Guides for Reading and Interpreting Systematic Reviews

II. How Did the Authors Find the Studies and Assess Their Quality?

Alejandro R. Jadad, MD, DPhil; David Moher, MSc; Terry P. Klassen, MD, MSc

One of the most powerful arguments used by the supporters of systematic reviews is that they overcome most of the limitations of narrative reviews by being the product of a scientific process to reduce bias and imprecision and by providing detailed information to allow replication by others.1,2 Two of the most effective mechanisms for a systematic review to reduce bias and imprecision are including the maximum possible number of relevant individual trials and providing a detailed description of their strengths and limitations. We have structured this article to serve 2 purposes. First, we describe the characteristics of the ideal search, the limitations and decisions that most reviewers face when deciding how to search the literature, and the aspects of a report that readers should evaluate to assess the comprehensiveness and appropriateness of the search strategy. Second, we describe the limitations and decisions that most reviewers face when deciding how to assess trial quality and the aspects of a report that readers should evaluate to determine how trial quality has been assessed and the appropriateness of the assessments.

DID THE AUTHORS SEARCH THE LITERATURE COMPREHENSIVELY?

Ideally, a systematic review includes all the relevant trials available.3 Identifying all the relevant trials for systematic reviews has been recognized as a “most fundamental challenge.”1 Identification of all relevant trials for a systematic review is possible only by scanning all records in available bibliographic databases; by hand searching all journals, theses, proceedings, and textbooks that are not indexed in any of the databases, including the bibliographies of all identified relevant reports; and by obtaining relevant, unpublished information from investigators and/or organizations involved in completed or ongoing trials.4

In practice, constrained by time and by cost, reviewers must strive to identify the maximum number of eligible trials, hoping that the studies included in the review will be a representative sample of all eligible trials.4 Given that searching the literature can be an onerous, resource-consuming task, reviewers have to set priorities regarding what sources to use to identify trials in the most cost-effective way. The rule of thumb is that reviewers should select as many sources as their resources allow.

Each of the 3 main sources of trials has advantages and disadvantages. Bibliographic databases are the most cost-effective source, followed by hand searches, and direct contact with investigators and sponsoring organizations.

Bibliographic Databases

Bibliographic databases vary in scope, currency, accessibility, and cost. The most widely used and studied database in health care is MEDLINE, which is produced by the US National Library of Medicine. It covers January 1, 1966, to the present, is updated weekly, indexes more than 4000 journals, and has citations for more than 7 000 000 articles.3 It is available on-line and in CD-ROM. The corresponding print
product is Index Medicus. Another commonly used electronic bibliographic database is EMBASE (Excerpta Medica Database), which is produced in Europe. Other databases include Canadian Government and Research Documents, CANCERLIT, CARL’s UnCover database, CINAHL, Conference Papers Index, Current Contents, Directory of Published Proceedings, Dissertations Abstracts Online, Federal Research in Progress, HEALTHStar, Masters Abstract International, Medical Outcome and Guideline Alert, PDQ, SciSearch, and EMBASE. The main advantage of electronic databases is that users may search for numerous concepts in many references simultaneously across multiple sources.3 Inadequate indexing to satisfy the needs of most reviewers is their main disadvantage. This inadequacy occurs because of the poor quality of reporting of the original trials, coding systems that cannot include every possible coding term, and inconsistencies between indexers. Multiple studies have shown, for instance, that identification of all published articles indexed is not possible, and that systematic reviews of randomized clinical trials that depend solely on MEDLINE searches to identify relevant studies will miss about half of them.6

In 1996, the Cochrane Collaboration7,8 released an electronic library, The Cochrane Library. It was designed to provide evidence to support informed health care decision making. This library contains 4 databases, one of which, The Cochrane Controlled Trials Register, contains references of more than 110,000 controlled trials identified through searches of electronic databases and hand searches of biomedical journals. This database is likely to become an important source of trials for reviewers in the future. As with any other database, however, adequate coding is needed if reviewers are to find what they want.

Hand-Search Sources

The main advantage of hand searching is accuracy. It can increase the yield of a search by retrieving trials that may have been missed in a database. The main disadvantage of hand searching is that it is time-consuming. Therefore, the potential increase in the search yield needs to be balanced with the required time and costs. The following 3 main types of sources can be hand searched: reference lists of all eligible trials identified by other methods, information sources that are not indexed in electronic databases, and sources that are indexed to ensure that eligible trials not captured by database searching are identified. Among these, the first is the simplest and should be a minimum requirement for every search. The latter two are much more time consuming and can be performed as time and resources allow. The burden of hand searching, however, will be progressively reduced as the number of journals hand searched by members of the Cochrane Collaboration increases.

Contacting Researchers and Sponsoring Organizations

Reviewers can also contact experts in the field who can identify additional relevant published and unpublished trials. Corresponding authors of relevant papers can also provide contacts for finding additional published and unpublished trials. Response, however, is variable, and multiple reminders are often required. The time elapsed between the publication of their primary studies and the reviewer’s request can also be a factor. Readers must be aware that the variable response of researchers to requests by reviewers provides yet another justification for the need to create registers of ongoing trials in all areas of health care research.9,10

LOOKING AT THE BREADTH AND DEPTH OF THE LITERATURE SEARCH

Reports of systematic reviews should include detailed information on the following 2 aspects of the literature search: the sources searched and the methods used to search them.

Ideally, the description of the sources searched should include the number and names of the sources, the date on which they were searched, and the period covered by the search. These elements allow the reader to judge the breadth of the search. For instance, a description of a bibliographic database should read “HEALTHStar (OVID for UNIX version 3.0) was searched in May 1996 for January 1975 to December 1995.”

Once the reader has identified the sources searched, the next step is to establish what strategies were used to search each source. The latter two are more time consuming. Therefore, the potential increase in the search yield needs to be balanced with the required time and costs. The following 3 main types of sources can be hand searched: reference lists of all eligible trials identified by other methods, information sources that are not indexed in electronic databases, and sources that are indexed to ensure that eligible trials not captured by database searching are identified. Among these, the first is the simplest and should be a minimum requirement for every search. The latter two are much more time consuming and can be performed as time and resources allow. The burden of hand searching, however, will be progressively reduced as the number of journals hand searched by members of the Cochrane Collaboration increases.

Contacting Researchers and Sponsoring Organizations

Reviewers can also contact experts in the field who can identify additional relevant published and unpublished trials. Corresponding authors of relevant papers can also provide contacts for finding additional published and unpublished trials. Response, however, is variable, and multiple reminders are often required. The time elapsed between the publication of their primary studies and the reviewer’s request can also be a factor. Readers must be aware that the variable response of researchers to requests by reviewers provides yet another justification for the need to create registers of ongoing trials in all areas of health care research.9,10

LOOKING AT THE BREADTH AND DEPTH OF THE LITERATURE SEARCH

Reports of systematic reviews should include detailed information on the following 2 aspects of the literature search: the sources searched and the methods used to search them.

Ideally, the description of the sources searched should include the number and names of the sources, the date on which they were searched, and the period covered by the search. These elements allow the reader to judge the breadth of the search. For instance, a description of a bibliographic database should read “HEALTHStar (OVID for UNIX version 3.0) was searched in May 1996 for January 1975 to December 1995.”

Once the reader has identified the sources searched, the next step is to establish what strategies were used to search each source. The following 3 main types of sources can be hand searched: reference lists of all eligible trials identified by other methods, information sources that are not indexed in electronic databases, and sources that are indexed to ensure that eligible trials not captured by database searching are identified. Among these, the first is the simplest and should be a minimum requirement for every search. The latter two are much more time consuming and can be performed as time and resources allow. The burden of hand searching, however, will be progressively reduced as the number of journals hand searched by members of the Cochrane Collaboration increases.

Contacting Researchers and Sponsoring Organizations

Reviewers can also contact experts in the field who can identify additional relevant published and unpublished trials. Corresponding authors of relevant papers can also provide contacts for finding additional published and unpublished trials. Response, however, is variable, and multiple reminders are often required. The time elapsed between the publication of their primary studies and the reviewer’s request can also be a factor. Readers must be aware that the variable response of researchers to requests by reviewers provides yet another justification for the need to create registers of ongoing trials in all areas of health care research.9,10

LOOKING AT THE BREADTH AND DEPTH OF THE LITERATURE SEARCH

Reports of systematic reviews should include detailed information on the following 2 aspects of the literature search: the sources searched and the methods used to search them.

Ideally, the description of the sources searched should include the number and names of the sources, the date on which they were searched, and the period covered by the search. These elements allow the reader to judge the breadth of the search. For instance, a description of a bibliographic database should read “HEALTHStar (OVID for UNIX version 3.0) was searched in May 1996 for January 1975 to December 1995.”

Once the reader has identified the sources searched, the next step is to establish what strategies were used to search each source. These will allow the reader to judge the depth of the search. A detailed description of how to effectively search the possible sources of trials is beyond the scope of this series. A number of tips that readers could use to judge the adequacy of the searches follow.

For bibliographic databases, search strategies usually include terms from the following 3 categories: terms related to the condition (eg, attention-deficit disorder with hyperactivity); terms related to the intervention (eg, clonidine or pemoline); and terms related to the method of the trials (eg, randomized, controlled trials). These terms are frequently linked with Boolean operators (usually AND or OR), which determine how the terms are combined.11 Synonyms are usually linked with OR. The OR operator ensures that any record containing any one of the terms will be retrieved; eg, combining the search terms clonidine OR pemoline means that all records containing either term will be retrieved. Thus, an OR operator broadens the search. Combining terms with AND allows searching for more than 1 concept at a time; eg, combining attention-deficit disorder with hyperactivity AND clonidine will retrieve articles that discuss both terms only. The AND operator narrows a search.

Different databases use terms called qualifiers to specify which part, or fields, of each record to search. For instance, text word qualifiers may be used to search for a term in the title and abstract fields. Ideally, the whole search strategy (the terms and how they were combined) should be provided in a table or appendix. The Table illustrates an example of a MEDLINE search strategy using different approaches.
and infrequently studied interven-
tors, of course) should keep in mind
the breadth and depth of the litera-
ture, depending on the direction of the
findings. This tendency, which appears to favor tri-
als with positive results, has been
called publication bias.15 Systematic
reviews that fail to identify and in-
clude unpublished trials, therefore,
are at risk of overestimating the ef-
fect of the interventions they evalu-
ate. Readers, however, must take into
account that identifying unpub-
lished trials is even more difficult
than identifying published trials.
Without compulsory registration of
trials at inception, it is almost im-
possible to know how many unpub-
lished trials exist. Readers must also
realize that even if unpublished tri-
als are identified, some journal edi-
tors are reluctant to publish re-
views that include unpublished data.17 In any case, readers should look for information on the review indicating whether the reviewers
made efforts to identify unpublished
data or whether they made any efforts to estimate (statistically
or graphically) the potential im-
 pact of unpublished studies on the
results of the review.

To be pragmatic (there is no
empirical evidence supporting this),
we suggest that a review with a com-
prehensive search uses at least 3
 sources and provides a description
of efforts to identify unpublished
 trials. A particularly effective com-
bination could be 1 bibliographic
database (eg, MEDLINE or The
Cochrane Library), a hand search of
reference lists of eligible trials, and
direct contact (by mail, fax, e-mail,
and/or telephone) with the corre-
sponding authors of eligible trials
asking for additional published or
unpublished trials. Such a review
should include a discussion of the
search’s limitations.

### HOW DID THE AUTHORS EVALUATE THE QUALITY OF THE TRIALS INCLUDED?

If there is 1 issue about which most
supporters and detractors of system-
atric reviews agree, it is the rel-
evance of the dictum “garbage in,
garbage out.” This means that the extent
to which a systematic review (or
more particularly, a meta-analysis)
could guide health care decisions de-
dpends on the quality of the trials
available. How to assess trial qual-
ity as part of a systematic review, or
even if it should be assessed at all,
remains surrounded by great con-
troversy. The positions of research-
erists in this field vary from those
who regard quality assessment as an im-
portant strategy to identify and re-
duce bias, to those who see quality
assessment as a source of bias or as
completely uninformative.18,19

It would be ideal if reviewers
only dealt with perfect trials, thus not
having to worry about the quality of
the trials included in their reviews.
Among other things, those trials would have (1) to answer clear and relevant clinical questions; (2) to be designed, conducted, and reported by researchers who did not have conflicts of interest; (3) to follow strict ethical principles; (4) to include all patients available; (5) to evaluate all possible interventions for all possible variations of the conditions of interest, in all possible types of patients, in all settings, and using all relevant outcome measures; (6) to include strategies to eliminate bias during the administration of the interventions, during the evaluation of the outcomes, and during reporting of the results, thus reflecting the true effect of the interventions; (7) to include perfect statistical analyses; and (8) to be described in clear and unambiguous language, including an exact account of all the events that occurred during the design and conduct of the trial, individual patient data, and an accurate description of the patients who were included, excluded, and withdrawn and who dropped out.

Unfortunately, real-life reviewers only have imperfect trials to review and face many barriers to determine their quality with confidence.

One of the main barriers that hinders the assessment of trial quality is the complex concept or construct of quality. As with any other construct (ie, anxiety, happiness, or love), quality can be acknowledged without difficulty, but it is not easy to define or measure.

Another major barrier hindering the assessment of trial quality is that, in most cases, the only way a reviewer may assess quality is by relying on the information contained in the written report. The problem is that a trial with a biased design that is well reported could be judged as having high quality, while a well-designed but poorly reported trial could be judged as having low quality.18

Both important barriers may dissuade some reviewers from assessing trial quality, whereas others may decide to include quality assessments in their systematic reviews.

Taking this into account, the first question that a reader of a systematic review must ask in relation to the assessment of trial quality is: DID THE REVIEWERS ASSESS TRIAL QUALITY?

This is usually an easy question to answer. The information is usually found in the “Methods” section of the reviews. Most reviewers will use the term quality, but others may use terms such as adequacy, validity, or rigor. If the reader does not find any of these words, efforts should be made to determine whether the reviewers included in the report any other information suggesting that they used specific criteria to value some trials more than others. If all of these efforts fail, the readers may conclude that the reviewers did not assess trial quality in that particular review. If this is the case, the readers should look for information in the report describing why the reviewers did not assess trial quality. If such information is not available—and this is usually the case—then the readers will have to judge, given the information available in the report, whether this decision was appropriate.

If the report contains a clear statement confirming that the reviewers assessed trial quality as part of the systematic review, or even information suggesting that trial quality was assessed, the next step is to establish whether the report includes a description of the methods used to assess trial quality. If such description is available, the reader should answer the following question:

HOW WAS QUALITY ASSESSED?

The assessment of trial quality includes at least the following 3 components: the definition of quality, the tool used to generate the assessments, and the circumstances in which the assessments are generated. We briefly describe the most relevant aspects of each of these components.

Definition of Quality

Quality may mean different things to different people. Reviewers can decide to assess the quality of trials focusing on any methodological aspect, from the design to the reporting. They may focus on the generalizability of the results, or on a combination of these.18,20 Specific aspects of trials that have been used to assess trial quality include the literary style of the report, the clinical relevance of the research question, the likelihood of producing biased results, the precision and extent to which it is possible to generalize the results, the appropriateness of the statistical analysis, the presentation of the data, and the ethical implications of the intervention evaluated.18,20 Our preference has been to focus on 1 aspect of methodological quality, internal validity, which we define as “the confidence that the trial design, conduct, analysis, and presentation have minimized or avoided biased comparisons of the interventions under evaluation.”18 Given that each approach has a different purpose, it would be inappropriate for the readers to judge whether any particular definition is right or wrong. Instead, the readers should establish whether the report includes enough information to allow them to understand what was assessed.

On occasions, however, determining the meaning of quality is not possible, because the reports may include only the name of the tool used by the reviewers and a bibliographic reference for it. Obtaining a hard copy of such a reference may be appropriate when the quality assessments may have had an important effect on the results of the review and when the results are likely to have a significant impact on important decisions. An alternative could be to obtain that information from a recently published annotated bibliography of scales and checklists.21 The number of tools, however, seems to be increasing as the number of systematic reviews increases.

Assessment Tools

Once the reader has established the meaning of quality used by the reviewers, the next step is to look for information on the tools they used to generate the assessments. Tools to assess trial quality could be classified broadly into those that include individual components and those that include groups of components.

A component represents an item that describes a single aspect of qual-
Assessing trial quality using a component can be achieved by scoring the component as present or absent or by judging the adequacy of the information available in the report in relation to the component. For instance, concealment of patient assignment could be judged as present or absent, or, depending on the information provided by the report, it could be judged as adequate, unclear, or inadequate.\(^2\) There is empirical evidence suggesting a relationship between specific components and the likelihood of bias in trials; eg, in trials in which patients are not allocated to the study groups randomly, concealment of allocation is inadequate or unclear, or binding is inadequate are more likely to exaggerate treatment effects.\(^2\) However, although individual components are quick and easy to score, they provide minimal information about the overall quality of trials.\(^1\)

The narrow view provided by individual components can be overcome by using several components grouped in checklists or scales. The main difference between a checklist and a scale is that, in a checklist, the components are evaluated separately and do not have numerical scores attached to them, whereas, in a scale, each item is scored numerically, and an overall quality score is generated.\(^1\) A recent systematic search of the literature identified 9 checklists and 25 scales.\(^2\) Only 1 scale has been developed following established methodological procedures.\(^2\) The number of tools to assess trial quality is increasing.

**Generation of the Assessments**

In addition to information on the definition of quality and the tools used to generate the assessments, ideal reports of the reviews contain information on the number and background of the individuals who generated the assessments, on the extent to which their assessments agreed, on the methods used to reconcile discrepant assessments, and on whether the assessments were generated under masked conditions.

After trying to establish how the quality of the trials has been assessed in the review, the reader must try to answer a final question:

**WHAT DID THE REVIEWERS DO WITH THE ASSESSMENTS?**

Reviewers can follow different approaches to incorporate quality assessments into systematic reviews. The simplest approach is the provision of a description of the quality assessments, usually in tabular form, with the only purpose to inform the reader about the credibility of the evidence provided by the trials reviewed.

There are at least 5 other approaches to incorporate quality assessments into systematic reviews.\(^2\)\(^-\)\(^2\)\(^4\) For instance, quality assessments could be used (1) to include or exclude trials from a review; (2) to conduct sensitivity analyses allowing comparisons between the results of trials with different quality; (3) to display graphically the results of each of the trials according to their quality (eg, the trials are displayed in descending order, starting with the one with the highest quality); (4) to perform cumulative meta-analyses using quality assessments as the input sequence; and (5) to weight trials according to their quality.

The reader must be aware that there is very little research evidence evaluating the impact of any of these methods or guiding their selection.\(^1\) Given the potential influence of quality assessments on the results of reviews, readers should rely perhaps more on those reviews that report the results with and without quality assessments and/or that provide enough information to allow the reader, if necessary, to calculate the effect estimates using different approaches to incorporate quality assessments.

**SHOULD THE ASSESSMENTS BE PERFORMED UNDER MASKED CONDITIONS?**

It has been suggested for more than 10 years that the quality of trial reports should be assessed without the knowledge of the authors, institutions, sponsorship, publication year and journal, or study results (ie, under masked conditions).\(^2\) However, there are only 2 published empirical studies addressing this issue.\(^1\)\(^,\)\(^2\)\(^7\) One of these studies showed that assessments under these conditions were more likely to yield lower and more consistent scores than assessments under open conditions.\(^2\) These results have been replicated, at least in relation to the lower scores.\(^2\) These results imply that bias could be introduced by assessments under open conditions. There is evidence, however, suggesting that data extraction under masked conditions does not affect the overall results of systematic reviews.\(^2\) Masked conditions would reduce the likelihood of bias, but would also increase the resources required to conduct the reviews. Given the methodological and financial implications of these findings, and the little empirical evidence available, we would not recommend it as a mandatory step in a systematic review.

**CONCLUSIONS**

Searching the literature for relevant individual studies and assessing their quality are 2 of the most important and controversial components of systematic reviews. They are also the foci of intense research activity. Therefore, as stated in the first article of this series, we do not pretend to provide the last word (ie, written in stone) and intend to update the series accordingly. In the meantime, however, we encourage readers to take into account the following principles and issues:

- As the literature continues to grow, it is unlikely that any reviewer or group of reviewers can search all the possible sources of trials for a review.
- Reviewers should include enough information in the reports of systematic reviews to inform readers of the breadth and depth of their searches.
- Readers must realize that poor reporting does not necessarily reflect a poor literature search. Conversely, an incomplete search could be well reported.
- The final evaluation of the comprehensiveness of a literature search is highly subjective. Reviews that include only 1 source, however, are unlikely to be comprehensive.
- The assessment of trial quality could provide valuable information to reviewers during the con-
duct of systematic reviews and to readers who rely on systematic reviews to guide their health care decisions, but could also lead to biased or inaccurate conclusions.

- An appropriate rule of thumb for readers is to rely more on those reviews that report clearly the definition of quality, describe the tools and circumstances to obtain the assessments and the methods used to incorporate such assessments in the review, and provide enough information to allow the reader to re-calculate effect estimates using different approaches.

This is the second of a 3-part series, the first of which appeared in our July issue (accompanied by an editorial) and the third of which will run in our September issue.

Accepted for publication December 29, 1997.

Reprints: David Moher, MSc, Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Ontario, Canada K1H 8L1 (e-mail: 107656,3375@compuserve.com).

REFERENCES


