

## **Narrowing the focus and broadening horizons: Complementary roles for systematic and nonsystematic reviews**

**David A. Cook**

Received: 25 September 2008 / Accepted: 25 September 2008 / Published online: 11 October 2008  
© Springer Science+Business Media B.V. 2008

In a recent editorial, Eva discussed the “limits of systematicity” (Eva 2008). His comments highlight a number of legitimate concerns regarding the validity and usefulness of systematic reviews. He notes that bias—“systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others” (Merriam-Webster)—is unavoidable. Given this, he posits that we ought to embrace our biases, avoid the pretence of systematicity altogether, and focus instead on nonsystematic syntheses. While I do not disagree with his critique, I believe his solution to the problem swings the pendulum further than needed. In this article, I will argue that both systematic and nonsystematic reviews play vital and complementary roles in advancing the art and science of medical education.

### **The limits of systematicity**

The purpose of a systematic review is to identify and summarize all research germane to a focused research question using methods that limit bias and random error (Cook et al. 1997). A systematic review begins with a clear statement of a focused question, followed by a comprehensive literature search to identify potentially relevant articles, application of predefined criteria to exclude irrelevant articles, abstraction of key information, and summarization of that information in a succinct and easily understood format. Some systematic reviews include a meta-analysis that statistically combines the quantitative results of many studies into a single average (pooled) estimate of effect. However, not all systematic reviews attempt meta-analyses. Indeed, if studies vary widely then pooling of results is inappropriate, and a narrative review of the systematically identified evidence is conducted.

While every attempt is made to minimize bias in a rigorous systematic review, bias is, as Eva notes, unavoidable. First, each included study will be biased (i.e., favor one

---

D. A. Cook (✉)

Division of General Internal Medicine, Mayo Clinic College of Medicine, Baldwin 4-A,  
200 First Street SW, Rochester, MN 55905, USA  
e-mail: cook.david33@mayo.edu

intervention over the other) to some degree. A recent study demonstrated that if these biases are not randomly distributed among the included studies (as is often the case) then a meta-analytic pooled estimate will reflect this bias and thus will be inaccurate (Colliver et al. 2008). Furthermore, such pooling glosses over important differences between studies in the participants, interventions, and outcomes. While this won't necessarily introduce bias, such differences certainly affect the conclusions that can be drawn. Second, studies that report statistically significant results or results that favor accepted practices are often easier to publish than studies with results that are not statistically significant or that are unexpected. When this occurs published studies reflect an inaccurate (biased) perspective (Montori et al. 2000). A good systematic review researcher will attempt to identify such limitations, and account for or acknowledge them.

In addition to the biases inherent in the literature, reviewers cannot avoid introducing their own biases. These preconceived notions may manifest in the criteria established for review, the manner in which such criteria are applied, the data selected for abstraction, the abstraction process itself, and the presentation and discussion of results. It is these limitations that prompted Eva (2008) to "place less emphasis on the 'systematic' and more emphasis on the 'narrative' [review]."

### **The merits of systematicity**

While these concerns regarding the presence of bias in systematic reviews are correct, such reviews nonetheless serve many useful purposes. If nothing else, systematic reviews provide a comprehensive (within the limitations noted above) list of articles relevant to the focused research question. Such lists—even without the accompanying synthesis (narrative or otherwise)—can be of great benefit to other researchers in the field and to educators seeking information on the topic. Such reviews also highlight gaps in the evidence, noting for example that nearly all studies in a given field have focused on medical students, or used widely divergent interventions, or employed weak research designs. Future researchers can use such information to focus their own research efforts to address these gaps.

More importantly, however, there are many instances in which systematicity—even with all its flaws—is better than the alternative. While systematic reviews are susceptible to bias in study selection, this effect is compounded (sometimes egregiously so) in non-systematic reviews. The ultimate intent of science is to inform practice. The translation of original research to practical recommendations typically requires a review and synthesis of evidence. However, if an author focuses on studies already in his or her file cabinet he or she may inadvertently miss important studies showing contrary results that might mitigate the recommendation, or supportive studies that might strengthen it. Systematic reviews, despite their flaws, lessen (even if not eliminating) the potential for such error.

### **A balanced approach**

Nonsystematic, critical syntheses of the sort Eva proposes are not without their own merit. Thoughtful reflection upon a theme, drawing upon research, frameworks, and philosophy both within one's field and from other fields, can yield insights that a systematic review could never achieve. Indeed, the very rules that enhance the systematic review's rigor blind the researcher to ideas outside the scope of the focused question and resultant search

strategy. A systematic review focused on original research on professionalism involving medical students will, by definition, exclude studies involving residents, relevant writings without empiric data, and evidence from outside the medical profession. Such focus is clearly both a strength and a liability.

Examples of nonsystematic, critical syntheses abound in medical education. These articles, often provocative and occasionally paradigm-shifting, have broadened our horizons and changed the way we look at clinical performance assessment (Williams et al. 2003), self-assessment (Eva and Regehr 2005), clinical reasoning (Eva et al. 1998; Norman and Schmidt 1992), and research in computer-assisted instruction (Cook 2005) to name a few.

Clearly, both systematic and nonsystematic approaches have merit.

This polarization reminds me of the either-or debate regarding the merits of quantitative and qualitative research paradigms. Most researchers now realize that each approach has advantages and disadvantages, and that they answer different questions and thereby complement each other (Ercikan and Roth 2006).

Taking this analogy further, there are many similarities between the quantitative research paradigm and the systematic review, and likewise many parallels between the qualitative tradition and the nonsystematic, critical synthesis (see Table 1).

Both quantitative research and systematic reviews prefer large samples (of subjects or research studies) and emphasize systematic sampling according to a prespecified algorithm, numerical data, and detachment of the researcher from the analysis. Homogeneity is desirable—in the population studied, the interventions (if any), the outcome measure, and (for reviews) the study designs. Differences between individual subjects/studies are viewed as error to be ignored if possible while averaging results to find a best estimate of effect.

In contrast, both qualitative research and nonsystematic reviews emphasize purposive, iterative sampling (of participants, research studies, and other data sources) that shapes and is shaped by the emerging insights. Differences between subjects/studies are seen, not as error, but as important inconsistencies that merit explanation, requiring additional data collection and often yielding novel insights. Rather than emphasizing large samples these approaches contrast information from multiple sources (triangulation), often including evidence that does not involve medical education learners. In both cases, results are presented as a rich narrative full of insights and often a critique of the status quo.

Each pair of approaches also shares a common purpose. Systematic reviews and quantitative research seek to summarize large amounts of data, typically to evaluate a priori hypotheses. Nonsystematic reviews and qualitative research, on the other hand, seek to produce novel insights and tend to be hypothesis-generating.

Perhaps, then, we ought not encourage one review approach over another. Rather, it seems that potential authors should clarify their purpose and then select the design most appropriate to meet that need. Perhaps too, as with original research, mixed methods—combining the best of both approaches—will be appropriate in many cases.

Standards for rigor in systematic reviews and meta-analyses are well-codified (Moher et al. 1999; Stroup et al. 2000). However, I am not aware of similar standards for nonsystematic reviews. Fortunately, standards for qualitative research have been developed to facilitate rigorous, defensible, trustworthy results (Côté and Turgeon 2005; Devers 1999; Elliott et al. 1999; Malterud 2001). Attention to principles of qualitative research such as clarification of the question or purpose, reflexivity (identifying researcher assumptions and perspectives), collaboration with other researchers, purposeful sampling of diverse data sources, and inductive critical analysis including examples, counterexamples (i.e., evidence that does not fit with the rest), triangulation of evidence, and a conscious effort to

**Table 1** Parallels between systematic/nonsystematic reviews and quantitative/qualitative original research

	Systematic reviews and quantitative original research	Nonsystematic reviews and qualitative original research
Purpose	Summarize large amounts of data, hypothesis confirming	Generate novel insights, hypothesis generating
Archetypical study design	Meta-analysis (review), randomized trial (original)	Insightful synthesis (review), inductive thematic analysis (original)
Data selection	Comprehensive, systematic	Purposeful, selective
Data inclusion criteria	Similar (homogenous) population, intervention, and outcome measure for all data sources	Perceived relevance to question; intentionally seeks and highlights dissimilar sources
Data analysis	Quantitative (tests of inference or meta-analysis)	Qualitative (narrative)
Results presented as ...	Single number or series of numbers	Rich, critical, insightful narrative
Attribution of disparities between data sources	Differences are error (heterogeneity)	Differences suggest novel insights
Validity derived from ...	Sufficient numbers of similar data sources, absence of bias in data sources	Triangulation (concordant findings from multiple sources)
Threats to validity	Systematic error (bias) in sampling, excessively dissimilar data sources (heterogeneity)	Researcher bias in analysis, failure to triangulate or sample to saturation
Role of researcher	Impartial, detached	Immersed, interactive
Useful when ...	Focused question, tentative answer has been identified (hypothesis), and sufficiently numerous and homogenous data sources are available	Question not amenable to quantitative methods, emerging field without clear hypotheses, or insufficient data sources available for hypothesis-confirming research

In all of the above, think of the “data source” as an individual research participant for original research (although non-human sources could be envisioned as well), and as an original research study for a review

consider alternate perspectives may likewise facilitate rigorous nonsystematic reviews. Ultimately, a nonsystematic review will be judged by the degree to which it “identifies knowledge that is well established, highlights gaps in understanding, and provides some guidance regarding what remains to be understood” (Eva 2008).

### Acknowledging bias and embracing diversity

In summary, both nonsystematic reviews and systematic reviews serve important roles. The former integrate research from diverse fields and identify new insights, while the latter summarize research on focused topics and highlight strengths and weaknesses in existing bodies of evidence. Both approaches are susceptible to bias that can be reduced but never eliminated. Rather than embracing such bias, perhaps a more appropriate response is to acknowledge these biases and encourage readers to interpret findings in that context.

Authors conducting systematic and nonsystematic reviews on a similar topic are likely to arrive at different conclusions, and thus both approaches can stimulate creative debate. If the factors that lead authors to interpret research differently are labeled bias then I agree that this should be embraced, since creative debates tend to sharpen thinking and stimulate the search for additional evidence. However, perhaps the reflective, rational, and intentional cognitive processes that drive divergent interpretations merit a less disparaging label. It is this *diversity*, and not the bias per se, that warrants embracing.

## References

- Colliver, J. A., Kucera, K., & Verhulst, S. J. (2008). Meta-analysis of quasi-experimental research: Are systematic narrative reviews indicated? *Medical Education*, *42*, 858–865.
- Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, *80*, 541–548.
- Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, *126*(5), 376–380.
- Côté, L., & Turgeon, J. (2005). Appraising qualitative research articles in medicine and medical education. *Medical Teacher*, *27*, 71–75.
- Devers, K. J. (1999). How will we know “good” qualitative research when we see it? Beginning the dialogue in health services research. *Health Services Research*, *34*(5 part 2), 1153–1188.
- Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, *38*, 215–229.
- Ercikan, K., & Roth, W.-M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, *35*(5), 14–23.
- Eva, K. W. (2008). On the limits of systematicity. *Medical Education*, *42*, 852–853.
- Eva, K. W., Neville, A. J., & Norman, G. R. (1998). Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Academic Medicine*, *73*(10 Suppl), S1–S5.
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine*, *80*(10 Suppl), S46–S54.
- Malterud, K. (2001). Qualitative research: Standards, challenges, guidelines. *Lancet*, *358*, 483–488.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet*, *354*, 1896–1900.
- Montori, V. M., Smieja, M., & Guyatt, G. H. (2000). Publication bias: A brief review for clinicians. *Mayo Clinic Proceedings*, *75*, 1284–1288.
- Norman, G. R., & Schmidt, H. G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine*, *67*, 557–565.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., et al. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA*, *283*, 2008–2012.
- Williams, R. G., Klamen, D. A., & McGaghie, W. C. (2003). Cognitive, social, and environmental sources of bias in clinical performance ratings. *Teaching & Learning in Medicine*, *15*, 270–292.