Virtual Mentor

MEDICAL EDUCATION
Teaching Critical Appraisal of Medical Evidence

Martha Carvour, MD, PhD

Teaching medical learners how to evaluate evidence for medical decision making represents both a critical and daunting task for curriculum designers. It is, after all, a desirable outcome for any curriculum in evidence-based medicine to produce graduates who not only understand the medical literature, but who can appraise, discuss, and apply what they read in a sound, thoughtful, and efficient manner [1].

When critical appraisal skills remain underdeveloped, students tend to resort to other, less useful approaches. One of these might be called *convenient appraisal*, in which the reader merely accepts, at face value, what the literature claims. Another, equally concerning, approach might be called *cynical appraisal*, in which the reader simply rejects the bulk of the literature, perhaps citing the potential for bias in all research studies. The critical reader, by contrast, takes a balanced approach to the medical literature, seeking to glean the most from each piece, approaching it with reason (not reflex), and supplementing the evidence from one article with data from other sources to build a thoughtful and defensible assessment of a study's content.

Instruction in critical appraisal skills is sometimes deferred in favor of teaching students *what* they read rather than demonstrating *how* to read it. Below are three practical methods for reversing this trend—that is, for actively teaching critical appraisal skills in an evidence-based medicine curriculum.

Teaching Critical Appraisal Skills
1. Start with what makes sense. It is hard to imagine a student who believes that basing the practice of medicine on evidence is *un*important. There are certainly many, however, who question whether incorporating the skills of evidence-based medicine into clinical decision making is practicable and who doubt that they can ever become proficient in these skills. Indeed, understanding of evidence-based medicine concepts among medical learners tends to be lower than desired [2].

Yet, the fact is that many principles of evidence-based medicine—including some of the most complex—mirror the way that many clinicians *already* think about medicine. For instance, the likelihood that a treatment will benefit a patient often depends on the characteristics of the patients in whom it is used. This parallels the evidence-based medicine concepts of "effect modification" or its practice of differentiating between effectiveness and efficacy.

Similarly, in clinical practice, a patient's symptoms may be mistakenly attributed to one diagnosis when the circumstances surrounding the case conceal the genuine cause. This is analogous to the evidence-based medicine concept of confounding. Likewise, individual patients and their experiences are, unsurprisingly, not all the same. This nonuniformity can be linked to discussions in evidence-based medicine about diagnostic validity and precision or about variation and central tendency. For instance, how might a particular diagnostic result be interpreted in a population at high risk for disease compared to a population at lower risk? How should the concept of error in diagnostic tests influence conversations with patients about diagnoses? When evaluating a research study, should clinicians direct more focus to the experience of the entire study population (the distribution or curve representing all patients in the study) or to the average experience (indicated by a mean or similar measure)?

Principles like these have the appeal of being both relevant and intuitive. Instruction in evidence-based medicine should start with what intuitively makes sense (e.g., the general principles of prognostic research) and use this context to explain more difficult concepts (e.g., the critical interpretation of particular prognostic statistics).

2. Don't treat continuous concepts as though they were categorical concepts. Evidence-based medical instruction should not shy away from complexity when complexity makes more sense than simplicity. Although there may not always be a single correct interpretation of a research article so much as several reasonable ones, the temptation for curriculum designers is to create simplicity where it may not be very useful in the long run—to create false dichotomies between good and bad studies, for instance, rather than encouraging learners to critically appraise what they read.

The problem is not that students start simple, but that they learn to think this way. Many students have been trained, intentionally or otherwise, to evaluate study designs, statistical analyses, and research findings reflexively.

Many students, at all levels of medical education, evaluate studies simply on the basis of their design, perhaps by an oversimplified dichotomization of randomized controlled trials and a lesser, nontrial, category, or by identifying the position of a study design on a hierarchical map. These approaches fail to evaluate the content or quality of any individual study and render students unprepared to evaluate important subsections of the medical literature. Critical readers of the evidence should have a basic understanding of the utility, purpose, and value of many designs, including nontrial studies, and be able to engage in meaningful discussions about them [3, 4].

Similarly, students may reflexively evaluate a study on the basis of a single number, such as a p-value, regardless of where that value originated, whether the design or analysis makes any sense, or what the statistical test was meant to assess in the first place. This kind of assessment also falls short of the standard of critical appraisal. Instead, students should learn—from the start—how to interpret the concepts of

evidence-based medicine, including p-values, *within their appropriate contexts*. For example: What was the purpose of the study? Was it appropriately designed to answer the question? What biases, confounding factors, and other considerations influence the interpretation of study findings? Do these factors strengthen or weaken the findings? Then (and only then), should they ask: How should I interpret the p-value in the context of a medical decision?

Notably, many educators have recognized the need to expand upon knee-jerk approaches to the medical literature by creating standardized algorithmic methods for students to follow when they read an article from the literature. This is a step in the right direction. But, here again, a temptation exists to conflate algorithmic assessments of research studies with critical evaluations of their content. These are simply not the same. Completing a checklist or calculating a score satisfies neither the intellectual rigor nor the thoughtful independence of a critical evaluation any more than adding Likert-based scores to a student evaluation offers a constructive view of a student's ability to practice medicine.

3. Adopt a SOAP approach. Admittedly, offering context-based learning opportunities in a concept-heavy curriculum presents a considerable challenge. It may be easier, after all, to define, even abstractly, a p-value or a statistical test and then imagine, also abstractly, that students will eventually learn to think critically about it. Here again, curriculum designers should start with what makes sense—that is, with the way clinician educators already think about medical education.

Consider the way that students are taught to approach the diagnosis of disease. Typically, the process begins with a chief concern—a problem to be solved or a question to be answered. This must then be supported by a series of questions, examination techniques, and laboratory or radiological analyses developed in a thoughtful, systematic, yet patient-specific manner in order to refine a differential diagnosis, identify a working diagnosis, and outline a plan of action together with the patient.

This process is expressly predicated on the notion that a single laboratory value—an international normalized ratio (INR), for instance—derives its practical meaning from the rest of the story that precedes and encompasses it. An INR value—2.5, let's say—is not very useful unless the practitioner knows where it originated and how it may influence decision making. (Is this a therapeutic INR in a patient on warfarin, or is it too elevated to justify nonurgent paracentesis in a patient with cirrhosis?)

The process by which students learn to think about diagnosis may also be applied in evidence-based medicine to promote contextual appraisal of the literature. One possible approach is shown below. While this outline may be used as an example, it does *not* represent an algorithm or formula for critical evaluation. Rather, just as the process of diagnosis must be based on standards of care tailored to individual circumstances, the process of critical evaluation should be founded on standards of

quality adapted to the clinical question of interest and the studies designed to address it.

Subjective
- What is the question to be answered? What information is needed to answer the question? How does the study approach the question? Is the design well matched with the question, and is the measured outcome relevant to it?
- What kind of information was collected? Where did the information originate?
- What analyses were used? What are the assumptions underlying the analyses?

Objective
- What are the results? How are these presented (e.g., p-values, confidence intervals, figures)?
- Is all of the necessary information provided to make an assessment? What is missing?
- How accurate or precise is the information (e.g., estimates of error, sensitivity analyses)?

Assessment
- What are the biases, confounding factors, and other considerations influencing appraisal of the subjective and objective information? Do these strengthen or weaken the findings?
- What relevant contextual questions are not answered by this study? How might these be addressed?
- Taking all of this information together, what is a reasonable interpretation of the study findings (even if it differs from that of others who read the same study)?

Plan
- Despite any limitations of the study, can this information be useful? If so, how and when might it be applied? What are the limitations or alternatives of this plan?
- What other information should be sought to aid in answering the original question? How can this information be obtained?

Conclusion
There is no substitute for genuine critical appraisal of the medical evidence. Medical learners who lack training in this skill may learn to rely instead on convenience (e.g., choosing a randomized trial to present to the group and, upon finding a low p-value, accepting its results at face value) or cynicism (e.g., identifying a shortcoming of a study without offering reasoned explanation about how it affects the results and disregarding anything to be gained from the research). However, critical appraisal skills may be successfully incorporated into evidence-based medical curricula by starting with what makes sense and using this as a context for more challenging concepts, by limiting oversimplification of the appraisal process, and by encouraging students to develop a systematic, yet nonalgorithmic, approach to evidence appraisal.

References

1. Bayley L, McLellan A, Petropoulos JA. Fluency-not competency or expertise-is needed to incorporate evidence into practice. *Acad Med.* 2012;87(11)1470.
2. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA.* 2007;298(9):1010-1022.
3. Baird JS. Journal clubs: what not to do. *Acad Med.* 2012;87(3):257.
4. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet.* 2008;372(9656):2152-2161.

Martha Carvour, MD, PhD, is a first-year resident in internal medicine at the University of Texas Southwestern Physician Scientist Training Program in Dallas. Prior to residency, Dr. Carvour completed her MD and PhD (epidemiology) degrees at the University of Iowa, where she first became involved in the design, development, and delivery of an evidence-based medicine curriculum.

Related in VM

Rating Evidence in Medical Literature, January 2011

Insights from Teaching Evidence-Based Medicine, January 2011

Authority, Health Advocacy Organizations, and Scientific Evidence, January 2013