Clinician’s Guide to Systematic Reviews and Meta-Analyses

David A. Feldstein, MD

ABSTRACT
Systematic reviews answer clinical questions by finding and evaluating all available evidence. The systematic review is a powerful tool to help clinicians use evidence for patient care decisions. There are many sources for high-quality systematic reviews. Like all scientific studies there are potential biases, but systematic reviews have many benefits over narrative reviews. To ensure appropriate use of systematic reviews, clinicians must evaluate them in a logical, step-by-step manner. This article will review the benefits of systematic reviews, how to locate them, and how to evaluate their quality and results.

SYSTEMATIC REVIEWS DEFINED
Clinical research evidence is rapidly expanding and clinicians may find multiple trials addressing the same clinical question. One method to summarize the large numbers of clinical trials is the systematic review—a critical evaluation of research that attempts to answer focused questions. An effort is made to evaluate all available research and reduce bias. Systematic reviews are scientific investigations and have pre-planned strategies to include all relevant articles, appraise primary trials, and synthesize data. Meta-analyses are systematic reviews that use statistical methods to combine primary trials and produce a single pooled estimate of an intervention’s effect.

Preventing implementation delays
Systematic reviews can prevent unnecessary delays in the implementation of new therapies; these delays can occur while waiting for large randomized controlled trials to be performed. The Cochrane Collaboration Logo (Figure 1) shows a striking example of this delay. It is a diagrammatic representation of systematic review data from 7 randomized controlled trials performed from 1972 to 1982, evaluating corticosteroids given to women expected to give birth prematurely. The horizontal lines are individual trials; the length of the lines represents confidence intervals. The diamond at the bottom is the pooled result of these 7 trials. The vertical line represents the point at which the 2 treatments are the same. A result on the left side of the vertical line shows that the treatment is beneficial; a result on the right side of the line shows that the treatment is harmful. By 1982 these studies showed a clear benefit with a 30%-50% reduction in the odds of the death of the baby in women given corticosteroids. Experts at the time should have been recommending routine use of corticosteroids for women in premature labor. Unfortunately they did not recognize the strength of the evidence supporting the use of corticosteroids. The first systematic review was not performed until 1989 and by 1991, 7 more trials were done. The delay in adopting this clearly beneficial therapy cost many premature babies their lives.

A similar example was seen in the delay of the adoption of thrombolytic therapy for acute myocardial infarction. It took 5 years and multiple systematic reviews from the time that meta-analyses clearly showed a reduction in mortality until the majority of narrative reviews recommended routine use of thrombolytics.
POTENTIAL PROBLEMS WITH SYSTEMATIC REVIEWS

Systematic reviews are scientific investigations, so when they are improperly performed their results can be biased and misleading. The act of combining many individual studies makes systematic reviews vulnerable to new biases not seen in clinical trials. Altering the studies that are included in a systematic review can dramatically change its results. Using different methods of statistically combining individual trials can also cause varying results. These differences can lead to conflicting conclusions among systematic reviews of the same question.

A systematic review is only as good as its individual studies. Combining multiple poorly-performed and biased studies will compound the bias and give misleading results. If there is very little quality evidence for a clinical question the systematic review will not be able to adequately answer the question. What is needed is a systematic way of looking for these potential problems; this will be discussed later in this article.

CLINICAL CASE

A 70-year-old man comes to the office with a complaint of being unsteady on his feet. A thorough evaluation reveals no cause for his imbalance. The patient lives alone and there are concerns about a potential fall. After counseling about environmental changes to decrease his risk of falling, the possibility of using vitamin D to help prevent falls is considered. The question the physician creates for the search is: “In elderly patients at risk for falls does vitamin D supplementation decrease the risk of having a fall?”

Finding the Evidence

Having established the clinical question, the important evidence resources specific to systematic reviews are the following:

- Cochrane Database of Systematic Reviews is a collection of full text systematic reviews produced by the Cochrane Collaboration, an international group committed to preparing, updating, and disseminating systematic reviews of health care interventions. The systematic reviews are structured and use explicit criteria to minimize bias. They are peer-reviewed throughout the process to ensure the highest quality and are updated at regular intervals. Cochrane reviews have been shown to be of higher methodologic quality than other systematic reviews found in MEDLINE.3

- Database of Abstracts of Reviews of Effects (DARE) is funded by the NHS in the United Kingdom. DARE contains structured summaries of published systematic reviews that have met strict quality criteria. Each summary also provides a critical commentary on the quality of the original review.4 Cochrane and DARE contain high-quality evidence, but both are relatively small databases, so they may not contain answers to many clinical questions.

- PubMed MEDLINE is a logical next step if no evidence is found in Cochrane or DARE. The “Clinical Queries” page can be accessed from a link on the left side menu of PubMed. This feature provides filters for limiting searches including one for systematic reviews.

A search of PubMed Clinical Queries using the terms “vitamin D AND falls” returns a recent systematic review by Bischoff-Ferrari that appears to answer the clinical question.5

Evaluating the Evidence

Once a systematic review is found it must be evaluated to decide if it can be used to treat the patient. There are 3 steps to critically evaluating evidence: (1) Assess the validity of the study (is it biased?),(2) Understand the results,(3) Assess the applicability of the evidence to the patient. A commonly used approach to evaluating systematic reviews has been published in the Journal of the American Medical Association (JAMA) User’s Guides to the Medical Literature series and more recently in book form.6

If evidence is available in 1 of the 2 pre-appraised resources (Cochrane and DARE) then there is no need to assess the validity of the study. Otherwise systematic reviews, like any scientific investigation, are open to bias and must be appraised. The systematic review by Bischoff-Ferrari will be used as an example of the areas to evaluate.
Assessing Study Validity

There are 4 questions that need to be evaluated to determine the validity of a systematic review.

1. Did the review explicitly address a sensible clinical question?
The authors must clearly state the question they are trying to answer as well as the inclusion and exclusion criteria for individual studies. This implies that the authors are including studies in an unbiased manner. It also allows the reader to determine if the intervention will have a similar effect across the range of included patients, interventions, and outcomes. Combining studies that have very different patients, interventions, and outcomes can cause a misleading or nonsensical result.

   The goal of the review by Bischoff-Ferrari is to determine the overall efficacy of vitamin D in preventing falls among older individuals. They looked at any type of vitamin D in patients 60 years or older, and the study defined falls and how they were ascertained. The authors excluded studies of patients with unstable health conditions, including post-hospitalization. This appears to be a sensible question, but the intervention may be considered somewhat broad since they included studies that used calcium plus vitamin D.

2. Was the search for relevant studies detailed and exhaustive?
By selecting only certain trials to include in a systematic review the authors can introduce a selection bias. To help prevent this, authors must try to find all available trials that might meet their inclusion criteria. The basic search should include databases such as MEDLINE, EMBASE (the Excerpta Medica database), Cumulative Index to Nursing and Allied Health Literature (CINHAL) and the Cochrane Controlled Trials Register. In addition, authors should also search for unpublished data by contacting experts in the field, and obtaining abstracts from recent scientific meetings and pharmaceutical manufacturers if indicated. Unpublished data is important since “negative” studies, or those where the treatment shows no significant benefit, are less likely to be published. Excluding unpublished data may overestimate the treatment effect by more than 10%.

   Bischoff-Ferrari searched MEDLINE, Cochrane Controlled Trials Register and EMBASE. They also contacted experts, and searched reference lists and abstracts presented at the American Society for Bone and Mineral Research. There were no limitations based on language. This is a thorough search strategy.

3. Were the primary studies of high methodologic quality?
A meta-analysis is only as good as the individual trials it includes. Lower quality studies may overestimate treatment effects by more than 30%. It is extremely important that the authors clearly report the quality of the individual trials. The criteria for evaluation of therapeutic trials should be similar to those demonstrated in the User’s Guides to the Medical Literature, with the most important criterion being proper randomization. The authors can use a quality rating scale to give each study a numerical quality score or just assess individual quality criteria.

   Bischoff-Ferrari’s study assessed individual trials for randomization, allocation concealment, blinding, and withdrawals. They only included randomized controlled trials and report the quality of the individual trials in tabular form. They do not use an overall quality rating scale. All of the studies are randomized and double blind. Two of the studies do not state whether there was allocation concealment and 1 study had a dropout rate greater than 20%. Overall the included studies were of good methodologic quality.

4. Were the results similar from study to study?
The first validity question determines whether the authors had a question narrow enough to expect a similar effect across the included studies. Once the included studies are combined, it must be determined if the studies’ results were actually similar. The statistical difference between the results of individual studies is referred to as heterogeneity. This is a somewhat confusing concept, since meta-analyses are often thought of as a way of combining multiple studies with different results to get the “real” result. The problem lies in combining studies that may be measuring very different things. If the same trial was repeated 10 times in the same patients a slightly different result would occur each time due to chance or random error. But if the results of the studies are heterogeneous, then the differences between the results vary due to more than chance alone. As mentioned earlier, combing studies measuring different parameters could cause a misleading result. If there is heterogeneity between the studies the authors should try to determine why it is occurring. The most common reasons are differences in patients, interventions, or outcome measurement, as well as the quality of the individual trials. Tests for heterogeneity can be performed visually looking at the forest plot of the results or with formal statistical testing. The most common test used for heterogeneity is the Q statistic.

   Bischoff-Ferrari report that they performed formal heterogeneity testing and did not find any significant heterogeneity.
RESULTS
Results of systematic reviews and meta-analyses are presented differently from those of individual trials of therapy. The pooled or combined estimate is usually given as an odds ratio or relative risk owing to the statistical calculation necessary for pooling studies. There is also usually a graphical representation of the results of each study and the combined result. This is in the form of a forest plot. Figure 2 is an adaptation of the forest plot from the Bischoff-Ferrari study. The individual studies are listed down the left side of the figure, often in chronological order. The result or point estimate of each study is represented by a box that's size is proportional to the weight of the study in the pooled result; the larger the box, the greater its weight. The horizontal lines on each side of the box represent the confidence intervals of the study. This is plotted on the x-axis that, in this example, is odds ratios on a logarithmic scale. An odds ratio of 1.0 means that there is no difference between the experimental and control groups. If the confidence intervals touch 1.0 then the results are not statistically significantly different. Finally, the pooled result is reported as a diamond on the bottom line of the forest plot. The center of the diamond represents the pooled result and the ends of the diamond are the confidence intervals. The forest plot allows the reader to see trends in the results of individual studies at a glance, as well as grossly evaluate for heterogeneity.

Examining the forest plot in Figure 2 shows us that there were 5 individual studies that made up the meta-analysis. Only 1 of these studies showed a statistically significant decrease in falls with vitamin D (the Gallagher 2001 study). The pooled result showed a statistically significant decrease in falls with an odds ratio of 0.69 (95% confidence intervals 0.53 to 0.88). This means that there was an approximately 31% decrease in the risk of a fall in patients taking vitamin D.

APPLICATION
When a systematic review is determined to be valid and the results of the meta-analysis have been evaluated, it is time to decide whether this evidence can be used to treat an individual patient. Systematic reviews pool results across many studies with slightly different populations so the reader can be more confident that the results can be generalized to their individual patient. Once you have determined that the evidence from the meta-analysis can be used to care for your patient, you need to decide what the potential benefits and harms of the treatment are and to incorporate your patient’s values about these benefits and harms.

The patient in the scenario is 70, similar in age to the patients in the individual studies. He is otherwise relatively healthy and the authors excluded studies of patients with unstable health conditions. Therefore, it appears reasonable to use the evidence in caring for this patient. Unfortunately the systematic review does not make any mention of side effects of the vitamin D supplementation, but the patient is quite concerned about falls and hip fractures and is willing to take vitamin D given the fairly large decrease in the risk of falls regardless of side effects.
CONCLUSION
Systematic reviews can summarize all of the available evidence for a clinical question. If appropriate, a meta-analysis can be performed to provide an estimate of the overall treatment effect for a given therapy. This is a tremendous timesaver and allows busy clinicians to incorporate evidence from trials that they may not have found otherwise. The combination of multiple studies also provides more powerful evidence for making clinical decisions than an individual study. Systematic reviews are scientific investigations and have many potential biases, so it is important to systematically approach their evaluation. Their results are usually reported graphically, which allows the reader to get a quick look at the intervention effect across all of the available evidence. Systematic reviews are becoming more common and will continue to play a major role in translating research evidence into patient care decisions.

REFERENCES
The mission of the Wisconsin Medical Journal is to provide a vehicle for professional communication and continuing education of Wisconsin physicians.

The Wisconsin Medical Journal (ISSN 1098-1861) is the official publication of the Wisconsin Medical Society and is devoted to the interests of the medical profession and health care in Wisconsin. The managing editor is responsible for overseeing the production, business operation and contents of Wisconsin Medical Journal. The editorial board, chaired by the medical editor, solicits and peer reviews all scientific articles; it does not screen public health, socioeconomic or organizational articles. Although letters to the editor are reviewed by the medical editor, all signed expressions of opinion belong to the author(s) for which neither the Wisconsin Medical Journal nor the Society take responsibility. The Wisconsin Medical Journal is indexed in Index Medicus, Hospital Literature Index and Cambridge Scientific Abstracts.

For reprints of this article, contact the Wisconsin Medical Journal at 866.442.3800 or e-mail wmj@wismed.org.

© 2005 Wisconsin Medical Society