

Editorial

Finding Truths in Clinical Medicine: Through the Looking Glass—Cracked

David L. Schriger, MD, MPH
UCLA Emergency Medicine Center
UCLA School of Medicine
Los Angeles, CA

Reprints not available from the author.

Address for correspondence: David L. Schriger, MD, MPH, UCLA Emergency Medicine Center, UCLA School of Medicine, Suite 300, 924 Westwood Blvd, Los Angeles, CA 90024; E-mail schriger@ucla.edu.

Copyright © 2001 by the American College of Emergency Physicians.

0196-0644/2001/\$35.00 + 0

47/1/119251

doi:10.1067/mem.2001.119251

Finding Truths in Clinical Medicine: Through the Looking Glass—Cracked

See related article, p. 518.

[Schriger DL. Finding truths in clinical medicine: through the looking glass—cracked. *Ann Emerg Med*. November 2001;38:566-569.]

“In general we look for a new law by the following process. First we guess it.”—Richard Feynman,¹ Nobel Laureate in Physics

An emergency physician, whether working in a private hospital trying to do the right thing for his or her patients, working in an academic program trying to teach the residents the right thing, or working on the American College of Emergency Physicians' clinical policy committee trying to discern and disseminate the right strategy, faces a conundrum—what is the right thing? We seek a singular, unambiguous truth. We encounter a chimera of incomplete, contradictory data and a host of interpretations and misinterpretations of that data.

There is nothing new or unique about this phenomenon; professionals are expected to act under circumstances of incomplete and contradictory information. What has varied over time is the medical community's response to uncertainty. In the mid-1900s, the majority of uncertainty was managed through paternalism. The practitioner absorbed the uncertainty and presented the patient with the diagnosis and a single treatment option.² How did the practitioner know what to do? More than likely, either an expert taught him during training or it was spelled out in a review article. The review article presented an expert's opinion. The opinion was supported by an intentionally selective list of references. Alternative hypotheses and positions were seldom mentioned. Equivocation was not a valued characteristic of such papers. Just as the physician “protected” the patient from the maelstrom of uncer-

tainty, the expert did the same for the physician. This system was not without merit. By hiding much of the uncertainty behind the experts' directives, the system freed patients and physicians to proceed under the illusion of certainty. This illusion decreased collective anxiety, even if, at times, it obscured the truth.

In the past few decades, some physicians and academics have shed this cloak of paternalism and brought uncertainty into the light where it can be examined, understood, and, hopefully, managed. Practitioners now commonly function as consultants, helping patients choose from a variety of options.² The current interest in errors in medicine (alternatively formulated as uncertainty about whether proper care will be provided) is another example.³ In recognizing that experts are often wrong, structured reviews, that is, systematic accountings of all of the evidence, have supplanted the old-fashioned review article as the in-vogue method for divining the truth.⁴⁻⁶

Although the disclosure of uncertainty seems honest and desirable, it may have unintended negative consequences. Some patients may not wish to bear the anxiety that comes with the acknowledgment of uncertainty.² Some practitioners may feel the same. The physician who wants practical guidance on how to respond to a clinical situation and follows good evidence-based medicine practice by searching for a quality structured review on the topic, may be disappointed if the review does not offer a definitive recommendation. Yet, the conclusions of a properly executed structured review should be more uncertain than those of an expert review, because contradictory evidence is not suppressed. Evidence-based medicine may be *au courant*, but a textbook of emergency medicine based solely on evidence would be too heavy to carry and would seldom provide useful direction. There is an inherent opposition between the desire for an unambiguous recommendation and the desire to openly acknowledge uncertainty.

Authors of structured reviews presumably initiate their effort with the goal of getting a definitive answer. When they encounter data that are too sparse, too contradictory, or too far removed from the specific clinical situation to permit synthesis of a meaningful recommendation, they may not be satisfied. The most honest conclusion, "we really don't know the answer," seems inadequate. The authors may be strongly tempted to make assumptions about the data that, by filling in holes and smoothing over contradictions, permit the derivation of an unambiguous conclusion. Most statistical modeling efforts do exactly this.⁷ However, we must recognize that, when the authors

of the structured review make their assumptions, they are doing something not at all dissimilar to what the expert did when he selectively culled the evidence to support his point of view.

Although its methods may be more explicit and seemingly rigorous than those of an expert review, the structured review also requires a host of judgments: what databases to search, what search terms to use, which articles to evaluate, how to weigh them, how to combine them, and what conclusions to reach. Each of these judgments requires assumptions, and each assumption can introduce bias. No review process eliminates judgment. No review process is immune to bias.

In this issue of *Annals*, Kelly et al⁸ examine the quality of structured reviews that have appeared in emergency medicine journals. They evaluate each review with respect to the aforementioned judgments. Their premise is that if proper methods are used to make each judgment, the review will be more likely to produce an accurate conclusion. From this evaluation process, they conclude that the quality of most reviews is fair to poor.

Although I am in complete agreement with their conclusion and believe that there is much room for improvement in the quality of emergency medicine structured reviews, I have reservations about their method of reaching it. I fear that readers may misread their paper and assume that compliance with their checklist ensures that a review is unbiased. I fear that readers will assume that the checklist has predictive value—that papers in compliance with the checklist are highly likely to provide an accurate estimate of the truth, whereas papers that score poorly are less likely to be accurate.

There is, however, little evidence that any particular method for making the judgments analyzed by their checklist uniformly produces a less-biased result. In fact, there is evidence to suggest that some of the processes advocated by their checklist, such as quality scoring, may increase bias.⁹⁻¹¹ Greenland¹⁰ has written, "Perhaps the most insidious form of subjectivity masquerading as objectivity in meta-analysis is 'quality scoring.' This practice subjectively merges objective information with arbitrary judgments in a manner that can obscure important heterogeneity among study results." These sentences strike at the Achilles' heel of structured reviews: the use of complex routines to create the appearance of objectivity.

Fortunately, the analysis of knowledge summarization efforts can be enhanced through the use of a more complex model recently offered by Maclure and Schneeweiss¹² (Figure 1). Using an astronomical analogy—the difference between a celestial object as seen through a telescope and

the celestial object as it truly is, they suggest that the evidence collection, selection, and review process is akin to using a telescope. Because they are concerned with the summation of knowledge regarding epidemiologic issues, they call their model the Episcopo.

Just as the wise astronomer recognizes that dirt on the lenses, imperfect lenses, atmospheric pollution, and gravitational pull on light may distort the image in his eyepiece from the true celestial form, the wise clinician understands that estimates of effect derived from a structured review may differ from the true effect. The Episcopo is composed of 10 discrete layers in which bias can be introduced. Biases that come from individual studies include: random events (b), confounding (c), misclassification of exposure status or outcome (d and e), data-handling errors (f and h), and analytic errors (g and i). The process of combining the results of individual studies can introduce bias at levels j and k. Although the Episcopo is concerned with observational studies, with a slight alteration in the definition of each level, the format can be applied to randomized trials. When this is done, it

becomes clear that randomization, although undeniably an important step in the creation of groups with equivalent expected outcomes, by no means eliminates the potential for bias. Several features of the Episcopo are worthy of notice. First, it is complex; a large number of assumptions must be made to traverse the 10 steps from truth (level a) to knowledge use (level k). Second, our confidence in the overall result is dependent on our ability to estimate bias at the level in which we are most uncertain. Said another way: a team is only as strong as its weakest player. We will often have insufficient knowledge to estimate accurately and adjust for bias at one or more levels. We should fully expect that structured reviews that produce biased, imprecise estimates would be the rule, not the exception.

Kelly et al⁸ evaluate structured reviews using a checklist developed by Oxman et al.¹³ Although the checklist has been subjected to evaluations of its interrater reliability¹³ and content and construct validity,¹⁴ there has been no attempt to establish that structured reviews that score higher on the checklist better approximate the truth. Of

Figure 1.

*The Episcopo. The user looks at the output of the device (level k) and sees the “known” risk difference (^kRD) (or any other measure of effect). The known RD results or information transmitted, like light waves through a telescope, from a causal (“etiologic”) RD (^aRD) in a target population, through layers of “lenses” and “filters.” Each layer is a distinct domain in which certain types of biases operate, potentially adding additional distortion. It is only by considering the biases introduced at each level that we can determine to what degree ^kRD is an accurate proxy for ^aRD. From Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology*. 2001;12:114-122. Used with permission from Lippincott Williams & Wilkins, Baltimore, MD.*



greatest concern is that the checklist is concerned solely with levels j and k of the Episcopo. The belief that if levels j and k are properly (ie, according to the rules of Oxman et al) carried out, then all of the biases in levels a-i will be accounted for is implicit in the analysis by Kelly et al. Is this a wise assumption? We kick a tire in the used car lot and quickly move on if the bumper falls off, but do we truly believe that a few cursory checks are as good as a systematic review of what's under the hood and chassis? If we are unwilling to dirty our hands and sort through each piece of evidence at each of the levels, are we likely to be confident that our estimates are correct, or have we merely committed the error described by Greenland?¹⁰ The model that Kelly et al propose is depicted in Figure 2. It is much simpler than the Episcopo. That is its appeal. Unfortunately, simplicity comes at a price; the method only works if the 5 dimensions adequately capture the complexity of the 11 levels in the model by Maclure and Schneeweiss.¹² There is little reason to believe it will. The only way to truly determine the value of a recommendation is to examine each assumption made in reaching it.

If structured reviews most often lead to an inconclusive answer or an answer dependent on as many assumptions as an expert review, why should we do them? Several features stand in their favor. First, their explicitness is a potential benefit. By breaking the analysis into small steps instead of global subjective judgments, the structured review establishes a firmer framework for the consideration of the underlying assumptions.¹⁵ Second, by showing

all of the data, the structured review empowers readers to reach their own conclusions. Third, explicitness may elevate the quality of discourse, as proponents and dissenters pinpoint the issues on which they disagree rather than engaging in more unfocused forms of debate. Finally, explicitness may better illustrate the crucial unknowns and set a better agenda for future research.

Readers should understand, however, that structured reviews are not the wonderful solution that some evidence-based medicine advocates suggest.¹⁶ Those done with limited assumptions are often inconclusive. Those that reach definitive conclusions may be as loaded with questionable assumptions as an expert review. The only way to assess the reasonableness of a clinical recommendation is to examine its underlying assumptions.¹⁷ There are no shortcuts. Things really haven't changed very much. Clinicians must either follow the advice of an expert they trust or be willing to go through a detailed examination of the evidence to reach their own conclusion.

1. Feynman RP. *The Character of Physical Law*. Cambridge, MA: MIT Press; 1965.
2. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. *JAMA*. 1992;267:2221-2226.
3. Leape L. Error in medicine. *JAMA*. 1994;272:1851-1857.
4. Eddy DM. Clinical decision making: from theory to practice. The challenge. *JAMA*. 1990;263:287-290.
5. Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA*. 1998;280:278-280.
6. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature VI. How to use an overview. *JAMA*. 1995;272:1367-1371.
7. Leamer EE. *Specification Searches: Ad Hoc Inference With Nonexperimental Data*. New York, NY: Wiley; 1978:1-3.
8. Kelly KD, Travers A, Dorgan M, et al. Evaluating the quality of systematic reviews in the emergency medicine literature. *Ann Emerg Med*. 2001;38:518-526.
9. Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054-1060.
10. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol*. 1994;140:290-296.
11. Greenland S. Quality scores are useless and potentially misleading [letter]. *Am J Epidemiol*. 1994;140:300-301.
12. Maclure M, Schneeweiss S. Causation of bias: the Episcopo. *Epidemiology*. 2001;12:114-122.
13. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol*. 1991;44:91-98.
14. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991;44:1271-1278.
15. Eddy DM. Practice policies: where do they come from? *JAMA*. 1990;263:1265, 1269, 1272, passim.
16. Schriger DL. One is the loneliest number: be skeptical of evidence summaries based on limited literature reviews. *Ann Emerg Med*. 2000;36:517-519.
17. Florin D. Scientific uncertainty and the role of expert advice: the case of health checks for coronary heart disease prevention by general practitioners in the UK. *Soc Sci Med*. 1999;49:1269-1283.

Figure 2.

A knowledge acquisition device modeled after Kelly et al.⁸ These authors suggest that following the steps in level b leads to an accurate assessment of the truth. Compared with the Episcopo, this model may be overly simplistic.

