DRAFT: 2009 06 12

SUPPORT Tools for evidence-informed health Policymaking (STP)

7. Assessing the reliability of systematic reviews

Simon Lewin[3]
Andrew D Oxman[1]
John N Lavis[2]
Atle Fretheim[4]

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway. Email: oxman@online.no
2. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5. Email: lavisj@mcmaster.ca
3. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway and Health Systems Research Unit, Medical Research Council of South Africa. Email: simon.lewin@nokc.no
4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway. Email: atle.fretheim@nokc.no

Corresponding author:
Dr Simon Lewin
Norwegian Knowledge Centre for the Health Services
P.O. Box 7004, St. Olavs plass
N-0130 Oslo, Norway

Email: simon.lewin@nokc.no

Abstract

Background: This article is number 7 in a series of 21 articles on tools for evidence-informed health policy making. The reliability of systematic reviews of health interventions is variable. Consequently, policy makers and others need to assess how much confidence they can place in this evidence. Using a systematic and transparent process to determine reliability can help to prevent introducing errors and biases into these judgements.

Objectives: In this article we suggest five questions to consider when assessing the reliability of a systematic review.

Key messages:

- The following questions should be considered when assessing the reliability of a systematic review of effects:
  1. Did the review explicitly address an appropriate policy or management question?
  2. Were appropriate criteria used to select the studies?
  3. Was the search for relevant studies detailed and reasonably comprehensive?
  4. Were assessments of the studies' relevance to the review topic and of their risk of bias studies reproducible?
  5. Were the results similar from study to study?
- Tools to assess the reliability of systematic reviews can only assess what is reported. While a number of available tools allow the reliability of reviews to be scored, these approaches should be avoided
- When making decisions informed by the evidence presented in a review, policy makers need to consider assessments of the reliability of a review alongside other information, such as the usefulness of the review in relation to the policy question and the local context

Background

This article is number 7 in a series of 21 articles on tools for evidence-informed health policy making [1]. It is also the first of six articles in this series on characterising the costs and consequences of potential policy and programme options. Its purpose is to suggest questions to guide those who wish to critically appraise the reliability of systematic reviews.

Systematic reviews of randomised controlled trials (RCTs) are widely accepted as providing the most reliable evidence about the effects of healthcare interventions [1, 2]. Increasingly, systematic reviews are also being used to identify, appraise and combine evidence on the economic consequences of interventions [3]. They are also used to summarise evidence from qualitative studies, for example of consumer or provider views of health interventions [4-7]. In this article we focus on systematic reviews of the effects of healthcare policies or programmes. However, we also provide some guidance for assessing the quality of reviews of qualitative studies and of reviews of economic studies (see Box 1).

Systematic reviews are characterised by their systematic and explicit approach to accessing, appraising and synthesising evidence. This approach is intended to reduce the risk of bias and errors that occur by chance and also to help facilitate critical appraisal of these syntheses [8, 9]. However, the rigour with which systematic reviews are conducted varies and reviews are therefore not all equally reliable. There are a number of reasons why the reliability of a systematic review may be limited, including a failure to:

- Specify the question and methods for the review in advance of undertaking the review, for example in a published review protocol
- Specify clear criteria for including and excluding studies
- Adequately describe the studies included in the review
- Assess the risk of bias for studies included in the review
- Assess the risk of publication bias, i.e. the possibility that some studies, typically those with positive ('statistically significant') results are more likely to be published, and therefore included in a review, than others
- Use appropriate methods for combining the results of the included studies ('meta-analysis'), when this is relevant
- Adequately examine differences in the findings of studies included in a review (i.e. the 'heterogeneity' of the findings)
- Base the conclusions of the review on the included data

Other potential limitations of systematic reviews include conflicts of interest, which can affect the reliability of a review in any of the ways listed above, and reviews being out of date.

For example, a study comparing the methodology and reporting components of Cochrane reviews with those published in paper-based journals found that the former included components that made them less prone to bias. Specifically, clear descriptions of reviews' inclusion and exclusion criteria and a formal assessment of the risk of bias of the studies included in each review decreased the overall risk of bias in the Cochrane reviews [10]. Another study compared the methodological quality and conclusions in Cochrane reviews of drug trials with those in industry-supported reviews of the same drugs. This study found that Cochrane reviews scored higher on a quality assessment. The potential for bias in a review was considered more frequently in Cochrane reviews than in those which were industry supported. Furthermore, industry-supported reviews were significantly more likely to recommend the drug in question without reservations [11]. A number of other studies of reviews have also reported differences in their quality and conclusions [12-15].

In using systematic reviews of effects to inform policy decisions, policy makers and others therefore need to judge how much confidence they can place in this evidence. Using a systematic and transparent process can help to prevent introducing errors and biases into these judgements. A systematic and transparent process also allows other stakeholders, including the public, to understand and appraise these judgements. This is particularly important where such assessments influence the recommendations or decisions regarding clinical interventions or services [16] or decisions to implement or stop programmes or policies.

While a number of tools have been developed for assessing the quality of systematic reviews [17, 18], the criteria included in these tools are similar. For example, the 11 items included in the AMSTAR tool (A MeaSurement Tool to Assess Reviews) cover key aspects of the systematic review process, including those listed above, that can potentially limit the reliability of a review (the AMSTAR tool is described in Table 1). A number of other tools are available including those developed by CASP (Critical Appraisal Skills Programme) [19] and by Oxman and Guyatt [20]. Several tools allow the reliability of reviews to be scored on a rating scale. This latter approach should be avoided, however, as an overall score does not necessarily indicate to the reader which particular aspects of the review were conducted reliably. The process also involves assigning so-called weightings to different items in the assessment tool, which is difficult to justify as it is not clear which items should be weighted more heavily [21]. It should also be noted that all such tools can only assess the reliability of what is reported. When key information about the methods used in a review is not reported, it may be unclear what was done – or the extent to which what was done constitutes an important limitation.

An *assessment* of the reliability of a review needs to be differentiated from an *understanding* of the results of the review itself. Box 2 provides guidance on what to look for in the results of a review. An assessment of reliability also needs to be differentiated from any assessment that might be done of the *relevance* of the review to the policy question at hand. Considerations of relevance relate to whether the review provides evidence of the effects of the different policy options under consideration, and whether the findings of the review are applicable to the setting in which the policy will be implemented. The process of assessing the applicability of the findings from systematic reviews is discussed further in Article 8 in this series [22].

In this article we suggest five questions to consider when assessing the reliability of systematic reviews of the effects of policies or programmes. The term ‘reliability’ rather than ‘quality’ is used in this article: the latter is typically used to refer to both the quality of a review and the quality of the evidence included in it while the former refers to the quality of the review only.

Questions to consider

The following questions can guide policy makers in assessing the reliability of a systematic review of effects:

1. Did the review explicitly address an appropriate policy or management question?
2. Were appropriate criteria used to select the studies?
3. Was the search for relevant studies detailed and reasonably comprehensive?
4. Were assessments of the studies’ relevance and of their risk of bias reproducible?
5. Were the results similar from study to study?

1. Did the review explicitly address an appropriate policy or management question?

An important first step in assessing the reliability of a systematic review is to examine the question addressed. While the technical design and conduct of a review may be excellent, the findings of the review are unlikely to be useful in decision making if they do not explicitly address an appropriate or sensible policy or management question.

An appropriate policy or management question will:

- Be explicit – in other words stated in detail rather than implied in the material presented. If the review question is not expressed explicitly and formulated clearly, it is difficult to assess adequately the conduct of the review given that this will need to be considered, at least in part, in relation to the question itself [23]. For example, an appraisal of whether the criteria used to select studies for a review were appropriate needs to be done in relation to the review question that the studies are intended to answer. A clear question also helps readers to assess whether the review is relevant to their work [23]
- Be established *a priori* – in other words *before* the review was conducted. It is important that the review question be specified before a review is conducted, for example in a published review protocol. If this is not done, there is a risk that the question may be altered to suit the evidence found, thereby undermining confidence in the findings
- Address a question of relevance to policy making or management. This will need to be assessed in a specific context, based on the range of issues that are important in a particular jurisdiction at a particular time. A review question may not be relevant if:
    - It is too narrow. For example, a review may consider the effects of a programme on a particular age group of participants only, in a particular setting or for a restricted range of outcomes. The results in this instance would therefore not be generalisable to other populations or settings
    - It is too broad. For example, a review defines a programme to include a very broad range of practices, not all of which may be relevant in a particular jurisdiction. Or it asks a very broad question that is not useful from a decision making perspective, such as whether nurses can effectively deliver health promotion programmes. This question may not be useful in deciding whether a particular cadre of nurses, such as enrolled nurses, can effectively deliver a health promotion programme for a specific health issue, such as HIV/AIDS prevention
    - It does not specify an appropriate comparison group. For example, the programme is compared to a 'no programme' scenario rather than to the current best treatment available for the condition

A well formulated review question should specify *all* of the following: the types of population and settings that the review will cover (e.g. children aged between one month and six years living in a malaria-endemic area); the types of programmes and comparisons considered (e.g. anti-malarial drugs given at regular intervals (the intervention) compared to placebo or no drug (the comparison)); and the types of outcomes that are of interest (e.g. clinical malaria and severe anaemia) [24, 25]. While the need for a well formulated review question seems obvious, many narrative reviews fail to provide this. A review of a sample of such reviews published in major medical journals showed that 20% failed to state their purpose clearly [26].
.

2. Were appropriate criteria used to select the studies?

The inclusion and exclusion criteria for a review refer to the detailed listing of the types of population, interventions, comparisons and outcomes that the review will consider. These criteria, specified in the review protocol, will determine which studies are included in the review. They will therefore influence strongly the findings of the review. It is important that these criteria are appropriate in relation to the review question.

The following questions should be examined when considering whether the criteria used to select studies are appropriate:

- Does the review specify clear inclusion and exclusion criteria? These are important to protect against bias related to the inclusion of studies in the review. A recent assessment of the methodological quality of systematic reviews in general surgery, for example, found that only 70% of these reported the criteria used for deciding which studies to include in the review [14]
- Are the inclusion and exclusion criteria explicit in relation to the following: the types of population considered, the types of interventions and comparisons considered, and the types of outcomes considered?
- Are the inclusion and exclusion criteria congruent with the review question? [27] For example, if the review aims to evaluate prophylaxis and intermittent treatment with anti-malarial drugs to prevent malaria in young children living in malaria-endemic areas, do the criteria indicate the inclusion of studies of children from the appropriate settings and specify the forms of prophylaxis and treatment that will be considered? [25]

3. Was the search for relevant studies detailed and reasonably comprehensive?

A key aspect of a systematic review is a thorough and reproducible search of the literature for studies that meet the eligibility criteria of the review. This approach is one of the elements that differentiates systematic reviews from narrative reviews. Systematic searching contributes to minimising bias in a review by ensuring that all relevant evidence is considered, and therefore helps to achieve reliable estimates of the effects of the policy or programme being examined [28].

Publication bias – that is, the selective publication of studies based on the direction and strength of their results [29] – is one route through which bias may be introduced into reviews. A recent review examined the extent to which the publication of randomised trials is influenced by whether positive results were found, and the perceived importance of the trial findings. It showed that trials with positive results were significantly more likely to be published than trials presenting negative findings [30]. Both this review and other work has also shown that trials reporting positive findings are published sooner than other trials [31]. A consequence of publication bias is that reviews may overestimate the positive effects of programmes if attempts are not made to identify both published and unpublished studies.

The extent to which systematic reviews include comprehensive searching varies. For example, a review of the reporting of published reviews on the treatment of asthma found that only 52% of the 33 reviews examined included a reasonably comprehensive search for evidence of effects [12]. It is therefore important to check how searches for relevant studies were conducted.

The following should be examined when considering whether the search for relevant studies was detailed and reasonably comprehensive [32]:

- Does the review describe in detail the strategy used to search for relevant studies? This reporting should include the list of sources searched, the key words used to search these sources (where applicable), and the years over which the sources were searched. Table 2 provides examples of the range of sources searched in reviews published in the Cochrane Library
- Did the search strategy include electronic databases of published studies? A wide range of electronic databases of published studies is available, and several can be searched at no or very low cost. Key databases include Pubmed/MEDLINE (compiled by the National Library of Medicine, USA), the Cochrane Central Register of Controlled Trials (CENTRAL – compiled by the Cochrane Collaboration), and regional databases such as LILACS (Latin American and Caribbean Health Sciences). Articles 3 [33] and 4 [34] in this series provide further information on finding relevant research literature
- Were the searches of electronic databases supplemented by additional searching? This might have included examining the reference lists for relevant studies, contacting authors and experts in the field, and consulting specialised registers of studies related to the topic area of the review
- Are the searches up-to-date? Does the review specify the period covered by the searches and are the searches current? A published review, while relevant to a policy question, may have used searches that are now several years old. It is therefore possible that the review does not include all the relevant evidence and may give an unreliable estimate of the effects of the programme

4. Were assessments of the studies' relevance to the review topic and of their risk of bias reproducible?

Authors of systematic reviews need to make two important judgements regarding each primary study that might be included in a review. Firstly, does the study meet the criteria for inclusion in their review – in other words, is it relevant to the review topic? Secondly, what is the risk of bias in the results of the study? Risk of bias refers to the risk of "a systematic error, or deviation from the truth, in results or inferences" [21]. It also relates to the question of whether the results of a study can be assumed to be accurate [21].

As noted above, reviews need to specify clear inclusion and exclusion criteria to protect against bias in the process of including studies. Review authors need to make judgements in assessing potentially eligible studies against these criteria. These judgements will affect the findings of the review by influencing the studies included in it. The chance of bias or errors in these judgements can be minimised as follows. Firstly, decisions on which studies to include in a review should be made independently by two reviewers. Including an additional reviewer, or holding additional discussions are two ways that can be used to resolve any disagreements related to the inclusion of a particular study. Secondly, the reasons for including a study, and for excluding a study that appears relevant, should be recorded in the published review. This allows the reader to make their own judgement regarding eligibility decisions. It also provides a transparent 'audit trail' for the review, ensuring that the process is reproducible.

The ability of a systematic review to reach conclusions regarding the effects of a policy or programme also depends on the validity of the data obtained from each included study. Pooling the results of the studies, or creating a summary of them in a review, may give a

misleading result if the validity of the individual studies included in the review is poor. Evaluating the risk of bias in the results of the included studies is therefore an important element of a systematic review. Such assessments should feed into the interpretation and conclusions of the review [21].

A number of different approaches have been developed for assessing quality or risk of bias for randomised trials [21, 35, 36]. While we will not discuss these different approaches here, it is important to note that a review should be explicit regarding the approach used and should apply this approach consistently.

In assessing the relevance of the included studies to the review topic and their risk of bias in the context of a systematic review the following should be considered:

- Was an explicit and transparent approach used to assess the relevance of studies to the review topic? As noted above, a review should state how relevance was assessed and provide a list of both included and excluded studies
- Was an explicit and transparent approach used to assess the risk of bias in the included studies? A review should report the tool used to assess the risk of bias, how the assessment was conducted, and the results of the assessment
- Were the results of the risk of bias assessment taken into account in interpreting the results of the review? For example, when the risk of bias in the included studies is high, we might have less confidence in the findings of the review

5. Were the results similar from study to study?

The findings of the studies included in a review may be very similar or they may vary, in terms of the effects of the programme on a particular outcome. This variability among the studies included in a review is usually referred to as 'heterogeneity' [21]. The variability among studies included in a review depends in part on the scope of the review. Where the scope is wide, it might be expected that the range of included studies, and therefore of variability, will be wide. In contrast, where the scope of a review is narrow, the included studies are likely to be more similar to one another.

If the participants, interventions or outcomes of studies included in a review are very different, this may lead to variation or heterogeneity if the intervention effect is affected by these factors. The true intervention effect will be different across these studies and therefore the average effect across studies will not be helpful.

Depending on the level of variability, reviews may use different approaches to summarising the information from the studies included. Such approaches include:

- Calculating the average effect across studies: This approach is useful when the variability across studies is low. For example, a systematic review of 'early discharge combined with hospital at home' programmes (i.e. programmes that provides active treatment by health providers in a patient's home for a health issue that otherwise would require acute hospital inpatient care) found that the studies included were sufficiently similar to be able to estimate the average effect of the programme. The review found insufficient evidence of economic or health benefits from 'early discharge hospital at home' programmes [37]
- Calculating the average effect for subgroups of studies included in a review: This may be useful when the overall variability of studies included in a review is high, and it is therefore unhelpful to calculate an average affect, but where variability is low among

- Describing the range of effects sizes: Where studies are not sufficiently similar to make calculating an average effect useful, it may still be possible to describe the range of effects found in the studies. For example, a review of the effects of audit and feedback on the practice of healthcare providers demonstrated that compliance with desired practice ranged from a decrease of 16% to an increase of 70%, with a median of 5%. The review indicated that audit and feedback can make practice more effective but that the effects are generally small to moderate [39]
- Cataloguing the types of interventions to address a particular issue: The wide scope of some reviews, and therefore the variability of the studies within them, means that it is not sensible to attempt to combine quantitatively the findings of the included studies or even to describe the range of effect sizes. In these cases, a narrative review can be undertaken. For example, a systematic review of the effectiveness of health service interventions aimed at reducing inequalities in health included studies that assessed programmes designed to reduce inequalities in health, and that could be implemented by the health services alone, or in collaboration with other agencies. The range of included studies was large, from programmes to improve control of blood pressure to health promotion interventions, and so no statistical pooling was attempted [40]

Where results differ from study to study, the following questions should be considered:

- Is there a compelling explanation for the differences that were found? This might include differences in the participants, interventions, comparison groups or outcomes across the included studies
- If a pooled estimate was made, is this likely to be meaningful? If the studies included in a review are very varied, a pooled estimate may not be meaningful. Further exploration of the data, through subgroup analysis, may be conducted but the results of such exploratory analyses need to be interpreted with caution

As the number of available systematic reviews increases, it is becoming more common to find more than one systematic review for a particular policy question. Sometimes the results or conclusions of these reviews may be different. Box 3 provides guidance on how policy makers might approach the problem of locating two or more reviews with conflicting findings.

Resources

Useful documents and further reading

- Higgins JPT, Altman DF: **Chapter 8: Assessing risk of bias in included studies.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008. Available at: http://www.cochrane-handbook.org/
- Counsell C: **Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews.** *Ann Intern Med* 1997, **127:** 380-387
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C *et al.*: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.** *BMC Med Res Methodol* 2007, **7:** 10. Available at: http://www.biomedcentral.com/1471-2288/7/10

Links to websites

- The *Rx for Change* database summarises current research evidence about the effects of strategies to improve drug prescribing practice and drug use. This database includes summaries, including reliability assessments, of systematic reviews that evaluate the effects of strategies targeting professionals, the organisation of healthcare, and consumers. Available at: http://www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions
- Cochrane Effective Practice and Organisation of Care (EPOC) Review Group. The Review Group provides guidance on assessing the reliability of different types of studies of effectiveness. Available at: http://www.epoc.cochrane.org/en/index.html
- The SUPPORT (SUPporting POlicy relevant Reviews and Trials) Collaboration produces summaries of high priority reviews for low- and middle-income countries. These include assessments of reliability. Available at: http://www.support-collaboration.org/index.htm

Table 1: AMSTAR – A MeaSurement Tool to Assess Reviews, 2007 (from [32])

1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review	□ Yes □ No □ Can't answer □ Not applicable
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors, and a consensus procedure for disagreements should be in place	□ Yes □ No □ Can't answer □ Not applicable
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include the years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and, where feasible, the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialised registers, or experts in the particular field of study, and by reviewing the references in the studies found	□ Yes □ No □ Can't answer □ Not applicable
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language, etc.	□ Yes □ No □ Can't answer □ Not applicable
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided	□ Yes □ No □ Can't answer □ Not applicable
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided about the participants, interventions and outcomes. The ranges of characteristics in all the studies analysed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported	□ Yes □ No □ Can't answer □ Not applicable
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g. for effectiveness studies if the author(s) chose to include only randomised, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria). For other types of studies, alternative items will be relevant	□ Yes □ No □ Can't answer □ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The methodological rigour and scientific quality of the studies should be considered in the analysis and the conclusions of the review, and explicitly stated when formulating recommendations	□ Yes □ No □ Can't answer □ Not applicable
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable and to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should also be taken into consideration (i.e. was it appropriate to combine the results?)	□ Yes □ No □ Can't answer □ Not applicable
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g. a funnel plot, other available tests) and/or statistical tests (e.g. Egger regression test)	□ Yes □ No □ Can't answer □ Not applicable
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies	□ Yes □ No □ Can't answer □ Not applicable

Table 2: Examples of sources searched in systematic reviews

Review	Sources searched
Health systems review Example: Systematic review of lay health worker interventions in primary and community healthcare [38]	1. Electronic databases of published studies: • MEDLINE • Cochrane Central Register of Controlled Trials (CENTRAL) and specialised Cochrane Registers (EPOC and Consumers and Communication Review Groups) • Science Citations • Embase • CINAHL • Healthstar • AMED • Leeds Health Education Effectiveness Database 2. Bibliographies of studies assessed for inclusion 3. All contacted authors were asked for details of additional studies
Public health review Example: Systematic review of male circumcision for prevention of heterosexual acquisition of HIV in men [41]	1. Electronic databases of published studies: • MEDLINE • EMBASE • Cochrane Central Register of Controlled Trials (CENTRAL) 2. Electronic databases of conference abstracts: • AIDSearch Conference databases 3. Electronic databases of ongoing trials: • ClinicalTrials.gov • Current Controlled Trials 4. Contacted researchers and relevant organisations in the field 5. Checked the reference lists of all studies identified by the above methods and examined any systematic reviews, meta-analyses, or prevention guidelines identified during the search process

Review	Sources searched
Clinical review Example: Systematic review of statins for the prevention of dementia [42]	1. Electronic databases: • The Specialized Register of the Cochrane Dementia and Cognitive Improvement Group • Cochrane Central Register of Controlled Trials (CENTRAL) • MEDLINE • EMBASE • PsycINFO • CINAHL • SIGLE (Grey Literature in Europe) • LILACS (Latin American and Caribbean Health Science Literature) 2. Electronic databases of conference abstracts: • ISTP (Index to Scientific and Technical Proceedings) • INSIDE (British Library Database of Conference Proceedings and Journals) 3. Electronic databases of theses: • Index to Theses (formerly ASLIB) (United Kingom and Ireland theses) • Australian Digital Theses Program • Canadian Theses and Dissertations • DATAD – Database of African Theses and Dissertations • Dissertation Abstract Online (USA) 4. Electronic databases of ongoing trials: searched a large range of such databases

Box 1: Assessing the quality of systematic reviews of qualitative studies and systematic reviews of economic studies

An increasing number of systematic reviews of qualitative studies are being undertaken. As well as providing important information in their own right, these reviews can also inform and supplement systematic reviews of effects [43, 44]. However, it is important for the reader to assess the reliability of these reviews. To date, few tools have been designed for this specific purpose. But many of the questions used to guide policy makers in assessing the reliability of systematic reviews of effects are also useful for reviews of qualitative studies:

1. *Did the review address an appropriate policy or management question?* The review question should be amenable to being addressed using qualitative data and should be relevant to policy making. Typical questions focus on the views and experiences stakeholders regarding health and healthcare
2. *Were the criteria used to select studies appropriate?* The description of how studies were selected should be appropriate in relation to the research question
3. *Was a clear and appropriate explanation provided for the search approach used?* Some reviews of qualitative studies undertake comprehensive literature searches while others may use sampling approaches. A clear description of, and justification for, the approach used should be provided
4. *Was the approach used to assess the reliability of the included studies appropriate?* The review should describe how the reliability of the included studies was taken into account
5. *Was a appropriate approach used to analyse the findings of the included studies?* The review should use an accepted approach to synthesis and should describe the rationale for the approach chosen

Questions to consider when assessing the reliability of reviews of economic studies include (from [45]):

1. *Is it unlikely that important relevant studies were missed?*
2. *Were the inclusion criteria used to select articles appropriate?*
3. *Was the assessment of studies reproducible?*
4. *Were the design and/or methods and/or topic of included studies broadly comparable?*
5. *How reproducible are the overall results?*
6. *Will the results help resource allocation in healthcare?*

Box 2: What are the results of the systematic reviews of effects?

As a guide to reading and interpreting the results of a systematic review of effects, policy makers may want to consider the following questions (adapted from [27, 46, 47])*:

- *What estimate of effect is presented?*
  Many reviews present an average estimate of effect across the included studies. This is often in the form of a risk ratio, odds ratio or standardised mean difference
- *Is an average estimate of effect across studies appropriate?*
  Reviews use statistical methods to summarise and combine outcome data from the studies included in the review. It is useful to consider whether the included studies were sufficiently similar, in terms of population, intervention, comparison and outcomes measured, to make combining the outcome data appropriate. Where an average estimate of effect is not possible, reviews usually present a narrative overview of the available data
- *Are confidence limits for the estimate of effect presented?*
  The review should present confidence intervals around the average estimate of effect. The wider the confidence interval, the less certain we can be about the true magnitude of the effect
- *If the results of subgroup analyses are reported, are these appropriate?*
  A review may present findings for a particular subgroup of participants across all trials or for a subgroup of studies [48]. For example, a review of interventions to reduce diarrhoeal diseases in children less than 5 years of age might also consider the effects of the interventions on children of less than 1 year of age. Similarly, a review may include a subgroup analysis of studies judged as having a low risk of bias. A subgroup analysis should make sense in relation to the overall review question and prior knowledge of factors that may influence or moderate the effects of the intervention. For example, it might be anticipated that a higher intensity intervention may produce larger effects. Subgroup analyses should be planned before the review is undertaken and their results should be interpreted with caution. This is because they are less reliable than analyses based on all of the included trials and because multiple statistical analyses may produce positive findings by chance alone
- *If there is 'no evidence of effect' is caution taken not to interpret this as 'evidence of no effect'?*
  'No evidence of effect' is not the same as 'evidence of no effect'. The former suggests that insufficient evidence is available to draw conclusions regarding the effects of the intervention in question. The latter suggests that there is clear evidence from the included studies that the intervention does not have the anticipated effects [49]
- *Do the conclusions and recommendations (if any) flow from both the original review question and the evidence that is presented in the review?*
  It is important to consider whether the conclusions presented by the review authors emerge directly from the data gathered from the review, and do not go beyond this evidence
- *Is the evidence applicable to the policy question under consideration?*
  Differences in health systems can mean that a programme or intervention that works in one setting may not work the same way in another. Policymakers need to assess whether the research evidence from a review applies in their setting. Guidance on this is presented in Article 8 in this series [22]

* There is some overlap between the questions listed here and those intended to guide assessment of the reliability of systematic reviews. This is because reliability is an important element of assessing and understanding the results of a systematic review

Box 3: What should policymakers do when different systematic reviews that address the same question have different results?

When looking for evidence to inform a particular policy decision, it is not uncommon to identify more than one relevant systematic review. Sometimes the results of these reviews may be different, leading the review authors to draw different conclusions regarding the effects of the intervention. This scenario differs from one in which the findings of two or more reviews agree, but researchers or others disagree on the interpretation of these findings [15].

There are many reasons why the results of different systematic reviews may differ. These include differences in: the question addressed by the review, the inclusion and exclusion criteria used, which data were extracted from the studies, how study quality was assessed, and decisions regarding (and methods for) statistical analysis of the data [15].

The following series of questions have been developed by Jadad and colleagues to assist in identifying and addressing the causes of discordance [15]:

- Do the reviews address the same question? If not, the review chosen should be that which addresses a question closest to that of the policy question for which evidence is needed, or which assesses outcomes most relevant to the policy question
- If the reviews address the same question, do they include the same trials or primary studies? If they do not include the same trials, the review including studies most relevant to the policy question being considered should be selected
- If the reviews include the same studies, are the reviews of the same quality? If not, the higher quality review should be used

Where both reviews are relevant, for example where they address different aspects of the same question, it may be useful to draw evidence from both.

References

1. Oxman A, Lavis JN, Fretheim A, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 1. What is evidence-informed policymaking**. Health Res Policy Syst, In Press

2. Lavis JN, Posada FB, Haines A, Osei E: **Use of research to inform public policymaking.** *Lancet* 2004, **364:** 1615-21.

3. Oxman A, Lavis JN, Fretheim A, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 10. Incorporating economic evidence**. Health Res Policy Syst, In Press

4. Carlsen B, Glenton C, Pope C: **Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines.** *Br J Gen Pract* 2007, **57:** 971-978.

5. Mays N, Pope C, Popay J: **Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field.** *J Health Serv Res Policy* 2005, **10 Suppl 1:** 6-20.

6. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J: **Patient adherence to tuberculosis treatment: a systematic review of qualitative research.** *PLoS Med* 2007, **4:** e238.

7. Pound P, Britten N, Morgan M, Yardley L, Pope C, Daker-White G *et al*.: **Resisting medicines: a synthesis of qualitative studies of medicine taking.** *Soc Sci Med* 2005, **61:** 133-155.

8. Mulrow CD: **Rationale for systematic reviews.** *BMJ* 1994, **309:** 597-599.

9. Oxman AD, Schunemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006, **4:** 20.

10. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M *et al*.: **Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals.** *JAMA* 1998, **280:** 278-280.

11. Jorgensen AW, Hilden J, Gotzsche PC: **Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review.** *BMJ* 2006, **333:** 782.

12. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M *et al*.: **Systematic reviews and meta-analyses on treatment of asthma: critical evaluation.** *BMJ* 2000, **320:** 537-540.

13. Linde K, Willich SN: **How objective are systematic reviews? Differences between reviews on complementary medicine.** *J R Soc Med* 2003, **96:** 17-22.

14. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C: **Evaluating meta-analyses in the general surgical literature: a critical appraisal.** *Ann Surg* 2005, **241:** 450-459.

15. Jadad AR, Cook DJ, Browman GP: **A guide to interpreting discordant systematic reviews.** *CMAJ* 1997, **156:** 1411-1416.

16. Oxman AD, Glasziou P, Williams JW, Jr.: **What should clinicians do when faced with conflicting recommendations?** *BMJ* 2008, **337:** a2530.

17. Canadian Coordinating Office for Health Technology Assessment. Proposed Evaluation Tools for COMPUS. 29-11-2005. Ottawa, Ottawa: Canadian Coordinating Office for Health Technology Assessment.

18. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF *et al*.. Systems to rate the strength of scientific evidence [Evidence report/technology assessment no 47]. Publication No. 02-E016. 2002. Rockville, MD, USA, Agency for Healthcare Research and Quality.

19. Critical Appraisal Skills Programme. 10 questions to help you make sense of reviews. 2006. United Kingdom, Public Health Resource Unit.

20. Oxman AD, Guyatt GH: **Validation of an index of the quality of review articles.** *J Clin Epidemiol* 1991, **44:** 1271-1278.

21. Higgins JPT, Altman DF: **Chapter 8: Assessing risk of bias in included studies.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.

22. Lavis JN. **SUPPORT Tools for evidence-informed health policymaking (STP). 8. Assessing the applicability of systematic reviews**. Health Res Policy Syst, In Press

23. Counsell C: **Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews.** *Ann Intern Med* 1997, **127:** 380-387.

24. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]*. The Cochrane Collaboration; 2008.

25. Meremikwu MM, Donegan S, Esu E: **Chemoprophylaxis and intermittent treatment for preventing malaria in children.** *Cochrane Database of Systematic Reviews* 2008, **2:** Art. No.: CD003756.

26. Mulrow CD: **The medical review article: state of the science.** *Ann Intern Med* 1987, **106:** 485-488.

27. Oxman AD: **Checklists for review articles.** *BMJ* 1994, **309:** 648-651.

28. Lefebvre C, Manheimer E, Glanville J, on behalf of the Cochrane Information Retrieval Methods Group: **Searching for studies.** In *Cochrane Handbook for systematic reviews of interventions. Version 5.0.1 [updated September 2008]*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.

29. Dickersin K, Min YI: **Publication bias: the problem that won't go away.** *Ann N Y Acad Sci* 1993, **703:** 135-146.

30. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: **Publication bias in clinical trials due to statistical significance or direction of trial results.** *Cochrane Database Syst Rev* 2009, MR000006.

31. Hopewell S, Clarke M, Stewart L, Tierney J: **Time to publication for results of clinical trials.** *Cochrane Database Syst Rev* 2007, MR000011.

32. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C *et al*.: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.** *BMC Med Res Methodol* 2007, **7:** 10.

33. Lavis JN. **SUPPORT Tools for evidence-informed health policymaking (STP). 3. Defining the problem**. Res Policy Syst, In Press

34. Lavis JN, Wilson M, Oxman A, Grimshaw J, Lewin S, Fretheim A. **SUPPORT Tools for evidence-informed health policymaking (STP). 4. Framing options to address a problem**. Health Res Policy Syst, In Press

35. Katrak P, Bialocerkowski A, Massy-Westropp N, Kumar VS, Grimmer K: **A systematic review of the content of critical appraisal tools.** *BMC Medical Research Methodology* 2004, **4:** 22.

36. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S: **Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists.** *Control Clin Trials* 1995, **16:** 62-73.

37. Shepperd S, Doll H, Broad J, Gladman J, Iliffe S, Langhorne P *et al*.: **Early discharge hospital at home.** *Cochrane Database Syst Rev* 2009, CD000356.

38. Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B *et al*.: **Lay health workers in primary and community health care.** *Cochrane Database Syst Rev* 2005, CD004015.

39. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD: **Audit and feedback: effects on professional practice and health care outcomes.** *Cochrane Database Syst Rev* 2006, CD000259.

40. Arblaster L, Lambert M, Entwistle V, Forster M, Fullerton D, Sheldon T *et al*.: **A systematic review of the effectiveness of health service interventions aimed at reducing inequalities in health.** *J Health Serv Res Policy* 1996, **1:** 93-103.

41. Siegfried N, Muller M, Deeks JJ, Volmink J: **Male circumcision for prevention of heterosexual acquisition of HIV in men.** *Cochrane Database Syst Rev* 2009, CD003362.

42. McGuinness B, Craig D, Bullock R, Passmore P: **Statins for the prevention of dementia.** *Cochrane Database Syst Rev* 2009, CD003160.

43. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A: **Synthesising qualitative and quantitative evidence: a review of possible methods.** *J Health Serv Res Policy* 2005, **10:** 45-53.

44. Noyes J, Popay J, Pearson A, Hannes K, Booth A: **Chapter 20: Qualitative research and Cochrane reviews.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008).* Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.

45. Jefferson T, Demicheli V, Vale L: **Quality of systematic reviews of economic evaluations in health care.** *JAMA* 2002, **287:** 2809-2812.

46. Moher D, Jadad AR, Klassen TP: **Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions?** *Arch Pediatr Adolesc Med* 1998, **152:** 915-920.

47. Oxman AD, Cook DJ, Guyatt GH: **Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group.** *JAMA* 1994, **272:** 1367-1371.

48. Oxman A, Lavis JN, Fretheim A, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 9. Incorporating equity considerations**. Health Res Policy Syst, In Press

49. Oxman A, Lavis JN, Fretheim A, Lewin S. **SUPPORT Tools for evidence-informed health policymaking (STP). 12. Dealing with insufficient evidence**. Health Res Policy Syst, In Press

Evidence-informed decision-making (EIDM) refers to the systematic and transparent use of research evidence, along with other considerations important in decision-making, such as context, acceptability to stakeholders, feasibility of implementation, and equity (Carter et al. 2008, Oxman et al.Â Standardized methods exist for the preparation of systematic reviews and, in particular, the Cochrane Collaboration has made important efforts in their methods, design, use and dissemination. Standardized processes and methodologies have also been developed for the other evidence products, including clinical practice guidelines, health technology assessments, and policy briefs.Â In the case of the formulation of evidence-informed policy options the SUPPORT. 3. SUPPORT Tools for evidence-informed health Policymaking (STP) 9: Assessing the applicability of the findings of a systematic review. Health Research Policy and Systems, 7(Suppl 1), S9. http://doi.org/10.1186/1478-4505-7-S1-S9 6. Potential Causes of Bias and Error in Evaluating Study Quality and Applicability. n Bias u Evaluator influenced by conflict of interest u Evaluator influenced by another reviewer. n Error u Inadequate specification of assessment criteria u Limited reliability of assessment instruments u Inadequate training and auditing. 7. What Can Be Done to Minimize Bias and Error in... The series addresses four broad areas: supporting evidence-informed policymaking; identifying needs for research evidence; finding and assessing research evidence; and, going from research evidence to decisions.Â This is a set of 18 tools that can be used by those involved in finding and using research evidence to support evidence-informed health policy making. The series addresses four broad areas: supporting evidence-informed policymaking; identifying needs for research evidence; finding and assessing research evidence; and, going from research evidence to decisions. Serial Title: Health Research Policy and Systems. Serial Part: Volume 7, Suppl 1. Notes: The introduction and each tool can be downloaded separately. View webpage for full text.