

Evidence-Based Medicine, Part 2. An Introduction to Critical Appraisal of Articles on Therapy

Roberto Cardarelli, DO, MPH
Richard F. Virgilio, DO
Lockwood Taylor, MPH

This article provides an introductory step-by-step process to appraise a therapeutic article. The authors introduce these principles using a systematic approach and case-based format. The process of assessing the validity of a therapeutic article, determining its importance, and applying it to an individual patient is reviewed. The concepts of randomization, blinding, and concealment are discussed to help physicians determine an article's validity. Instruction on calculating relative risk reduction, absolute risk reduction, and number needed to treat is provided and applied to the clinical scenario. Finally, information that is learned from the previous two steps is applied to patient care. The skills learned from appraising a therapeutic article in the manner outlined provides a basis for life-long learning and improved patient care.

[Editor's note: This article is part 2 of a six-article series intended to introduce the principles of evidence-based medicine (EBM) to busy clinicians, physician residents, and medical students. Because the application of EBM is a career-long process, further training is needed beyond the information provided within this article and series. A foundation of knowledge about research methods is critical in understanding EBM; however, such details, though introduced, are beyond the scope of this series.]

J Am Osteopath Assoc. 2007;107:299-303

Physicians face numerous clinical decisions at the point-of-care. New medical treatments and technological innovations have made practicing medicine exciting and more challenging than ever. During clinical visits, patients are active participants in their healthcare and regularly inquire about new therapies and diagnostic tests. It is imperative for physicians to locate, interpret, and apply new research

quickly. Evidence-based medicine (EBM) is the practice of assessing the medical literature in a time-efficient manner to answer a clinical question about, and on behalf of, one's patients.¹

In this article, we introduce a strategy for busy physicians, physician residents, and medical students to critically assess the medical literature on therapy. In-depth details of research methods are beyond the scope of this introductory series on EBM. Readers are encouraged to seek further training on these topics with supplemental learning opportunities and continuing medical education. Finally, the clinical scenario described has been simplified to provide readers with an illustrative example for the general concepts introduced.

Levels of Evidence

Although this article reviews how practitioners can critically appraise an individual article on therapy, strong and valid systematic reviews and meta-analytic studies are preferable to therapeutic reports in the clinical decision-making process. Systematic reviews and meta-analyses collectively summarize similar articles on therapy for a common medical problem to provide conclusions or recommendations. However, a collective summary is only as good as each individual article that is used. Prior to using treatments recommended in such articles, physicians are encouraged first to determine if the systematic review uses an EBM approach before presenting conclusions.

Although describing systematic reviews and guidelines are beyond the scope of this article, they are addressed in the first article in this series, "Evidence-based medicine, part 1. An introduction to creating an answerable question and searching the evidence" by Richard F. Virgilio, DO; Ana Luz Chiapa, MS; and Elizabeth A. Palmarozzi, DO.² The strongest study design for therapeutic interventional investigations is randomized controlled trials (RCTs), which are discussed briefly below. It is important to have an understanding of the various types of study designs, when each is most appropriate, and the various strengths and weaknesses of each model. Good sources for further study include the Web sites for Oxford's Centre for Evidence-Based Medicine (<http://www.cebm.net/index.aspx?o=1039>) and the National Cancer Institute at the US National Institutes of Health (<http://www.cancer.gov/cancertopics/pdq/levels-evidence-cam/HealthProfessional/page2>).

From the Department of Family Medicine (Drs Cardarelli and Virgilio) at the University of North Texas Health Science Center—Texas College of Osteopathic Medicine in Fort Worth and the School of Public Health (Dr Taylor) at the University of Texas Health Science Center at Houston.

Address correspondence to Roberto Cardarelli, DO, MPH, Department of Family Medicine, Texas College of Osteopathic Medicine, University of North Texas Health Science Center, 855 Montgomery St, Patient Care Center, 2nd Fl, Fort Worth, TX 76107-2553.

E-mail: rcardare@hsc.unt.edu

Submitted February 1, 2007; revision received June 14, 2007; accepted June 18, 2007.

SPECIAL COMMUNICATION

Validity of Articles on Therapy

To assess the validity of a study is to ask if its findings are true and accurate. This is a crucial step in the validation process because physicians must determine whether the article outcomes were influenced by known or unknown sources of bias (Figure 1). This task can be accomplished by answering a set of questions:

■ Were subjects randomly assigned to the treatment group?

Randomization is a pivotal step in determining the validity of a study.³ It ensures that each subject has the same probability of being selected for active treatment protocols rather than for a control treatment or placebo. It also allows study results to be generalized to a larger population of interest. Randomization can be as simple as “flipping a coin” or using a random number generator.

The strengths and weaknesses of each article must be evaluated independently. If clinical practice recommendations based on RCTs are not available, physicians may choose to look at the results of nonrandomized studies. Causation (“X causes Y”) cannot be established using observational studies (eg, case control study, cohort study), however. In most instances, only an association between a therapeutic intervention and desired outcome can be interpreted from the results of observational studies.⁴

■ Were all the subjects accounted for and attributed at the end of the study?

All enrolled subjects must be accounted for at the end of the investigation. A large “lost to follow-up” group may lead researchers to present biased study results. For example, if very ill subjects do not complete a study and are not ultimately accounted for, the study’s outcome may appear favorable when a more thorough analysis of the data gathered may have led researchers to very different conclusions.

All study participants should be analyzed in the groups to which they were originally assigned. This principle is called “intention-to-treat.”⁵ This research model allows group randomization to be preserved, and the known and unknown factors affecting patient prognosis have an equal probability of impacting subjects assigned to each study group. Researchers who conduct their analysis excluding very ill subjects would obtain results that suggest the therapy was efficacious—though those study results would be compromised through bias.

■ Was the study “blinded”?

Study participants, clinicians, and investigators should not know which subjects are assigned to the intervention or control group. Participants who are aware of their group assignments may bias study results by behaving and/or responding differently. Clinicians who are aware of subject assignments may treat individuals in the intervention group differently than those in the control group, unconsciously (or consciously) manipulating the study design or analysis.

Clinical Scenario

A 47-year-old man visits your clinic for a routine physical examination. He appears to be in excellent physical health, and his total cholesterol is in the normal range. He tells you his father died of a myocardial infarction at the age of 63 years and his mother died of a stroke at 74 years. He asks about a treatment with a hydroxymethylglutaryl-CoA reductase inhibitor (statin) he saw advertised on television. He is interested in knowing more about the use of statins for heart-attack prevention.

Normally, you would suggest a modification in diet, encourage him to maintain regular exercise, and check his lipid levels yearly. However, you remember hearing about a study that investigated the primary prevention of coronary heart disease with statins, but you are unsure of its results. You let him know that you would like to do some research before making a recommendation. You promise that you will have an answer for him at his follow-up visit in 4 weeks.

When you return to your office, you type the following text into PubMed:

statins in primary prevention of coronary disease
and event reduction

You restrict the search to studies in English that are randomized controlled trials. Your search yields 6 items. You select and review the study that seems to be most appropriate to your clinical question.

(continued)

Figure 1. Clinical scenario.

Double-blinding refers to the processes of keeping group assignments concealed from study subjects and investigators. Sometimes, however, a double-blinded research protocol is simply not possible given the intervention used. For example, a study that investigates the efficacy of surgical vs nonsurgical procedures is unable to conceal group assignments from the subjects or surgeons. When possible, however, all measures of progress and improvement for such studies should be concealed from primary investigators through the use of independent evaluators.

■ Were the study groups similar at the start of the investigation?

The demographics and description of study participants are usually found in the first table of an article. There are always some established risk factors that may affect the study outcome. Therefore, it is important to determine whether these factors are equally balanced between the intervention and control groups. If one wanted to determine the overall benefit of a surgical procedure, it would be important to know if other comorbid conditions are balanced between the two groups (eg, coronary heart disease, diabetes mellitus). Randomization does not always guarantee an equal balance of demographic factors and medical history between groups. If the difference of a variable (eg, age) is large, it may bias study results. It is important for clinical investigations to have a sufficient number of subjects (sample size) so that the results would be able to find a desired difference in the outcome (sufficient power).⁶ Small studies have a greater probability of having an unequal distribution of baseline subject characteristics. At times, however, a deficiency of this kind may be overcome using appropriate statistical tools (eg, regression analyses).

■ Were the study groups treated equally?

As one might imagine, if subjects in the intervention and control groups were treated differently, maintaining blinding protocols would be difficult. For example, if subjects in the intervention group had more frequent follow-up visits than other study participants, they may inadvertently be “unmasked” to group assignments—suddenly aware that they are receiving the intervention. Data gathered from a study in which the intervention group receives more frequent follow-up visits may also skew data by virtue of the fact that subjects in such a group would have more opportunities to report adverse events (ie, significant between-groups differences). It is important that study groups are treated equally in all aspects of the study (Figure 2).

■ How big was the treatment effect?

Once it has been determined that a particular article on therapy is valid, the physician should evaluate the magnitude of the treatment effect and its precision. Only basic mathematical and statistical skills are required for this kind of postpublication review and analysis of the medical literature.

Most journal articles report outcomes in a dichotomous fashion. For example, one may choose to evaluate whether or not daily use of aspirin prolongs life 2 years after an initial myocardial infarction. Therefore, a researcher might then compare an event of interest (eg, mortality) among those who received aspirin and those who received nothing (or placebo). The proportion of those who died in the placebo group determine what is called the *control event rate* (CER). The CER is considered the baseline risk for patients who meet study inclusion and exclusion criteria. The proportion of study subjects in the intervention group who died determines the *experimental event rate* (EER). The truncated table shown in Figure 3 provides a

hypothetical example of just one possible study outcome (ie, mortality ≤ 2 y postinfarction).

Although the difference as reported in the table between the EER and CER may appear to be statistically significant, the data presented does not provide us with any clinically useful information.

Numerical terms that can be applied to our patients and allow us to explain potential outcomes are needed. It is for this reason that many articles report the relative risk reduction (RRR). When the EER is subtracted from the CER and that total is then divided by the CER, the result is the RRR.⁷ To elaborate using the example provided by the data in the table (Figure 3), the $RRR = [0.15 - 0.05] / 0.15 = 0.67$ or 67%. A physician reading this number in a medical journal can then safely say that aspirin, relative to placebo, decreased patients' risk of death in the 2 years after an initial myocardial infarction.

Although this finding appears to be impressive since it confers a large treatment effect, it still conveys incomplete information to the reader because it does not attempt to evaluate patients' baseline risk (ie, CER) of death during the 24 months postinfarction. One cannot discriminate large treatment effects from small ones. Therefore, for example, with a postinfarction CER of 0.00015% and an EER of 0.00005%, the RRR will still be 67%. Because the baseline risk (ie, CER) is small, a further decrease in risk will have only minimal clinical impact. It is for this reason that the RRR is not the best calculation to use in clinical practice.

The absolute risk reduction (ARR), which is calculated by subtracting the EER from the CER, takes the baseline risk into account.⁷ Using the same example (Figure 3), the ARR would be calculated by subtracting the 0.05 EER from the 0.15 CER for an ARR of 0.10 or 10%. This analysis would lead physicians practicing EBM to reach a very different conclusion from physicians who consider only the RRR of 67%.

At times, it is difficult to recall an ARR value. In addition, physicians want to convey the information available to their patients in a manner that is easy for them to understand.

Clinical Scenario (continued)

You review the article and find it to be a well-designed study. The study was randomized, blinded, and all participants were accounted for and analyzed in the groups to which they were initially assigned. In addition, all participants were treated the same throughout the study. Therefore, you determine the study to be valid.

(continued)

Figure 2. Clinical scenario (continued).

SPECIAL COMMUNICATION

Table Aspirin Use in Postinfarction Subjects (N=200)		
	Treatment Group (Aspirin, 81 g/d), n=100	Control Group (Placebo), n=100
■ Subject Health Status at 2-year Follow-up		
□ Died	5 (5)	15 (15)
□ Hypertensive	9 (9)	12 (12)

Figure 3. Example of poor presentation of efficacy data from a study investigating aspirin use in postinfarction subjects. The calculation for the experimental event rate is as follows: $5/100=0.05$ or 5%. The calculation for the control event rate is as follows: $15/100=0.15$ or 15%.

For these purposes, the number needed to treat (NNT) can be calculated. The NNT is computed by taking the reciprocal of the AAR, or dividing 1 by the ARR.¹ To calculate the NNT for the study reported in the table (Figure 3), one would divide 1 by 0.10 for an NNT total of 10. In other words, 10 patients need to take 81 g of aspirin daily for 2 years postinfarction to prevent one mortality.

Physicians must then decide whether an NNT of 10 is clinically significant or remarkable. This determination can be made by comparing the number to other NNTs for interventions with a similar therapy duration. The disease itself and the severity of the outcome must also be taken into consideration. For example, one may be willing to administer a particular therapy when an outcome is severe (eg, NNT of 50 at 1 year of treatment for cancer). Yet, with the same NNT, one may be reluctant to prescribe an antibiotic to manage a mild upper respiratory infection when it is known that the medication would shorten the symptomatic phase of the illness by only 1 or 2 days (Figure 4).

■ How precise is the estimate of the treatment effect?

When a result is calculated (CER, EER, RRR, ARR, and NNT), it represents an estimate of some theoretical true value. Ideally, the calculated result should be close to this true value as much as possible. A range of values is used to estimate where the true measure would lie. Normally, this range of values is expressed by a 95% confidence interval (CI) and can be interpreted as: "We are 95% confident that the true value lies within the given interval."⁷ The narrower the 95% CI, the more precise the result is considered to be. Although the *P* value is a statistical expression of significance (eg, $P<.05$), it does not provide any information on the magnitude of the effect or precision of the results. Therefore, the 95% CI is the most useful mode to express the precision of a treatment effect.

Practical Use

Determining the practical application of study results is an important step that is frequently overlooked by authors and not considered during postpublication reader assessments—or during postpublication peer review. Typically, as noted, the first table in an article outlines the study population's demographic data, allowing readers to determine quickly whether the researchers' findings can be applied to any given "real world" patient.

If the study is applicable to patient care, the overall potential benefit for the patient must be assessed. Every treatment has its risks (ie, adverse effects) and benefits. More importantly, the patient's feelings and perceptions must be taken into account. For example, taking a pill every day for the rest of one's life may be perceived by some patients as a greater risk than not taking the medication at all. Such patients may "refuse" treatment before the prescription pad is even out of his or her physician's pocket. Questions that physicians practicing EBM would ask to determine the practical use of study results might include the following:

□ Can I apply these results to my patient?

Physicians practicing EBM can answer this question by determining if his or her patient meets inclusion and exclusion criteria of the study and if the patient has any comorbidities that would bias or contraindicate the desired outcomes of the investigation. One should be able to generalize the results of a study to a particular patient most times—but

Clinical Scenario (continued)

You decide to use the primary endpoints of the study as your outcome of interest. This includes the incidence of first acute major coronary events, defined as fatal or nonfatal myocardial infarction, unstable angina, or sudden cardiac death. According to the truncated data presented in the authors' table, the control event rate for mortalities can be calculated at 0.055, or 5.5%; the experimental event rate, 0.035, or 3.5%. These numbers result in a relative risk reduction of 36.4% and an absolute risk reduction of 2%. The number needed to treat is 50, meaning 50 individuals must be treated with the statin over 5.2 years to prevent one primary endpoint.

(continued)

Figure 4. Clinical scenario (continued).

this should never be taken for granted. The key is not to be overly stringent with study inclusion or exclusion criteria, but rather, to determine if there are any compelling reasons that one should *not* generalize the results to the patient under consideration.

- *Were “disease-oriented” or “patient-oriented” outcomes considered?*
It is important to assess the types of outcomes studied. Many times researchers strive to improve markers of disease (eg, systolic blood pressure, pulse rate) rather than patient-oriented outcomes (eg, mortality rates, pain resolution). In our example with aspirin therapy, the reduction in deaths is an important clinical outcome. If the outcome was related only to platelet counts (a disease-oriented outcome), researchers would not have informed their readers about any significant clinical outcomes.
- *What are the benefits versus costs of this treatment?*
Patients may procrastinate in starting a new treatment due to concerns about potential adverse events and financial costs. Deleterious effects of therapies must be considered in every clinical decision, and they should be addressed directly and collaboratively with patients. For example, although taking aspirin may decrease overall mortality rates postinfarction, the potential for an increase in gastrointestinal or cerebral bleeding rates must also be considered, especially in conjunction with patient comorbidities.

The decision to initiate therapy must be reached collaboratively between the patient and physician. Ideally, the decision is based on an individualized treatment plan suggested by a physician practicing EBM (*Figure 5*).

Physicians attempting to practice EBM may find, on occasion, that they will need to adjust their expectations for patient acceptance of the proposed treatment plans. For example, a young athletic man with no past medical care is probably more likely to be interested in lifestyle modifications for elevated blood pressure when compared with a 65-year-old woman who had a previous heart attack and is a current smoker. And yet, physicians may also discover that some patients who initially appear to reject proposed treatment plans are simply functioning with alternative internal timelines and need only to hear the same recommendations repeated regularly over time before they decide to commit to a treatment plan.

Conclusion

Although most clinicians are already incorporating EBM principles in their practices, often instinctively, some physicians may require a more organized approach to integrating this relatively new model of self-education. Improved comfort levels and true expertise in the practice of EBM are the result of additional education, repetition, and self-assessment. The principles of EBM allow physicians to stay informed while also improving the quality of the information communicated to

Clinical Scenario (continued)

The patient returns for his 4-week follow-up visit. You explain the results of your research to the patient and review the adverse events that were found in the study as well.

After further discussion with you, the patient decides that he does not want to take a daily medication. In addition, he had been concurrently investigating the financial ramifications of this potential treatment option on his own. He reports to you that he discovered that his health insurance was not going to cover cost of the medication for preventive purposes.

You review lifestyle options and a thorough behavioral modification plan with your patient to reduce his risk for future cardiovascular events.

Figure 5. *Clinical scenario (continued).*

patients during patient encounters. The systematic approach that is used to appraise an article on therapy is but one step in practicing EBM. Remember, the goal is always to provide the best care possible to patients—using one’s clinical expertise to address patient values and expectations for treatment.

References

1. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB, eds. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. Edinburgh, Scotland: Churchill Livingstone; 2000.
2. Virgilio RF, Chiapa AL, Palmarozzi EA. Evidence-based medicine, part 1. An introduction to creating an answerable question and searching the evidence. *J Am Osteopath Assoc*. 2007;107:295-297. Available at: <http://www.jaoa.org/cgi/content/full/107/8/295>. Accessed August 20, 2007.
3. Guyatt GH, Sackett DL, Cook DJ; Evidence-Based Medicine Working Group. Users’ guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA*. 1993;270:2598-2602.
4. Rothman KJ, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1998:7-28.
5. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319:670-674. Available at: <http://www.bmj.com/cgi/content/full/319/7211/670>. Accessed August 1, 2007.
6. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*. 1994;272:122-124.
7. Guyatt GH, Sackett DL, Cook DJ; Evidence-Based Medicine Working Group. Users’ guides to the medical literature. II. How to use an article about therapy or prevention B. What were the results and will they help me in caring for my patients? *JAMA*. 1994;271:59-64.