

The role of systematic reviews in evidence-based practice

Knowing what questions to ask and information to include is critical

By Charles Turkelson, PhD, and Jill Elaine Hughes, MA

Evidence-based practice (EBP) has been around for nearly 25 years, but the questions about what constitutes EBP, particularly good EBP, are still being asked. Good evidence-based medicine is not simply selecting an article to prove a preconceived opinion or assigning levels of evidence to a journal article.

Good evidence-based medicine is a transparent system for ensuring that evidence is objectively obtained, appraised and analyzed, and then systematically applied to practice decisions.

Systematic review

The first step in EBP is the systematic review. Systematic reviews are stand-alone documents that address the effectiveness of a single drug, device or procedure. They underpin many clinical practice guidelines such as the ones being prepared by the AAOS for guidelines and performance measure development projects. A systematic review must use an evidence review protocol that fosters objectivity, which is the foundation of modern science. Furthermore, systematic reviews must be conducted using the scientific method.

According to Frank Wolfs, PhD, a physics professor at the University of Rochester, “The scientific method attempts to minimize the influence of . . . bias on the outcome of an experiment.”¹

Preventing bias and promoting objectivity

A common mistake, according to Dr. Wolfs, “is to ignore or rule out data which do not support the hypothesis. Ideally, the experimenter is open to the possibility that the hypothesis is correct or incorrect. Sometimes, however, a scientist may have a strong belief that the hypothesis is true (or false), or feels internal or external pressure to get a specific result. In that case, there may be a psychological tendency to find ‘something wrong’ . . . with data which do not support the scientist’s expectations, while data which do agree with those expectations may not be checked as carefully. The lesson is that all data must be handled in the same way.”¹

Objective data handling is fundamental to any systematic review. Systematic reviews begin by specifying hypotheses to test in the form of key questions. An evidence-based key question must be possible to answer using the results of a scientifically conducted experiment. The experiment need not have already been conducted; after all, science is built on the results of new experiments. Systematic reviews often ask questions that have

not been addressed by a clinical trial or some other experiment to highlight current gaps in our knowledge and provide suggestions for future research.

Framing scientific key questions is not easy, partly because such questions must be very specific. “Can treatment A be used for disease X?” is not a good key question; it is too vague. It is not clear if the question is about whether it is merely possible to use treatment A, whether treatment A is more effective than some or all other treatments for disease X, or whether treatment A is so effective that it should be used in most patients with disease X.

A better phrased question would be “Is treatment A an effective treatment for disease X?” This question forces the reviewer to consider and specify how effectiveness will be determined. Will it be from radiographic data? From measuring patients’ ability to walk? From some biochemical marker? From patients’ self-reports of their improvement after treatment? Some measures of effectiveness are more reliable than others, and only these relatively reliable measures should be considered in a systematic review.

Another example of a less-than-satisfactory key question is “Is treatment A an appropriate treatment for patients with disease X?” This is not a scientific question because “appropriate” is a term that requires subjective judgment. (In the case of EBP, this means clinical judgment.) However, much of medicine is about determining whether a treatment is appropriate, particularly for a specific patient, so questions about whether a treatment is appropriate are often posed in clinical practice guidelines. In preparing evidence-based clinical practice guidelines, clinicians should use only the results of a systematic review to form their judgments about appropriateness, instead of trying to form subjective judgments in advance, and then tailoring studies to support that preconceived result.

Inclusion/exclusion criteria

The next step in determining what kind of information to allow in a systematic review is to establish inclusion/exclusion criteria. For example, advertising, product promotional material, testimonials or anecdotal reports are not considered appropriate scientific information and are almost always excluded from primary literature searches. For many conditions, data from animal studies are excluded because they provide little information about the magnitude of the health benefits (pain reduction or improvement in functional status) experienced by people. Sometimes, the protocol requires inclusion of only relatively recent studies, so that studies conducted prior to a certain date, or using outmoded methods or technologies, are excluded. Meeting abstracts are also often excluded, partly because they usually don’t contain enough information to make good decisions about the quality of a study. Often the search criteria may include only the best available evidence (such as Level II evidence, when it exists, instead of Level III evidence). (For a description of what constitutes Levels of Evidence, see the [April 2005 Bulletin](#).)

Once the reviewer sets search parameters, the list of allowable data is formalized into a specific inclusion/exclusion criteria set within the search program(s) (PubMed, Embase, Ovid) being used to retrieve articles. These criteria must be transparent and auditable, which provides reviewers with an incentive to be objective when creating them. Creating the rules in advance also helps prevent bias.

Objectivity and bias in clinical study design

Searches for information—specifically for articles that describe the design, conduct and results of scientific studies—should also be conducted in an objective way. In general, systematic reviews are not confined to the National Library of Medicine’s PubMed/Medline, because using a single database may miss too many studies.²⁻³ Extensive searching is part of an overall effort to minimize bias, and helps prevent reviewers from primarily considering studies that support what they already believe while ignoring the studies with results that run counter to their beliefs.

Having identified and obtained articles that meet the required criteria, the reviewer then appraises their quality. Quality appraisal is a crucial, necessary part of evidence-based medicine, because, in general, the quality of clinical studies is often less than optimal.

There are several approaches to appraising study quality, none of which is universally accepted. What is accepted is that any method used should be transparent and auditable.

The AAOS appraises study quality using a “Levels of Evidence” approach. Higher levels of evidence (Levels I or II) provide more reliable evidence than lower levels (III or IV). Levels of evidence are typically used in the preparation of clinical practice guidelines, and their prevalence has perhaps led some to view them as the predominant—if not the sole—component of EBP. Therefore, it is worth repeating that EBP, particularly as applied to systematic reviews, is a system, with levels of evidence as one component of that system.

Another common misconception is that EBP does not consider expert opinion. In fact, EBP will consider expert opinion when no relevant experimental data are available (when expert opinion is the “best available evidence”). Expert opinion is commonly used in evidence-based clinical practice guidelines on topics where little clinical research exists.

Data analysis

Once the study quality has been appraised, the available data must be analyzed. Sometimes, these analyses are statistical (meta-) analyses geared towards determining how well a given drug, device or procedure works. When the data do not permit meta-analysis, a qualitative analysis is conducted. Qualitative analyses can sometimes yield an approximate answer to “How well does it work?” but otherwise may be limited to determining whether “it works.”

The data used in the analysis are provided in evidence tables, so that readers can also examine the data used in the analysis. Data are not limited to a study's results, but include information about the patients enrolled, the design and the conduct of the study. To avoid being influenced by any biases an article's authors might have, many systematic reviewers read only the "Methods" and "Results" sections of a published article, and do not consider information in the article's "Introduction" or "Conclusions" sections.

Before turning the answers to key questions of a systematic review into clinical practice guideline recommendations, an additional step is needed. Inevitably, different recommendations are based on studies of differing quality, differing amounts of data (both in the number of studies and the number of patients) and results that may or may not be consistent. These factors influence one's confidence in both the data and the recommendations. Recommendations based on large amounts of consistent, high-quality data are more reliable than recommendations that are based on small amounts of inconsistent, low-quality data.

Sometimes, a recommendation is not warranted. Authors of clinical practice guidelines often express their confidence with a grade of recommendation. The AAOS has formally adopted a grades of recommendation system (see Figure A on pg. 14). Unlike a level of evidence, which is applied to an individual study, a grade of recommendation is used to characterize the entire body of literature addressing a question. Applying transparent grading criteria to a systematic review is another way to promote objectivity as part of EBP.

Charles Turkelson, PhD, is director of the AAOS research and scientific affairs department and can be reached at turkelson@aaos.org. Jill Elaine Hughes, MA, is a clinical quality improvement coordinator who can be reached at hughes@aaos.org.

References

1. Wolfs, F. Introduction to the Scientific Method. Physics 113 Classroom and Lab Materials [online textbook publication]. University of Rochester, 1996; Appendix E.
2. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidem* (2003) 56:943-955.
3. Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying Clinical Trials in the Medical Literature with Electronic Databases: MEDLINE Alone is Not Enough. *Cont Clin Trials* (2000) 21:476-487.

Figure A

AAOS Grades of Recommendation

Each guideline recommendation is graded using the following system:

A — Good evidence (Level I studies with consistent findings) for or against recommending intervention.

B — Fair evidence (Level II or III studies with consistent findings) for or against recommending intervention.

C — Poor quality evidence (Level IV or V) for or against recommending intervention.

I — Insufficient or conflicting evidence not allowing a recommendation for or against intervention.