Essential Skills for Evidence-based Practice Understanding and Using Systematic Reviews

Jeanne Grace

Abstract
Systematic reviews are comprehensive reviews of quantitative research literature conducted according to explicit guidelines established in advance. The strengths of systematic reviews as sources of evidence for practice are the exhaustive nature of the literature search, the objective criteria for determining which research will be included, and the methods used to synthesize the findings from each study into comprehensive conclusions. Guidelines for appraising systematic reviews focus on these strengths and whether the results of the systematic review can be applied to practice.

Keywords: systematic reviews, nursing practice, evidence-based practice
Within the “5S” hierarchy of sources of evidence, systematic reviews occupy the second step from the bottom, syntheses. When clinicians have synopses, summaries or expert systems available to guide practice, they may not need to refer directly to the systematic reviews or single studies that form the basis for these more usable sources of evidence. In many areas of nursing practice, however, synopses, summaries and expert systems are not yet available. The systematic review then stands as the strongest source of evidence available to address a clinical problem.

Reviews of the literature are performed for many purposes, and those purposes influence the selection of literature to be included. A review article may seek to summarize the “state of the science” for a broad question or illustrate the history of the development of some field of scholarship. Research studies may be chosen for inclusion in an integrative or conceptual literature review because they support a theoretical framework the reviewer is proposing. Some review articles are intended to summarize the background information needed to understand a clinical problem. Still others are intended to support the reviewer’s claim that a specific research study needs to be conducted. In contrast, a systematic review is intended to synthesize all the “best evidence” to guide clinical care for a specific health problem. Most systematic reviews address therapy questions, that is, the effective way(s) to achieve some desired health goal. Systematic reviews of diagnosis, prognosis and harm questions are much less common.

Systematic reviews may be published in professional journals or made available through evidence-based practice institutes, like the Joanna Briggs Institute. The largest single source is the Cochrane Collaboration, which commissions original systematic reviews on a wide range of clinical topics (Cochrane Reviews) and maintains a listing of systematic reviews created elsewhere that meet Cochrane standards (Database of Reviews of Effectiveness – DARE).

The unique aspect of a systematic review as a source of evidence is the approach taken to identifying and appraising evidence for inclusion. In much the same way that a rigorous clinical trial has a study plan to control confounding influences on study outcomes (i.e. bias), a systematic review establishes an advance plan to control confounding influences on the identification and inclusion of evidence. The goal is to assure that the selection and analysis of research in the review will allow accurate conclusions to be drawn from the resulting synthesis.

### Reporting Bias

What sorts of confounding influences could affect a review of clinical evidence? There are a series of influences known collectively as reporting bias. Researchers are more likely to seek publication for studies when the results support their theories and/or are statistically significant. Moreover, findings that are novel or strongly statistically significant are more likely to be published in English, appear in journals that are indexed in electronic databases and receive “fast track” publication. Further, these findings are more likely to be cited by other authors. As a result, studies identified by means of even the most comprehensive electronic literature search may present an overly optimistic sample from the true knowledge base for some clinical problem.

As a protection against drawing conclusions based only on published evidence, the plan (often called a “protocol”) for a systematic review defines the ways in which the authors of the review will seek all the evidence, not just that published in indexed journals. This may include searching conference proceedings, reviewing the reference lists of published articles for additional studies, inspecting the table of contents for non-indexed journals and asking researchers known to be interested in the clinical problem to provide details and data for studies that were never published.

Once all the studies to be reviewed have been identified, the review authors may construct a funnel plot as a further check for reporting bias. For each independent study, some measure of the size of the treatment effect (x-axis) is plotted against some measure of the accuracy (for example, the standard error) of that estimate of the effect (y-axis). The reasoning is this: Study results reflect both the
“true” effect of the therapy and random error associated with small sample size, inexact measurement and other issues. Therefore, we would expect the results from comparable studies with the smallest error component (i.e. most accurate) to be quite similar to each other, while the results from studies with large error components would have a much wider, almost random distribution of result values. When a funnel plot is constructed, we would therefore expect that the points representing the results from the most accurate studies would form a tight cluster at the top of the plot, with results from less accurate studies spreading out below to both sides in the shape of an isosceles triangle or inverted funnel. The less accurate the results estimates are, the greater the spread. If one side of the funnel shape is not filled in with plotted points as well as the other side, reporting bias may be the explanation.

Quality Bias

The strength of evidence to support any clinical recommendation has three dimensions: quantity, quality and consistency. The synthesis of study results in a systematic review may also be confounded if the findings of studies with poor design (higher risk that results are biased) are considered equally with the findings of studies with good design.

For therapy questions, the highest quality design for studies is a true experiment conducted in a clinical setting (randomized clinical trial). Additional protections against confounding influences reduce the risk of bias within this general design. Strategies to assure that researchers cannot influence the assignment of subjects to supposedly randomized clinical trial groups would be one example. Rigorous randomization procedures are the best guarantee that the groups being compared were similar in all important ways at the beginning of the study. Another example would be concealing each subject’s treatment assignment from the subjects and the researchers who interact with them. When this is done, subjects’ and researchers’ expectations of what should happen cannot influence reports of what actually did happen.

A third set of concerns involves reporting of results. When the outcome data for some study subjects is missing, the remaining subjects with complete data may no longer represent the experience of the entire group. Thus, groups that were similar at the start of the study may have different outcomes from each other at the end of the study for some reason other than the therapy being tested (selective attrition). The ideal solution to avoid this confounding influence is to retain all subjects in the study until the end, but this is not always possible. When it is not, the researcher can do analyses of the available outcome data to estimate what impact the missing data may have on the conclusions of the study.

Even when there is no selective attrition among study subjects, researchers may be selective in the results they report. In a study with four planned outcomes, for example, the researcher may report the results for the only outcome where statistically significant group differences were found and neglect to report that no such differences were found for the other three outcomes. The ability to compare the original plan of the study to the reported outcomes is the best protection against this sort of bias and also protects against the reporting biases discussed above for entire studies. Many international scientific journals will no longer publish the results of clinical trials unless the plan of the research was registered in a public database before the study began.5

The authors of a systematic review must evaluate the risk of bias on the results of each study they are considering for inclusion in their review. This is often done using a rating tool of some sort so that every study is judged on the same criteria.4 The authors of a systematic review may deal with the confounding influences of poor study quality by excluding those studies from the evidence synthesis and basing all conclusions on high quality studies, only. Another option is retaining all studies, but conducting additional analyses. These sensitivity analyses determine whether including the results of poor quality studies with those of high quality studies changes the conclusions of the synthesis.
**Appraisal Bias**

The particular benefit of a systematic review as an objective source of synthesized “best evidence” is threatened if the review authors let personal bias influence their selection and evaluation of the evidence. The strategies used to reduce the risk of this bias include developing an advance plan for conducting the review, documenting the decisions made during the review and arranging independent confirmation for those decisions.

The advance plan, or protocol, for a systematic review is a detailed map of the road the authors of a systematic review intend to follow to identify and synthesize evidence for practice. Typically, the protocol specifies the clinical problem for which evidence is desired, makes the case for why the problem is important, and details the actions the authors intend to take to identify existing evidence. This detail may include the exact search terms that will be used to explore electronic databases and the databases that will be consulted. Much like the subject inclusion and exclusion criteria in the protocol for a clinical trial, the protocol for a systematic review will include explicit rules for deciding which studies to include in the review.

A particularly important part of the systematic review protocol is the advance plan for analyzing the results of included studies. The body of evidence about therapies for a focused clinical problem, for example, community-based group interventions for smoking cessation, may include studies sufficiently different from each other that they should not be directly compared. The authors of a systematic review will use their theoretical and clinical knowledge to define sub-groups that will be analyzed separately. For smoking cessation interventions, for example, studies that target pregnant women might be analyzed separately from studies that target adult smokers in general.

(The protocol for a Cochrane Collaboration systematic review is published before the review is conducted, so that other persons with interest in the specific clinical problem can suggest improvements or connect the review authors with unpublished studies to consider for inclusion. An entry for “protocol”, rather than “systematic review” in the Cochrane database indicates that the review on this topic is a work in progress, rather than a completed synthesis.)

Once the advance plan is established, the authors of a systematic review document the decisions they make as they follow that plan. The complete text of a Cochrane systematic review, for example, will identify all the studies retrieved and provide the reasons why each one was retained for the analyses or excluded from consideration. If there have been deviations from the analysis plan, the authors will explain their reasons for the change. This level of detail allows readers of the review to judge whether personal biases have influenced the review but may not be available in the shorter format for a systematic review published in a journal.

Another strategy used to limit the effect of personal bias is for two or more authors to make judgments about study inclusion and quality independently from each other. When each review author applies previously established standards to a study and reaches the same conclusion as the other author(s), we are reassured that it is the standard, not personal opinion, which guided the judgment.

**Evaluating Systematic Reviews**

The critical appraisal of many types of evidence for practice has a common organization, divided into three main sections:

- Are the results valid?
- What are the results?
- How can I apply the results to patient care?

For systematic reviews, questions about the validity of the review address the choice of topic, the search for studies and the quality of the studies found. A high quality systematic review has a clear focus on a specific and sensible clinical question, usually one about a therapy of some sort. The search for relevant studies has been thorough. The review has explicit and appropriate standards for selecting studies to be included, and ideally only high-quality studies (i.e. those with a low risk of biased results) are retained.
The evaluation of the quality of those studies, moreover, is objective, as demonstrated by use of checklists and agreement of independent raters.

While the quantity and quality of the evidence are addressed in the validity section of the evaluation, the consistency of the evidence across studies is addressed in the “What are the results?” section. In their analysis plan, the authors of the systematic review have identified which outcomes and studies are truly comparable to each other. If they have correctly grouped studies of the same phenomenon in similar study samples together, the results of those studies should be similar to each other.

The effect of some therapy on each individual outcome in each individual study is typically represented as a comparison of the value or likelihood of that outcome between the groups being studied. This results in a point estimate for the group difference and a range estimate of the accuracy of that point estimate (the confidence interval). The technique of meta-analysis combines this information for comparable studies, usually weighting each result by the sample size, and computes an average point estimate and confidence interval that best represents the studies as a whole. The original results and this computed statistical synthesis are often presented graphically as a forest plot.

Visual inspection of the forest plot may reveal whether the findings of individual studies seem consistent, or whether the results are so varied that it is not clear the studies really are comparable to each other. When displayed on a forest plot, the point estimates (short vertical lines in Figure 1) for the group comparison (the effect size) for each study should resemble a vertical stack if the results are truly consistent. The horizontal lines representing the confidence intervals for each study’s findings should all overlap each other if the results are consistent. (Figure 1). Overlap of confidence intervals for results across studies indicates that the comparisons between groups are probably really the same across the studies, except for random differences among the study samples.

Figure 1  Forest plots of consistent and inconsistent study results

![Forest plots of consistent and inconsistent study results](image)

In addition to the visual information about result consistency provided by the forest plot of comparable studies, the meta-analysis may include a statistical test of heterogeneity. This is similar to a significance test, because it estimates whether the differences among the results of individual studies are the result of random sampling only, and not some “real” differences among the studies. If the results across studies are consistent, the test of heterogeneity will NOT be statistically significant (will not have a p value of .05 or smaller). Do not confuse consistency of results across studies with the significance of group differences within studies. Consistent results may “agree” that a therapy has a
significant effect or the results may “agree” that the therapy had no effect.

Once the reader has considered whether the study results are consistent, the remaining questions in this section of the critical appraisal are about the nature of those results. How large are the treatment effects? Are those effects consistent across all subgroups of patients? If the outcomes are measured categorically, how many patients would need to receive the treatment to obtain one more desired outcome (number needed to treat - NNT)? The final question in this section of the critical appraisal addresses the precision of the results. (See 7 for a further discussion of this concept). When the results of comparable studies are synthesized, how great is the range of likely outcomes for any treatment?

The final section of the critical appraisal of a systematic review is the application of the synthesized results to patient care. This section asks questions that clinician / readers must answer based on their knowledge of the patients they serve. Given the size of the treatment effect and the likely range of variations in that size (precision), would this treatment result in a clinically important outcome for my patients? If so, for which ones? If subgroup analyses have been performed, would I expect my patients’ outcomes to be more like the subgroup outcomes than the overall outcomes?

Any treatment may have undesired outcomes (side effects) as well as desired ones. A systematic review is most useful for clinicians when it addresses all the clinically important outcomes, both desirable and undesirable. This information is essential when clinicians consider with their patients whether the costs and risks associated with the treatment are justified by the potential benefits. These considerations are the final questions to be answered in the critical appraisal of a systematic review.

Systematic reviews are primarily intended to be of use to clinicians and their patients, but they have implications for researchers and journal peer reviewers, as well. The study quality criteria that authors of systematic reviews apply to the available literature determine, in effect, which studies will be added to the accepted body of knowledge about the solutions to a clinical problem. Researchers who wish to contribute to that body of knowledge need to be aware of and address the identified systematic review concerns of rigorous randomization, concealed treatment allocation and subject retention in their study designs. Both researchers and peer reviewers can work to reduce the impact of reporting bias on the synthesis of knowledge. High quality treatment studies with no statistically significant outcomes tell us what does not work and deserve publication.

Production of a rigorous systematic review is a significant scholarly achievement with direct benefits for clinical practice. Clinical expertise is reflected in the choice of clinical problem, the specification of therapies and outcomes to be included and the decisions about which studies can be directly compared with each other. Research expertise is reflected in the comprehensive search for studies, the appraisal of individual study quality, the decisions made about synthesizing evidence from studies with varying quality levels and the methods used to control bias during the synthesis process. A systematic review is an important step in moving evidence from discovery to application.

References