

Bilaga 2 Granskningsmallar

QUADAS modifierad

Study identification <i>Including author, title, reference, year of publication</i>					
Review question no:	Checklist completed by:				
1	Was the spectrum of participants representative of the patients who will receive the test in practice?	Yes	No	Unclear	N/A
2	Were selection criteria clearly described?	Yes	No	Unclear	N/A
3	Was the reference standard likely to classify the target condition correctly?	Yes	No	Unclear	N/A
4	Was the period between performance of the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?	Yes	No	Unclear	N/A
5	Did the whole sample or a random selection of the sample receive verification using the reference standard?	Yes	No	Unclear	N/A
6	Did participants receive the same reference standard regardless of the index test result?	Yes	No	Unclear	N/A
7	Was the reference standard independent of the index test? (that is, the index test did not form part of the reference standard)	Yes	No	Unclear	N/A
8	Was the execution of the index test described in sufficient detail to permit its replication?	Yes	No	Unclear	N/A
9	Was the execution of the reference standard described in sufficient detail to permit its replication?	Yes	No	Unclear	N/A
10	Were the index test results interpreted without knowledge of the results of the reference standard?	Yes	No	Unclear	N/A
11	Were the reference standard results interpreted without knowledge of the results of the index test?	Yes	No	Unclear	N/A
12	Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?	Yes	No	Unclear	N/A
13	Were uninterpretable, indeterminate or intermediate test results reported?	Yes	No	Unclear	N/A
14	Were withdrawals from the study explained?	Yes	No	Unclear	N/A
15	Were the number, training, and expertise of the persons executing and reading the index tests and the reference standard described?	Yes	No	Unclear	N/A
16	Were methods for calculating test reproducibility described?	Yes	No	Unclear	N/A
17	Were estimates of diagnostic accuracy and measures of statistical uncertainty (e.g., 95% confidence intervals) reported?	Yes	No	Unclear	N/A
18	Was the conflict of interest stated?	Yes	No	Unclear	N/A

1-14 ursprung QUADAS; 15-17 ursprung: STARD; 18 ursprung: AMSTAR.

Notes on use of Methodology checklist: studies of diagnostic test accuracy

This checklist is designed for the evaluation of studies assessing the accuracy of specific diagnostic tests. It does **not** address questions of the usefulness of the test in practice, or how the test compares with alternatives. Such questions should be assessed using the checklists for studies on interventions (see appendices D, E and F).

The questions in this checklist are aimed at establishing the validity of the study under review – that is, making sure that it has been carried out carefully, and that the conclusions represent an unbiased assessment of the accuracy and reliability of the test being evaluated. Each question covers an aspect of methodology that is thought to make a difference to the reliability of a study.

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 not reported clearly. ‘N/A’ should be used when a study of diagnostic test accuracy cannot give an answer of ‘yes’ no matter how well it has been done.

1. Was the spectrum of participants representative of the patients who will receive the test in practice?

What is meant by this item

Differences between populations in demographic and clinical features may produce measures of diagnostic accuracy that vary considerably; this is known as spectrum bias. Reported estimates of diagnostic test accuracy may have limited clinical applicability (generalisability) if the spectrum of participants tested is not representative of the patients on whom the test will be used in practice. The spectrum of participants takes into account not only the severity of the underlying target condition but also demographic features and the presence of differential diagnoses and/or comorbidities.

How to score this item

Studies should score ‘yes’ for this item if you believe, based on the information reported, that the spectrum of participants included in the study was representative of those in whom the test will be used in practice. This judgement should be based on both the method for recruitment and the characteristics of those recruited. Studies that recruited a group of healthy controls and a group known to have the target disorder will be coded as ‘no’ on this item in nearly all circumstances. Reviewers should pre-specify what spectrum of participants would be acceptable, taking into account factors such as disease prevalence and severity, age and sex. Clinical input may be required from the Guideline Development Group (GDG). If you think that the population studied does not fit into what you specified as acceptable, the study should be scored as ‘no’. If there is insufficient information available to make a judgement, this item should be scored as ‘unclear’.

2. Were selection criteria clearly described?

What is meant by this item

This refers to whether studies have reported criteria for entry into the study.

How to score this item

If you think that all relevant information regarding how participants were selected for inclusion in the study has been provided, then this item should be scored as 'yes'. If study selection criteria are not clearly reported, then this item should be scored as 'no'. In situations where selection criteria are partially reported and you feel that you do not have enough information to score this item as 'yes', then it should be scored as 'unclear'.

3. Was the reference standard likely to classify the target condition correctly?

What is meant by this item

The reference standard is the method used to determine the presence or absence of the target condition. Indicators of diagnostic test accuracy are calculated by comparing the results of the index test with the results of the reference standard. Estimates of test performance are based on the assumption that the index test is being compared with a reference standard that is 100% sensitive and specific. If there are any disagreements between the reference standard and the index test, it is assumed that the index test is incorrect. Thus the use of an inappropriate reference standard can bias estimation of the diagnostic accuracy of the index test.

How to score this item

Making a judgement about the accuracy of the reference standard may not be straightforward. You may need to consult a member of the GDG to determine whether a test is an appropriate reference standard. If a combination of tests is used, you may have to consider carefully whether these were appropriate.

If you believe that the reference standard is likely to classify the target condition correctly, then this item should be scored as 'yes'. If you do not think that the reference standard is likely to have classified the target condition correctly, then this item should be scored as 'no'. If there is insufficient information to make a judgement, then it should be scored as 'unclear'.

4. Was the period between performance of the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?

What is meant by this item

Ideally, the results of the index test and the reference standard are collected on the same participants at the same time. If this is not possible and there is a delay, misclassification may occur because of either spontaneous recovery or progression of the disease. This is known as disease progression bias. The length of the period that may cause such bias will vary between conditions. For example, a delay of a few days is unlikely to be a problem for chronic conditions. However, for infectious diseases a delay of only a few days between performance of the index test and the reference standard may be important. This type of

bias may also occur in chronic conditions in which the reference standard involves clinical follow-up of several years.

You will have to make judgements about what is considered 'short enough'. You should think about this **before** beginning your review, and define what you consider to be short enough for the specific topic area that you are reviewing. You may need clinical input to decide this.

How to score this item

When to score this item as 'yes' is related to the target condition. For conditions that progress rapidly, a delay of a even few days may be important. For such conditions this item should be scored as 'yes' if the delay between the performance of the index test and the reference standard is very short – a matter of hours or days. However, for chronic conditions, disease status is unlikely to change in a week, a month or even longer. For such conditions, longer delays between performance of the index test and reference standard may be scored as 'yes'. If you think that the period between the performance of the index test and the reference standard was sufficiently long that disease status may have changed between the performance of the two tests, then this item should be scored as 'no'. If insufficient information is provided, it should be scored as 'unclear'.

5. Did the whole sample or a random selection of the sample receive verification using the reference standard?

What is meant by this item

Partial verification bias (also known as work-up bias, [primary] selection bias or sequential ordering bias) occurs when not all of the study group receive confirmation of the diagnosis by a reference standard. If the results of the index test influence the decision to perform the reference standard, then biased estimates of test performance may arise. If participants are randomly selected to receive the reference standard, the overall diagnostic performance of the test is, in theory, unchanged. However, in most cases this selection is not random, possibly leading to biased estimates of the overall diagnostic accuracy. Partial verification bias generally only occurs in diagnostic cohort studies in which participants are tested using the index test before the reference standard.

How to score this item

If it is clear from the study that all participants (or a random selection) who received the index test went on to receive verification of their disease status using a reference standard, even if this reference standard was not the same for all participants, then this item should be scored as 'yes'. If some of the participants who received the index test did not receive verification of their true disease state (or the selection was not random), then this item should be scored as 'no'. If this information is not reported, this item should be scored as 'unclear'.

6. Did participants receive the same reference standard regardless of the index test result?

What is meant by this item

Differential verification bias occurs when some of the index test results are verified by a different reference standard. This is a particular problem if these reference standards differ in their definition of the target condition; for example, histopathology of the appendix and natural history for the detection of appendicitis. This usually occurs when participants who test positive on the index test undergo a more accurate, often invasive, reference standard test than those with negative results on the index test. The link (correlation) between a particular (negative) test result and being verified by a less accurate reference standard can lead to biased estimates of test accuracy. Differential verification bias generally only occurs in diagnostic cohort studies in which all participants are tested using the index test before the reference standard is performed.

How to score this item

If it is clear that participants received verification of their true disease status using the same reference standard, then this item should be scored as 'yes'. If some participants received verification using a different reference standard, then this item should be scored as 'no'. If this information is not reported, this item should be scored as 'unclear'.

7. Was the reference standard independent of the index test? (that is, the index test did not form part of the reference standard)

What is meant by this item

When the result of the index test is used in establishing the final diagnosis, incorporation bias may occur. This incorporation will probably increase the amount of agreement between index test results and the outcome of the reference standard, and hence result in overestimation of the various measures of diagnostic accuracy. For example, a study investigating magnetic resonance imaging (MRI) for the diagnosis of multiple sclerosis could have a reference standard composed of clinical follow-up, cerebrospinal fluid analysis and MRI. In this case, the index test forms part of the reference standard. It is important to note that knowledge of the results of the index test does not automatically mean that these results are incorporated in the reference standard. This item will only apply when a composite reference standard is used to verify disease status. In such cases it is essential that a full definition of how disease status is verified and which tests form part of the reference standard is provided.

How to score this item

For studies in which a single reference standard is used, this item will not be relevant and should be scored as 'N/A'. If it is clear that the index test did not form part of the reference standard, then this item should be scored as 'yes'. If it appears that the index test formed part of the reference standard, then this item should be scored as 'no'. If this information is not reported, this item should be scored as 'unclear'.

8 & 9. Was the execution of the index test described in sufficient detail to permit its replication? Was the execution of the reference standard described in sufficient detail to permit its replication?

What is meant by these items

A sufficiently detailed description of the execution of the index test and the reference standard is important for two reasons. Firstly, variation in measures of diagnostic accuracy can sometimes be traced back to differences in the execution of index tests and reference standards. Secondly, a clear and detailed description (or references) is needed to implement a certain test in another setting. If tests are executed in different ways then this would be expected to have an impact on test performance. The extent to which this would be expected to affect results depends on the type of test being investigated.

How to score these items

If the study reports sufficient details to permit replication of the index test and the reference standard, then these items should be scored as 'yes'. In other cases these items should be scored as 'no'. In situations where details of test performance are partially reported and you consider that you do not have enough information to score these items as 'yes', then they should be scored as 'unclear'.

10 & 11. Were the index test results interpreted without knowledge of the results of the reference standard? Were the reference standard results interpreted without knowledge of the results of the index test?

What is meant by these items

This issue is similar to the blinding of the people who assess outcomes in intervention studies. Interpretation of the results of the index test may be influenced by knowledge of the results of the reference standard, and vice versa. This is known as review bias, and may lead to inflated measures of diagnostic test accuracy. The extent to which this can affect test results will be related to the degree of subjectivity in the interpretation of the test result – the more subjective the interpretation, the more likely that the interpreter can be influenced by the results of the index test in interpreting the results of the reference standard, and vice versa. It is therefore important to consider the topic area that you are reviewing and to determine whether interpretation of the results of the index test or the reference standard could be influenced by knowledge of the results of the other test.

How to score these items

If the study clearly states that the test results (index test or reference standard) were interpreted blind to the results of the other test, then these items should be scored as 'yes'. If this does not appear to be the case, then they should be scored as 'no'. If this information is not reported, these items should be scored as 'unclear'. If in the topic area that you are reviewing the index test is always performed first, then interpretation of the results of the index test will usually be done without knowledge of the results of the reference standard. Similarly, if the reference standard is always performed first, then the results will be interpreted without knowledge of the results of the index test. In situations where one form of review bias does not apply, the item should be scored as 'N/A'. If interpretation of test results is entirely objective, then test interpretation is not susceptible to review bias and the item should be scored as 'N/A'. Another situation in which this form of bias may not apply is when test results are interpreted in an independent laboratory. In such situations it is unlikely that the person interpreting the test results will have knowledge of the results of the other test (either index test or reference standard).

12. Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?

What is meant by this item

The availability of information on clinical data during the interpretation of test results may affect estimates of test performance. In this context, clinical data are defined broadly to include any information relating to the participant that is obtained by direct observation, such as age, sex and symptoms. The knowledge of such factors can influence the diagnostic test result if the test involves an interpretative component. If clinical data will be available when the test is interpreted in practice, then these should also be available when the test is evaluated. However, if the index test is intended to replace other clinical tests, then clinical data should not be available. Thus, before assessing studies for this item it is important to determine what information will be available when test results are interpreted in practice. You should consult the GDG to identify this information.

How to score this item

If clinical data would normally be available when the test results are interpreted in practice and similar data were available when interpreting the index test results in the study, then this item should be scored as 'yes'. Similarly, if clinical data would not be available in practice and these data were not available when the index test results were interpreted, then this item should be scored as 'yes'. If this is not the case, then this item should be scored as 'no'. If this information is not reported, this item should be scored as 'unclear'. If interpretation of the index test is fully automated, this item may not be relevant and can be scored 'N/A'.

13. Were uninterpretable, indeterminate or intermediate test results reported?

What is meant by this item

A diagnostic test can produce an uninterpretable, indeterminate or intermediate result with varying frequency, depending on the test. These problems are often not reported in studies on diagnostic test accuracy, the uninterpretable results simply being removed from the analysis. This may lead to the biased assessment of the test characteristics. Whether bias will arise depends on the possible correlation between uninterpretable test results and the true disease status. If uninterpretable results occur randomly and are not related to the true disease status of the individual then, in theory, these should not have any effect on test performance. It is important that uninterpretable results are reported so that the impact on test performance can be considered; however, poor quality of reporting means that this is not always the case.

How to score this item

If it is clear that all test results, including uninterpretable, indeterminate or intermediate results, are reported, then this item should be scored as 'yes'. If the authors do not report any uninterpretable, indeterminate or intermediate results, and if the results are reported for all participants who were described as having been entered into the study, then this item should also be scored as 'yes'. If you think that such results occurred but have not been reported, then this item should be scored as 'no'. If it is not clear whether all study results have been reported, then this item should be scored as 'unclear'.

14. Were withdrawals from the study explained?

What is meant by this item

This occurs when participants withdraw from the study before the results of both the index test and the reference standard are known. If participants lost to follow-up differ systematically from those who remain, for whatever reason, then estimates of test performance may be biased. Poor quality of reporting of withdrawals may make the impact on estimates of test performance difficult to determine.

How to score this item

If it is clear what happened to all participants who entered the study, for example if a flow diagram of study participants is reported, then this item should be scored as 'yes'. If the authors do not report any withdrawals and if results are available for all participants who were reported to have been entered into the study, then this item should also be scored as 'yes'. If it appears that some of the participants who entered the study did not complete the study (that is, did not receive both the index test and the reference standard), and these participants were not accounted for, then this item should be scored as 'no'. If it is not clear whether all participants who entered the study were accounted for, then this item should be scored as 'unclear'.

15-17 *ursprung*: STARD; 18 *ursprung*: AMSTAR.

15. Variability in the manipulation, processing, or reading of the index test or reference standard will affect measures of diagnostic accuracy (41, 42). Many studies have shown reader variability, especially in the field of imaging (43, 44). The amount of the readers training can help readers to judge whether similar results are attainable in their own settings, with possibly less experienced readers. Professional background, expertise, and prior training to improve interpretation and to reduce inter-observer variation all affect the quality of reading (45, 46). Readers are more likely to interpret results from (subjective) tests as abnormal in settings with higher prevalences of the target condition, a tendency known as context bias (47). The example describes the reference standard in a study of a model that uses results of commonly performed laboratory tests to identify men who are heavy drinkers.

16. The index test and the reference standard are seldom perfect. Their reproducibility varies, and limited reproducibility adversely affects diagnostic accuracy (52). Observer variability can arise with imaging tests when the reader must summarize visual observations in a statement about the presence of disease. It also arises during classification, when the reader must use the data to place patients into diagnostic categories (41). Instrument variability concerns the amount of variation that arises during the operation of devices or systems, such as automated laboratory measurements. Other terms for this form of variation include imprecision, analytic methodological variation, or analytical noise (error). Poor reproducibility adversely affects diagnostic accuracy. If possible, authors should evaluate the reproducibility of the test methods used in their study and report their procedure to do so. For quantitative assays, it is useful to report imprecision as the coefficient of variation at two or more specified mean values near clinical decision points as obtained by repeating the test over a specified number of days. Withinrun coefficients of variation are appropriate if all patient samples were analyzed in a single run. In the example, the authors used the kappa statistic to express interobserver variability for conventional angiography and MRA in the detection of renovascular disease.

17. The final aim of a study of diagnostic accuracy is to produce an expression of how well the test results corresponded with the presence or absence of the target condition, as established by the reference standard. The values presented in the report should be taken as estimates. Due to chance variations in the patients submitted to the tests and other factors, the results are likely to differ over replications of the study in the same study population (51). The reporting of precision will show the reader the range of likely values around an estimate of diagnostic accuracy. Many journals require or strongly encourage the use of confidence intervals as measures of precision. A 95% confidence interval is conventional. Only 50% of the reports of diagnostic evaluations published in 1996 or 1997 in the *British Medical Journal* reported precision for the estimates of diagnostic accuracy (78).

18. Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

Referenser

QUADAS: National Institute for Health and Clinical Excellence. The guidelines manual, Appendix G: Methodology checklist: the QUADAS tool. 2009

STARD: Bossuyt et al. The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration, *Ann Intern Med.* 2003;138:W1-W12.

AMSTAR: Shea BJ et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews, *BMC Medical Research Methodology* 2007; 7:10.

Granskningsmall för systematiska översikter enligt AMSTAR

1. Redovisas en förbestämd studiemetod?

Forskningsfrågan och inklusionskriterierna ska vara fastställda innan översikten genomförs.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

2. Gjordes studieurval och dataextraktion av två oberoende granskare?

Minst två oberoende granskare ska ha utfört dataextraktionen, och ett konsensusförfarande bör vara definierat för att lösa oenigheter.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

3. Var litteratursökningen av tillfredsställande omfattning?

Sökningen bör göras i minst två elektroniska databaser. Översikten ska ange de årtal och databaser som ingår (t ex Central, Embase och Medline). Ämnesord (keywords) och/eller MESH-termer ska anges och i tillämpliga fall sökstrategin.

Alla sökningar bör kompletteras med genomgång av översiktsartiklar, läroböcker, aktuella innehållsförteckningar, ämnesspecifika databaser och register eller rådfrågning av experter, samt av referenslistorna i de framtagna studierna.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

4. Användes publikationsformen som ett inklusionskriterium?

Om inte alla rapporter som identifierades (vid den systematiska sökningen) inkluderats ska detta anges liksom skälen (publikationsform, språk osv). Författarna bör ange om alla rapporter oavsett publikationsform omfattas av sökningen, inklusive grå litteratur*.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

* SBU:s kommentar: Med begreppet grå litteratur menas information som producerats på alla nivåer inom statsapparaten, den akademiska sfären, företag eller industri i elektroniskt eller tryckt format som inte är kontrollerade av kommersiell publicering, alltså där publicering inte är det främsta syftet för producenten av texten.

5. Finns en förteckning över studierna i översikten (inkluderade och exkluderade)?

En förteckning över medtagna respektive uteslutna studier bör finnas i rapporten.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

6. Redovisas de ingående studiernas karakteristika och resultat i översikten?

Kända faktorer hos deltagarna i de utvärderade studierna (patient characteristics), såsom ålder, etnicitet, kön, relevanta socioekonomiska data, sjukdomstillstånd, varaktighet, svårighetsgrad och andra sjukdomar, bör anges i rapporten. Uppgifter om deltagarna, åtgärd/ behandling och utfall i studierna bör presenteras i sammanfattad form, t ex i en tabell.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

7. Har den vetenskapliga kvaliteten hos de ingående studierna utvärderats och dokumenterats?

Förbestämda metoder för kvalitetsvärderingen ska anges.

För effektstudier bör exempelvis framgå om författarna valt att bara ta med randomiserade, dubbelblinda studier med kontrollgrupper som får placebo eller studier med dold fördelning. För andra typer av studier gäller andra ställningstaganden.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

8. Har vederbörlig hänsyn tagits till studiernas vetenskapliga kvalitet vid formulering av slutsatserna?

Utvärderingen av metodologisk stringens och vetenskaplig kvalitet ska framgå i översiktens analys och dess slutsatser, och tydligt anges vid utformning av rekommendationer.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

9. Användes lämpliga metoder för sammanvägning av studiernas resultat?

Det bör säkerställas om det är lämpligt att lägga samman resultat från olika studier genom en bedömning av de ingående studiernas homogenitet (dvs Chi-2 test för beräkning av homogenitet, I^2). Om heterogenitet finns bör man använda en modell som tar hänsyn till slumpeffekter (random effects model) och/eller överväga om det ur klinisk synpunkt är lämpligt att slå ihop resultaten.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

10. Har man bedömt sannolikheten för publikations bias?**

En bedömning av publikations bias bör omfatta en kombination av grafiska hjälpmedel (t ex funnel plot eller andra tester) och/eller statistiska metoder (t ex Eggers regressionsanalys).

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

*SBU:s kommentar: Publikations bias leder till snedvriden publikation, såsom att positiva resultat publiceras oftare än negativa resultat.

11. Finns eventuella intressekonflikter angivna?

Eventuella sponsorer och bidragsgivare bör tillkännages både i den systematiska översikten och i de ingående studierna.

- Ja
- Nej
- Kan inte svara
- Ej tillämpligt

Referenser:

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D and Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews, *BMC Medical Research Methodology* 2007, 7:10.

Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS ONE* 2:e1350.

Granskningsmall för randomiserad kontrollerad prövning

Författare, år alternativt SBU:s identifikationsnummer:

Anvisningar:

- Alternativet ”oklart” används när uppgiften inte går att få fram från texten.
- Alternativet ”ej tillämpligt” väljs när frågan inte är relevant.
- Det finns förtydligande kommentarer till vissa delfrågor. Dessa anges med en fotnot.

Studiekvalitet	Ja	Nej	Oklart	Ej tillämpligt
1. Studiepopulation				
a) Framgår det hur många personer som exkluderades före randomiseringen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Är redovisningen av personer som inte randomiserades, trots att de var valbara, adekvat?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Tilldelning av åtgärd/intervention/behandling				
a) Användes en randomiseringsmetod som på ett acceptabelt sätt minimerar risken för manipulation?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Utfördes randomiseringen så att fördelningen blev oförutsägbar och slumpmässig? ¹	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Påbörjade samtliga deltagare, som randomiserades, behandlingen? ²	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Gruppernas jämförbarhet				
a) Var grupperna vid baseline rimligt lika avseende egenskaper som kan påverka resultatet (t ex ålder, kön, sjukdoms svårighetsgrad)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Blindning (maskering) ³				
Blindades följande på tillfredsställande sätt:				
a) Patienter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Prövare/behandlare	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Utvärderare av resultat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Bortfall (antalet randomiserade deltagare som inte har följts upp enligt studieprotokollet) ⁴				
a) Går det att följa deltagarnas väg genom studien t ex i ett flödesschema?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Studiekvalitet	Ja	Nej	Oklart	Ej tillämpligt
b) Är storleken på bortfallet efter randomisering acceptabelt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Är orsakerna till bortfallet acceptabla?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Följsamhet (compliance, adherence, concordance) ⁵				
a) Framgår det i vilken utsträckning deltagarna fullföljde behandlingen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Var andelen som fullföljde behandlingen acceptabel?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Rapportering av effektmått och biverkningar				
a) Var det primära effektmåttet definierat i förväg <u>och</u> adekvat rapporterat?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Var de sekundära effektmåtten definierade i förväg <u>och</u> adekvat rapporterade?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Baserades slutsatserna på enbart i förväg definierade effektmått och subgruppsanalyser? ⁶	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Har utfallen av samtliga viktiga effektmått redovisats på ett adekvat sätt? ⁷	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Rapporteras biverkningar/komplikationer på ett tillfredsställande sätt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Resultat och precision				
a) Redovisas resultaten på ett adekvat sätt? ⁸	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Har resultaten beräknats med lämplig analysmetod? ⁹	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Var den minsta kliniskt relevanta effekten definierad på förhand?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Är den valda minsta kliniskt relevanta effekten av rimlig storlek?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Har man använt acceptabla metoder för att mäta effekterna?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f) Mättes observatörsöverensstämmelsen på ett acceptabelt sätt? ¹⁰	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g) Är de överväganden och beräkningar som ligger till grund för antal deltagare acceptabla (poweranalys)? ¹¹	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Bindningar och jäv				
a) Anges eventuella bindningar och jäv (conflicts of interest)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Bedömer du att studiens resultat inte påverkats av intressekonflikter?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Studiekvalitet	Ja	Nej	Oklart	Ej tillämpligt
Total bedömning av studiekvalitet:				
Hög <input type="checkbox"/>		Medelhög <input type="checkbox"/>		Låg <input type="checkbox"/>

Kommentarer till granskningsmallen för RCT

1. Risk för att randomiseringen blir förutsägbar för prövare eller deltagare kan t ex finnas vid blockrandomisering. Blockrandomisering används bl a i multicenterstudier för att motverka slumpmässiga snedfördelningar mellan olika centra eller länder.
2. Här bedöms risken för att resultaten har påverkats av att deltagare selektivt exkluderats från studien efter randomiseringen, men innan behandlingen har påbörjats. Antalet deltagare som har lämnat studien bör sättas i relation till studiens storlek. Om antalet är jämnt fördelade mellan grupperna och skälen är angivna och godtagbara, så bedöms risken att resultaten kan ha påverkats som liten. Om mer än 5 procent av de randomiserade patienterna fallit bort eller om orsakerna inte är angivna eller godtagbara så bedöms risken som stor.
3. Det är önskvärt att både patienter och prövare är blindade i en studie. Ibland kan det av praktiska skäl vara svårt eller omöjligt att dölja för prövare och/eller patient vilken behandling som ges. Däremot är det oftast möjligt att den som utvärderar effekten är blindad.

De alternativ som finns är följande:

Öppen prövning: ingen är blindad

Enkelblind: a) deltagarna är blindade; b) prövare och/eller resultatutvärderare är blindade

Dubbelblind: a) deltagarna och prövare och/eller resultatutvärderare är blindade;

b) deltagarna och prövare och/eller resultatutvärderare är blindade samt det framgår av beskrivningen att all resultatbearbetning utfördes innan prövningskoden bröts.

Det finns flera exempel på studier där blindningen inte har lyckats pga karakteristiska effekter eller biverkningar av aktiv behandling, exempelvis muntorrhet vid behandling med neuroleptika och underlivsblödningar vid behandling med östrogen. I vissa fall är det möjligt att ge biverkningsmotverkande medel som tillägg till aktiv behandling för att minska risken för att blindningen äventyras. Andra faktorer som kan försvåra blindningen är bristande likhet mellan tabletter, inhalationspreparat etc avseende utseende eller smak. En stor ”placeboeffekt” i kontrollgruppen kan tala för en lyckad blindning. I vissa studier låter man studiedeltagarna gissa om de fått aktiv behandling eller kontroll.

4. Det bortfall som bedöms här avser bortfall efter randomisering. Det kan dock finnas tillfällen då även ett större bortfall sannolikt är slumpmässigt. Nedanstående exempel bör därför ses som riktvärden.

Litet (<10%)

Måttligt (10–19%)

Stort (20–29%)

Mycket stort ($\geq 30\%$). Undersökningen bedöms ofta sakna informationsvärde vilket kan innebära att studien bör exkluderas.

Bortfallet varierar mellan olika tidpunkter i en studie och kan variera mellan olika effektmått.

Bortfallet är ofta större ju längre tid som har gått. Därmed kan behandlingsresultaten från de sista besöken vara av tveksam validitet, medan resultaten från de första besöken kan vara giltiga.

5. Kontroll av följsamheten är särskilt viktig då man inte ser en signifikant effektskillnad mellan grupperna. En bristande följsamhet kan minska såväl interventionens effekter som bieffekter. Om interventionen har en signifikant effekt är kontroll av följsamheten ofta av mindre betydelse.

Undantag är om det var sämre följsamhet i gruppen som fick referensbehandling. Det senare är tänkbart i en placebokontrollerad studie om blindningen varit otillräcklig, alternativt om en referensbehandling har mycket högre frekvens av biverkningar.

Ett riktmärke för vad som anses vara acceptabel följsamhet är att mer än 80 procent av deltagarna ska ha mottagit mer än 80 procent av behandlingen.

6. Det är inte ovanligt att studier med negativa resultat inkluderar förklarande efteranalyser (explanatory- eller post hoc-analyser) för att t ex finna vissa subgrupper inom den studerade patientgruppen som kan ha nytta av behandlingen. Dessa analyser kan fylla en viktig hypotes-genererande funktion, men är naturligtvis behäftade med stor osäkerhet. Slutsatserna i en studie får därför aldrig baseras på sådana analyser.
7. Även om redovisade effektmått är rimliga, definierade i förväg och adekvat rapporterade kan det finnas andra viktiga effektmått som utelämnats. Oftast gäller det effektmått för att bedöma risker som också bedöms under fotnot 8.
8. De vanligaste mått som används för dikotoma variabler är riskkvot (risk ratio, RR), oddskvot (odds ratio, OR), absolut riskreduktion/riskskillnad (risk difference) och number needed to treat (NNT). För kontinuerliga variabler används vanligen absolut skillnad i medelvärde (difference in means, mean difference). Alla måtten ska redovisas med lämpligt spridningsmått, företrädesvis 95 procents konfidensintervall.
9. Resultaten kan analyseras enligt intention to treat (ITT) och/eller per protokoll (PP). En ITT-analys innebär att alla personer som randomiserats följs upp inom sin behandlingsarm oavsett om de fått tilldelad behandling eller inte och är oftast den metod som bör användas. Om resultaten är beräknade på annat sätt än med ITT finns det risk för att behandlingseffekten blir överskattad. ITT-analysen kan kompletteras med en känslighetsanalys enligt "worst case scenario" där sämsta tänkbara utfall tillskrivs saknade patienter i den grupp som uppvisar bäst effekt och bästa tänkbara utfall tillskrivs saknade patienter i den grupp som uppvisar sämst effekt. Ibland är det önskvärt att även en PP-analys redovisas, vilket innebär att bara de som följt hela studieprotokollet ingår i analysen. Vid bortfall i studier där man använder kontinuerliga variabler eller skalor använder man ibland en beräkningsmetod där det senast uppmätta resultatet antas gälla även för senare tidpunkter då data saknas (last observation carried forward, LOCF).
10. Vid utfallsregistrering i en behandlingsstudie kan observatörsvariationen vara en svaghet. Ett exempel är om flera observatörer ska utvärdera röntgenbilder eller cytologiska prov. Då ska observatörsöverensstämmelse mellan alla, eller ett större antal av observatörerna, vara rapporterat. Detta kan ske i form av Kappa-överensstämmelse eller "Intra-class correlation coefficient" (ICC), beroende på vilken skala som använts.
11. Powerberäkningar används för att beräkna studiers statistiska styrka, dvs för att räkna ut hur stor en studie bör vara för att med rimlig sannolikhet kunna påvisa en behandlingseffekt. Det är viktigt att författarna redovisar hur de kommit fram till vald studiestorlek och att beräkningen skett före studiestart. Annars kan det inte uteslutas att författarna successivt adderat deltagare till studien tills statistisk signifikans har uppnåtts.

Granskningsmall för kohortstudier med kontrollgrupper

Författare, år alternativt SBU:s identifikationsnummer:

1. Jämförbarhet	Ja	Nej	Oklart	Ej tillämpligt
1.1 De jämförda grupperna				
a) Är de jämförda grupperna adekvat valda? ¹	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Är kontrollgruppen relevant?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Är det sannolikt att interventions- och kontrollgruppen valdes ut och diagnostiserades på ett likartat sätt? ²	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.2 Jämförbarhet och förväxlingsfaktorer (confounders)				
a) Har författarna identifierat viktiga förväxlingsfaktorer? ³	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Var skillnaderna mellan grupperna i baslinjedata små (se förväxlingsfaktorer nedan)? ³	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Har författarna tagit hänsyn till eventuella skillnader i analyserna? ³	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Är risken för selektions- eller indikationsbias liten? ⁴	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.3 Interventionen				
a) Är interventionens innehåll tydligt definierad? ⁵	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Är jämförelseinterventionen tydligt definierad? ⁵	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Har författarna på ett adekvat sätt mätt kvaliteten på genomförandet av behandlingen? ⁵	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Följsamhet, bortfall	Ja	Nej	Oklart	Ej tillämpligt
2.1 Följsamhet (compliance, adherence)				
a) Framgår det i vilken utsträckning deltagarna som tillfrågades om studien och tackade ja fullföljde behandlingen?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Var andelen som fullföljde behandlingen acceptabel?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.2 Bortfall (antalet deltagare som inte har följts upp enligt studieprotokollet)				
a) Redovisas hur stort bortfallet är? ⁶	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Anges orsakerna till bortfallet? ⁶	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Är bortfallet acceptabelt? ⁶	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Blindning	Ja	Nej	Oklart	Ej tillämpligt
Var de som bedömde utfallen omedvetna om de undersökta tillhörde interventions- eller kontrollgruppen? ⁷	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Statistisk styrka	Ja	Nej	Oklart	Ej tillämpligt
a) Är de överväganden och beräkningar som ligger till grund för urvalsstorleken (sample size) tydligt beskrivna? ⁸	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Är den statistiska styrkan (power) tillfredsställande hög? ⁸	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Effektmått och statistisk analys	Ja	Nej	Oklart	Ej tillämpligt
a) Har den statistiska analysen av osäkerhet hanterats på ett adekvat sätt? ⁹	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Har författarna på ett adekvat sätt korrigerat för obalanser mellan grupperna med avseende på förväxlingsfaktorer/confounders? ¹⁰	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Har hänsyn tagits till avhoppare från behandlingen (treatment dropouts)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6. Biverkningar	Ja	Nej	Oklart	Ej tillämpligt
Mättes biverkningar/komplikationer på ett tillfredsställande sätt?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Bindningar och jäv	Ja	Nej	Oklart	Ej tillämpligt
a) Finns en förteckning över eventuella bindningar och jäv (conflict of interests)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Bedömer du att studiens resultat inte påverkats av intressekonflikter?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Total bedömning av studiekvalitet:
Hög <input type="checkbox"/> Medelhög <input type="checkbox"/> Låg <input type="checkbox"/>

Kommentarer till granskningsmallen för kohortstudier med kontrollgrupper

Kohortstudier med kontrollgrupp innebär att man följer minst två grupper longitudinellt, dvs framåt i tiden för att studera hur det går för dem. Det kan gälla såväl icke-randomiserade kontrollerade studier som andra observationsstudier där man antingen studerar behandlingsinsatser eller riskfaktorer.

Synonyma begrepp är:

Interventionsgrupp = exponerad grupp = riskfaktor = fall

Kontrollgrupp = oexponerad grupp = jämförelsegrupp = referensgrupp

1. Finns det en klar definition av jämförelsegruppen? Jämförs interventionen mot en annan intervention eller mot ingen intervention alls? Har jämförelsegruppen hämtats från den allmänna befolkningen eller från ett begränsat urval? Om jämförelsegruppen är historisk kontroll finns det anledning att vara särskilt försiktig vid värderingen.
2. En viktig fråga är om samma metodik användes för att gruppera till interventions- respektive kontrollgrupp?
3. Förväxlingsfaktorer (confounders) är bakgrundsvariabler som påverkar utfallet. De kan vara ojämnt fördelade mellan grupperna och stör då det ”sanna” resultatet. Några viktiga confounders är ålder, kön, bakomliggande sjukdomshistoria, samsjuklighet, riskfaktorer och inte minst socioekonomisk status. Socioekonomi är sannolikt den starkaste riskfaktorn till sjuklighet och för tidig död.

Uppgifter som kan visa väsentliga skillnader mellan grupperna finns ofta i en inledande tabell eller under bakgrundsdata (baseline characteristics).

4. Selektionsbias uppkommer när det finns en eller flera inbyggda skillnader mellan grupperna som i sig kan förklara resultaten. Risken är särskilt stor när det gäller preventiva eller symtomlindrande åtgärder som välinformerade patientgrupper kan efterfråga. Risken för selektionsbias är också hög om åtgärden är särskilt lämplig att sätta in på högrisk- eller lågriskpatienter.
5. Interventionernas innehåll kan definieras avseende t ex tydligt beskrivet upplägg, antal träffar, tidsomfattning, manualiserat vs icke-manualiserat upplägg. Kvaliteten på interventionerna kan vara mätt avseende behandlingsintegritet, via t ex videospelningar, reliabilitetsmått, checklistor av genomförandet.
6. Stora bortfall ökar generellt sett risken för att resultaten kan vara påverkade av systematiska fel. Det kan dock finnas tillfällen då även ett större bortfall sannolikt är slumpmässigt. Som ett riktvärde för läkemedelsstudier är risken liten om bortfallet är mindre än 10 procent, medelstor om bortfallet ligger mellan 10 och 19 procent och stor om bortfallet är mellan 20 och 29 procent. Om bortfallet i läkemedelsstudier är 30 procent eller mer är informationsvärdet tveksamt och studien kan eventuellt sorteras bort. Bortfallet kan variera mellan olika tidpunkter och olika effektmått. Vid långtidsuppföljning kan man få acceptera något högre bortfall.
7. Om de som registrerar utfallen känner till vilka som fått en viss behandling kan det öka risken för en systematisk felregistrering.

8. Små studier där proverna inte har beräknat hur många deltagare som krävs för att få ett statistiskt signifikant resultat på det primära effektmåttet har ofta stora kvalitetsbrister. Det är viktigt att bedöma studiens statistiska styrka (power) för varje separat utfallsmått. Ett exempel är biverkningsrapportering. Studierna är oftast planerade för att belysa de positiva effekterna och har kanske inte tagit hänsyn till hur många deltagare som behövs för att få statistiskt säkerställda negativa effekter.
9. Bedöm om konfidensintervall eller andra relevanta mått redovisas på ett adekvat sätt eller om det finns en motivering för att sådana uppgifter saknas. Det kan t ex gälla vid totalundersökningar av stora datamaterial.
10. Metoder som kan användas i detta sammanhang är matchning/restriktion, stratifierad analys, multivariat modellanalys (t ex regressionsanalys) eller propensity score-metodik.