criteria used in this checklist are adapted from: Hayden JA, Cote P, Bombardier C (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine* 144: 427-37.

Checklist

Study identification <i>Include author, title, reference, year of publication</i>				
Guideline topic:		Review question no:		
Checklist completed by:				
		<i>Circle or highlight one option for each question</i>		
<u>1.1</u>	The study sample represents the population of interest with regard to key characteristics, sufficient to limit potential bias to the results	Yes	No	Unclear
<u>1.2</u>	Loss to follow-up is unrelated to key characteristics (that is, the study data adequately represent the sample), sufficient to limit potential bias	Yes	No	Unclear
<u>1.3</u>	The prognostic factor of interest is adequately measured in study participants, sufficient to limit potential bias	Yes	No	Unclear
<u>1.4</u>	The outcome of interest is adequately measured in study participants, sufficient to limit potential bias	Yes	No	Unclear
<u>1.5</u>	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest	Yes	No	Unclear
<u>1.6</u>	The statistical analysis is appropriate for the design of the study, limiting potential for the presentation of invalid results	Yes	No	Unclear

Notes on use of Methodology checklist: prognostic studies

The studies covered by this checklist are designed to answer questions about prognosis. Such questions address the likelihood of an outcome for patients from a population at risk for that

outcome, based on the presence of a proposed prognostic factor. Prognostic factors may be disease-specific (for example, presence or absence of particular disease feature), demographic (for example, age, sex), or relate to the likely response to treatment or the presence of comorbidities.

This checklist is based on a checklist for the quality appraisal of studies about prognosis developed by Hayden and co-workers (2006).

Checklist items are worded so that a 'yes' response always indicates that the study has been designed and conducted in such a way as to minimise the risk of bias for that item. An 'unclear' response to a question may arise when the answer to an item is not reported or is not reported clearly.

1.1 The study sample represents the population of interest with regard to key characteristics, sufficient to limit potential bias to the results

Measures of prognosis can vary substantially when obtained from populations with different clinical or demographic features. Estimates of prognosis are not useful without information about the population from which they were obtained.

To minimise bias, the study population should be clearly defined and described and should represent the source population of interest. Points to consider include the following:

- Are the source population or the population of interest adequately described with respect to key characteristics?
- Are the sampling frame and recruitment adequately described, possibly including methods to identify the sample (number and type used; for example, referral patterns in healthcare), period of recruitment and place of recruitment (setting and geographical location)?
- Are inclusion and exclusion criteria adequately described (for example, including explicit diagnostic criteria or a description of participants at the start of the follow-up period)?
- Is participation in the study by eligible individuals adequate?
- Is the baseline study sample (that is, individuals entering the study) adequately described with respect to key characteristics?

1.2 Loss to follow-up is unrelated to key characteristics (that is, the study data adequately represent the sample), sufficient to limit potential bias

Attrition refers to the loss of participants during the course of a study. Consideration should be given to why participants dropped out, as well as how many dropped out. Attrition bias occurs when there are systematic differences between participants lost to the study and those who remain.

To minimise bias, completeness of follow-up should be described and adequate. Points to consider include the following:

- Is the response rate (that is, proportion of study sample completing the study and providing outcome data) adequate?
- Are attempts to collect information on participants who dropped out of the study described?
- Are reasons for loss to follow-up provided?
- Are the key characteristics of participants lost to follow-up adequately described?
- Are there any important differences in key characteristics and outcomes between participants who completed the study and those who did not?

If your review addresses more than one outcome, you should score this item for each outcome individually.

1.3 The prognostic factor of interest is adequately measured in study participants, sufficient to limit potential bias

The prognostic factor under study should be well defined. It should be clear how the investigators determined whether participants were exposed or not to the factor. The same definition and measurement should be used for all participants in the study. Often there may be more than one way of determining the presence or absence of the factor (for example, physical or laboratory tests, questionnaire, reporting of symptoms). The method of measurement should be valid (that is, it measures what it is claimed to measure) and reliable (that is, it measures something consistently).

To minimise bias, prognostic factors should have been defined and measured appropriately. Points to consider include the following:

- Is a clear definition or description of the prognostic factor(s) measured provided (including dose, level, duration of exposure, and clear specification of the method of measurement)?
- Are continuous variables reported, or appropriate cut-off points (that is, not data-dependent) used?
- Are the prognostic factors measured and the method of measurement valid and reliable enough to limit misclassification bias? (This may include relevant outside sources of information on measurement properties, as well as characteristics such as blind measurement and limited reliance on recall.)
- Are complete data for prognostic factors available for an adequate proportion of the study sample?
- Are the method and setting of measurement the same for all study participants?
- Are appropriate methods employed if imputation is used for missing data on prognostic factors?

1.4 The outcome of interest is adequately measured in study participants, sufficient to limit potential bias

The outcome under study should be well defined. It should be clear how the investigators determined whether participants experienced, or did not experience, the outcome. The same methods for defining and measuring outcome should be used for all participants in the study. Often there may be more than one way of measuring an outcome (for example, physical or laboratory tests, questionnaire, reporting of symptoms). The method of measurement used should be valid and reliable.

To minimise bias, the outcome(s) of interest should be defined and measured appropriately. Points to consider include the following:

- Is a clear definition of the outcome of interest provided, including duration of follow-up?
- Are the outcomes that were measured and the method of measurement valid and reliable enough to limit misclassification bias? (This may include relevant outside sources of information on measurement properties, as well as characteristics such as 'blind' measurement and limited reliance on recall.)
- Are the method and setting of measurement the same for all study participants?

If your review addresses more than one outcome, you should score this item for each outcome individually.

1.5 Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest

Confounding can occur when there are differences between participants, apart from the presence or absence of the prognostic factor, that are related to both the outcome and the prognostic factor. An example of this is if the participants are recruited at different stages of disease progression. The design and analysis of prognostic studies are usually based on some conceptual model about how factors interact to lead to the outcome.

This question is not relevant where the study is being reviewed for the purposes of identifying the absolute risk of the outcome in the group with the prognostic factor.

To minimise bias, important confounders should be defined and measured, and confounding should be accounted for in the design or analysis. Points to consider include the following:

- Are all important confounders, including treatments (key variables in the conceptual model), measured? Are clear definitions of the important confounders measured (including dose, level and duration of exposures) provided?
- Is measurement of all important confounders valid and reliable? (This may include relevant outside sources of information on measurement properties, as well as characteristics such as 'blind' measurement and limited reliance on recall.)
- Are the method and setting of measurement of confounders the same for all study participants?
- Are appropriate methods employed if imputation is used for missing data on confounders?
- Are important potential confounders accounted for in the study design (for example, matching for key variables, stratification or initial assembly of comparable groups)?
- Are important potential confounders accounted for in the analysis (that is, appropriate adjustment)?

If your review addresses more than one outcome, you should score this item for each outcome individually.

1.6 The statistical analysis is appropriate for the design of the study, limiting potential for the presentation of invalid results

Analysis undertaken within the study that is incorrect or inappropriate for the study design may result in false conclusions being drawn from the data.

To minimise bias, the statistical analysis undertaken should be clearly described and appropriate for the design of the study. Points to consider include the following:

- Is the presentation of data sufficient to assess the adequacy of the analysis?
- Where several prognostic factors are investigated, is the strategy for model building (that is, the inclusion of variables) appropriate and based on a conceptual framework or model?
- Is the selected model adequate for the design of the study?
- Is there any selective reporting of results?
- Are only pre-specified hypotheses investigated in the analyses?

In some circumstances it may be possible to reanalyse the data using the information supplied in the study report, in order to remove bias.