

Huth EJ (2008). The move toward setting scientific standards for the content of medical review articles.



© Edward Huth, 1124 Morris Avenue, Bryn Mawr, PA 19010-1712, USA. E-mail: ejhuth23@verizon.net

The dependence of medical practitioners on medical reviews

In 1987, while Editor of the *Annals of Internal Medicine*, I wrote and published an editorial with the title *Needed: review articles with more scientific rigor* (Huth 1987). I was prompted to do this by a paper I had initially rejected but eventually accepted for publication in the *Annals* with my accompanying editorial commentary. In retrospect, Cynthia Mulrow's *The Medical Review Article: State of the Science* (Mulrow 1987) can now be seen as a landmark in medicine's long road from 'experience and expertise' to 'evidence' as the justification for particular medical treatments.

Cite as:

Huth EJ (2008). The move toward setting scientific standards for the content of medical review articles. *JLL Bulletin: Commentaries on the history of treatment evaluation* (www.jameslindlibrary.org).

When and where this road begins will remain a matter for argument: The James Lind Library offers many candidate 'beginnings'. Despite the growing development during the 18th, 19th, and 20th centuries of quantitative data for judgements on treatments, clinicians in these periods continued to rely on 'expert' judgements for choices of treatments. The challenge facing doctors who wish to identify evidence relevant to their practice among the plethora of potentially relevant reports was recognised as a problem at least as early as the 18th century. Andrew Duncan's editorial *Introduction to the first issue of Philosophical and Medical Commentaries*, published in 1773, has a remarkably familiar ring:

Medicine has long been cultivated with assiduity and attention, but is still capable of farther improvement. Attentive observation, and the collection of useful facts, are the means by which this end may be most readily obtained. In no age [...] does greater regard seem to have been paid to these particulars, than in the present. From the liberal spirit of inquiry which universally prevails, it is not surprising that scarce a day should pass without something being communicated to the public as a discovery or an improvement in medicine. It is, however, to be regretted, that the information which can by this means be acquired, is scattered through a great number of volumes, many of which are so expensive, that they can be purchased for the libraries of public societies only, or of very wealthy individuals.[...]

[...]No one, who wishes to practise medicine, either with safety to others, or credit to himself, will incline to remain ignorant of any discovery which time or attention has brought to light. But it is well known that the greatest part of those who are engaged in the actual prosecution of this art, have neither leisure nor opportunity for very extensive reading. (quoted in Chalmers and Tröhler 2000).

Because most doctors lack sufficient opportunity "for very extensive reading", they turn to summary views of evidence and expertise presented in synoptic form as textbooks, review articles, and medical meetings. Although the quality of reports of clinical trials in the second half of the 20th century has raised the value of journals for doctors' judgements about treatments, the relentless growth of the number of medical journals has meant that doctors seeking in them reliable data and conclusions have faced a daunting task. Some unpublished research I carried out about twenty years ago found that the ratio of the total number of medical journals to the number of physicians in the United States was actually fairly constant through many years. Hence one might suppose that the task of searching for the desired synoptic views of treatments would not go up. But the increased scattering of journals among more subspecialties of medicine meant that papers on any particular topic would be highly likely to be in journals not seen routinely by a physician or in journals difficult of access. This increased scattering of journals among medical specialties would, indeed, further raise the difficulties of finding all reports possibly relevant to his or her interests. And, far from solving the problem of 'information overload', the Internet often seems to have exacerbated it. In his 1981 book entitled *Coping with the biomedical literature*, Kenneth Warren stated the problem pithily.

... no matter what strategy is involved, attempts to deal with the literature in a comprehensive way are time-consuming indeed, perhaps leaving little time for practice and research. (Warren 1981, p 18).

Most clinicians are far too busy to find relevant articles reporting clinical trials, let alone to read them and digest their conclusions for use in clinical decisions about treatment for their patients. Most of them, of necessity, have continued to rely on synoptic views of proper treatments for particular problems, such as those appearing as review articles and, less frequently, as editorials.

Promoting awareness of the scientific quality of medical reviews

Because physicians have to rely on synoptic views of available treatments and their efficacy, the question of the reliability of review articles is obviously important. How sound are the data assembled on which review authors draw, and how free from biases are authors' methods in arriving at their judgements? In essence, are the authors truly reliable 'experts'?

Readers of journals have tended to trust their editors, editorial boards, and peer-reviewers to ensure the reliability and value of the synoptic views they publish. But how far can they be trusted? How thoroughly have the authors of such synoptic views searched medical literature for pertinent sources? How critically have they judged the reliability and quality of reports of treatments from which they will assemble the evidence for their synoptic conclusions?

It is clear from the content of review articles in clinical journals through many years that such questions were rarely, if ever, answered in them. As the Editor between 1971 and 1990 of a major clinical journal, *Annals of Internal Medicine*, the journal of the American College of Physicians, I can testify that authors of review articles were regarded as 'experts' to whom questions about the methods used in their reviews need not be raised. Even 'experts' can turn out to be non-experts in judging the validity of reviews solely from an apparent 'non-expertise' of the author of a review. I can draw a relevant example from my term as the Associate Editor of that journal from 1965 to 1971. A neurologist in Philadelphia submitted to the *Annals* a review of studies of cerebral blood flow in patients with neurologic diseases. The then-Editor sent an inquiry to a member of the *Annals'* Editorial Board, an internationally known expert on cerebral vascular disease on the staff of an internationally renowned medical center, asking him whether he would be willing to peer-review the review article. His reply was "Don't bother with considering that review; I have never heard of the author." Accepting this advice, the Editor of the *Annals* returned the review to the author with no further consideration of it. Ironically, the author then submitted his review to another even more eminent journal, which published it! The review became widely cited despite its 'non-expert' author. Whether the review answered the questions posed above about the reliability of its conclusions is not directly pertinent. If the author was not an 'expert', such questions need not be asked; if he was an 'expert', they need not be asked!

Questions relevant to judgements on the reliability of the conclusions reached in review articles were posed earlier in the social sciences than they were in the medical sciences (Chalmers et al. 2002), and some social scientists were aware of the relevance of their thinking to medicine. In *Summing Up: The Science of Reviewing Research*, for example, Light and Pillemer (1984) wrote:

For many years, the 'literature review' has been a routine step along the way to presenting a new study or laying the groundwork for an innovation. Journals such as Psychological Bulletin, Review of Educational Research, American Public Health Journal, and New England Journal of Medicine publish the best of such reviews. Traditionally, these efforts to accumulate information have been unsystematic. Studies are presented in serial fashion, with strengths and weaknesses discussed selectively and informally. These informal reviews often have several shortcomings:

1. *The traditional review is subjective....*
2. *The traditional review is scientifically unsound....*
3. *The traditional review is an inefficient way to extract useful information.... (Light and Pillemer 1984, p 3,4).*

The five chapters that follow discuss in detail the procedures authors of reviews should follow in preparing reviews. Their concluding chapter poses ten specific questions that authors of reviews should answer for readers:

1. What is the precise purpose of the review?
2. How were the studies selected?
3. Is there publication bias?
4. Are treatments similar enough to combine?
5. Are control groups similar enough to combine?
6. What is the distribution of study outcomes?
7. Are outcomes related to research design?
8. Are outcomes related to characteristics of programs, participants, and settings?
9. Is the unit of analysis similar across studies?
10. What are the guidelines for future research? (Light and Pillemer 1984, p 160, 161).

Cynthia Mulrow's 1987 article documented and exposed the poor scientific quality of medical reviews. She made clear

in the Methods section of her paper that her assessment of the quality of the review articles covered in her study drew on Light and Pillemer's recommendations, although she narrowed their list of ten questions to eight. But the Light and Pillemer book was not the initial impetus for beginning her study. In response to my request that she describe why she undertook the study that led to her 1987 *Annals* article, this is what she had to say (Mulrow 2008):

As a general medicine fellow at Duke in 1983, I wrote a review. I did much library work (searching and sorting) to find trials that had evaluated digitalis for heart failure and then critically appraised that evidence. I had never heard of 'systematic reviews' or 'meta-analyses' at that time. Of note, Annals published that review (Mulrow et al. 1984).

I then went to the London School of Hygiene and Tropical Medicine on a Milbank Scholarship and got a Masters in Epidemiology. While there, I heard Richard Peto present a meta-analysis about aspirin and CAD [coronary artery disease]. It was the first time that I had ever heard of 'meta-analyses'. I remember being very sceptical about combining data regarding different doses of aspirin given at different times (after myocardial infarction I think - but my memory is foggy).

*I then returned to the States as a junior faculty person at the University of Texas Health Science Center at San Antonio. I remember attending multiple grand rounds where 'experts' dogmatically presented overviews of topics. I suspected that much/some of what they were saying was based on opinion rather than evidence. Somehow that spurred me to think about systematically finding and critiquing evidence (which is what in retrospect I had done in a crude way with the digitalis paper). I began to look for literature on reviews - and found much good work in the social science field. I applied that work to thinking about reviews published in medical journals and [voilà]--the *Annals* article.*

*Unbeknownst to me, Andy Oxman (who I had not yet heard of or met) was thinking about systematic reviews at the same time (and perhaps even earlier than I). He submitted work similar to mine (albeit his work was probably a bit better than mine) a few months after *Annals* took my article. My memory is that *Annals* ended up not publishing Andy's article because mine was submitted first.*

*So I don't have a good quote for you - only the above story. Multiple experiences, reading work outside of my primary area, and luck, I guess, were behind the *Annals* systematic review article.*

Mulrow does not mention in this account that, as the Editor of the *Annals of Internal Medicine* when she submitted her paper to the journal, I did not at first accept her paper. Why I did not, I cannot recall. I have asked her to search her own files to see if she could find correspondence we exchanged about her paper that might explain our initial rejection; she has not been able to find any. I have asked the current managers of the *Annals* to look in its files for a possible answer; apparently the *Annals*'s file on her paper no longer exists. I doubt that our decision was based on a disbelief in her conclusions and a judgement that they were inadequately supported in the paper. Some of our so-called 'rejections' were in fact what we internally called 'rejected; revision will be considered'. Perhaps some weaknesses in the presentation of her methods and her conclusions led to an initial decision of this kind. In addition, I hasten to admit that I was probably guilty of faith in the expertness of 'experts' writing their reviews, as were other editors at that time.

My ultimate decision to accept the Mulrow paper may have been due in part to my recalling my awareness of the value of good and apparently-reliable review articles in supporting a journal's usefulness and reputation. In 1986 the *Annals* published a paper by Eugene Garfield (1986) on the influence of the various types of articles on a journal's Impact Factor, as reported annually in *Science Citation Index*. In the period 1977 through 1982, 93.4% of the reviews published in *Annals* were cited in other journals and they contributed 16.0% of total citations, second only to the 56.0% contributed by original reports of research and other studies.

Despite whatever led to my initial rejection of Mulrow's paper, we changed our minds and went on to publish the paper — *Deo gratias!* In my editorial (Huth 1987) supporting our decision to publish her paper and lauding her conclusions I stated clearly the responsibilities of editors in publishing any paper, be it a report of clinical or laboratory research, or a review article.

Editors, including those of this journal, must share blame for the defects Mulrow reports; editors are responsible for judging the adequacy of evidence in papers they accept for publication.

Eugene Garfield, founder of *Science Citation Index*, generously carried out at my request a study (2008) which found that the Mulrow paper was cited 375 times in the period from its publication up to 2008. This prompted him to

comment to me that: "The 1987 article by CD Mulrow has been extremely popular". The largest numbers of citations have been in major journals such as the *British Medical Journal* (14), *Annals of Internal Medicine* (13), the *Journal of the American Medical Association* (13), and the *Journal of Clinical Epidemiology* (10). The citations of Mulrow's paper have been mostly in two types of papers: methodologic recommendations for reviews, and review articles prepared with Mulrow's standards in mind. Here are two citations of these types, examples taken from the *Journal of the National Cancer Institute*.

Weed DL. Methodologic guidelines for review papers. *Journal of the National Cancer Institute* 1997;89 (1):6-7. [a methodologic paper].

Trock BJ, Leonessa F, Clarke R. Multidrug resistance in breast cancer: a meta-analysis of MDR 1/gp170 expression and its possible functional significance. *Journal of the National Cancer Institute* 1997;89(13):917-931. [a review citing Mulrow's criteria for finding and judging evidence relevant to reliable conclusions].

Interestingly, Mulrow's article was referred to only once in *The New England Journal of Medicine* in this period, possibly reflecting an editorial antipathy to publishing systematic reviews and meta-analyses during the 1990s (Chalmers 2001).

Comment

Cynthia Mulrow's 1987 consciousness-raising article about the poor scientific quality of medical reviews was not the first published document drawing attention to the need to address this problem. Six years previously, Ed Kass (1981) had emphasized this need in general terms in his contribution to Kenneth Warren's book, noting that reviews "... need to be evaluated as critically as are primary scientific papers but with slightly different guidelines ..." (Kass 1981, p 82). As Mulrow herself notes, contemporaneously with her study, Andy Oxman was developing guidelines for improving the quality of medical reviews, building on the example that had been set by social scientists (Oxman and Guyatt 1988).

The important common feature of the contributions made by Kass, Mulrow and Oxman, however, is that they focus on measures needed to control biases in reviews (Chalmers et al. 2002). They avoided giving inappropriate prominence to 'meta-analysis', the statistical synthesis of data from separate but similar studies. 'Meta-analysis' as a term had been introduced a decade before Mulrow's article (Glass 1976), but, with a few notable exceptions (Jenicek 1987), use of the term was too often restricted to considerations of statistical synthesis (O'Rourke 2006), with insufficiently explicit attention given to the measures needed to reduce biases.

What Mulrow found in her survey of the qualities of medical review articles was that, to some degree, all of them lacked the essential structure of the scientific version of 'critical argument'. I have summarized elsewhere (Huth 1999) the components of an adequately sequenced and structured scientific paper.

- Statement of problem: posing of a question or stating a hypothesis
- Presentation of the [relevant] evidence
- Validity of the evidence
- Implications of the evidence: initial answer or judgement on the validity of the hypothesis
- Assessment of the answer's validity in the face of conflicting evidence
- Conclusion

The central message of Mulrow's and Oxman's papers are that these components and their structure should be expected not only in reports of clinical trials or laboratory research but also in review articles and similar synoptic documents as well.

One of the standards for a synoptic document like the medical review article was applied as far back as the mid-eighteenth century. Lind's *Treatise of the Scurvy* is known mainly for his account of a controlled clinical trial, but it is worth noting that most of his book was a review of what was known about the disease. Lind observes in his introduction that "before the subject could be set in clear and proper light, it was necessary to remove a great deal of rubbish" (Lind 1753). He goes on to document his strategy for locating potentially relevant evidence and his selection of 54 books meriting critical appraisal, and he provides abstracts summarising his incisive views of the chosen books (Milne and Chalmers 2004). Only rarely, if ever, in the following two and a half centuries was Lind's standard applied in medical reviews. Only now, more than a quarter of a millennium later, are more determined efforts being made to improve the quality of reviews through the setting of standards for their content, and Cynthia Mulrow's paper has undoubtedly been a milestone in these developments.

This *James Lind Library* commentary has been republished in the *Journal of the Royal Society of Medicine* 2009;102:247-51.

Acknowledgments

I thank Eugene Garfield for providing his invaluable data on medical-journal article citations of Cynthia Mulrow's 1987 *Annals* article and Dr. Mulrow for the quoted account of why and how she came to write it. Neither Dr. Garfield nor Dr. Mulrow should be assigned any responsibility for the content of this commentary.

References

- Chalmers I (2001). Foreword. In: Egger M, Davey Smith G, Altman D, eds. *Systematic reviews in health care: Meta-analysis in context*. 2nd Edition of *Systematic Reviews*. London: BMJ Books, p xiii-xviii.
- Chalmers I, Tröhler U (2000). Helping physicians to keep abreast of the medical literature: 'Medical and Philosophical Commentaries', 1773-1795. *Annals of Internal Medicine* 133:238-243.
- Chalmers I, Hedges LV, Cooper H (2002). A brief history of research synthesis. *Evaluation & The Health Professions*. 25:12-37.
- Garfield E (1986). Which medical journals have the greatest impact? *Annals of Internal Medicine* 105:313-320.
- Garfield E (2008). Personal communication to Edward J. Huth; 17 July 2008.
- Glass GV (1976). Primary, secondary and meta-analysis of research. *Educational Researcher* 10:3-8.
- Huth EJ (1987). Needed: review articles with more scientific rigor. *Annals of Internal Medicine* 106:470-471.
- Huth EJ (1999). What is critical argument? In: *Writing and Publishing in Medicine*. Third edition. Baltimore: Williams & Wilkins, p 60-62.
- Jenicek M (1987). *Méta-analyse en médecine. Évaluation et synthèse de l'information clinique et épidémiologique*. St. Hyacinthe and Paris: EDISEM and Maloine Éditeurs.
- Kass EH (1981). Reviewing reviews. In: Warren KS, editor. *Coping with the biomedical literature: a primer for the scientist and the clinician*. New York: Praeger.
- Light RJ, Pillemer DB (1984). *Summing up: the science of reviewing research*. Cambridge, Massachusetts: Harvard University Press.
- Lind J (1753). *A treatise of the scurvy*. Edinburgh: Sands, Murray & Cochran.
- Milne I, Chalmers I (2004). Documenting the evidence: the case of scurvy. *Bulletin of the World Health Organisation* 82:791-792.
- Mulrow C (1987). The medical review article: state of the science. *Ann Intern Med* 1987;106:485-488.
- Mulrow C (2008). Personal communication to Edward J. Huth; 18 July 2008.
- Mulrow CD, Feussner JR, Velez R (1984). Reevaluation of digitalis efficacy: New light on an old leaf [review]. *Ann Intern Med* 101:113-7.
- O'Rourke K (2006). An historical perspective on meta-analysis: dealing quantitatively with varying study results. The James Lind Library (www.jameslindlibrary.org).
- Oxman AD, Guyatt GH (1988). Guidelines for reading literature reviews. *Canadian Medical Association Journal* 138:697-703.
- Warren KS (1981). Selective aspects of the biomedical literature. In: Warren KS, editor. *Coping with the biomedical literature: a primer for the scientist and the clinician*. New York: Praeger.