Critical Appraisal of Review Articles

BRIAN G. HUTCHISON, MD, CCFP

SUMMARY

Review articles are an important source of clinical information for family physicians. However, the volume of available reviews is vast and their scientific quality varies enormously. Family physicians must be able to identify trustworthy reviews quickly. This article outlines practical and flexible guidelines for critical appraisal and discusses the respective roles of review articles and original research reports in guiding clinical practice.

RÉSUMÉ

Les articles de synthèse sont une source importante d'informations cliniques pour les médecins de famille. Toutefois, le volume de ce type d'articles est trop vaste et leur qualité scientifique varie énormément. Les médecins de famille se doivent d'identifier rapidement les articles fiables. Cet article décrit un guide pratique et flexible facilitant l'évaluation critique et discute des rôles respectifs des articles de synthèse et des articles de nouveautés en recherche pour mieux orienter la pratique clinique.

Can Fam Physician 1993;39:1097-1102.

Because of the vast scope of family medicine, none of us can hope to be aware of (much less synthesize) all of the original research relevant to our discipline. As a result, family physicians, whether in the course of regular journal reading or when examining the literature to resolve a clinical dilemma, make extensive use of review articles as well as reports of original research. Inspired perhaps by the early literature on critical appraisal,[1-8] which dealt exclusively with original research, physicians – particularly academic family physicians – have tended to be apologetic about depending on review articles.

However, neither original research nor review articles have a monopoly on truth (or usefulness). Each has strengths and weaknesses, both practical and scientific. This paper will discuss the place of review articles for keeping up-to-date and for answering questions that arise in clinical practice, and will outline a practical approach to critical appraisal of review articles.

Dr Hutchison *is an Associate Professor in the Department of Family Medicine and is an associate member of the Department of Clinical Epidemiology and Biostatistics at McMaster University in Hamilton, Ont.*

Original research versus review articles

Original research encountered during the course of regular journal reading provides information that is more current than that available in review articles. In most cases, the difference is probably inconsequential. However, it assumes importance when a single study provides strong evidence on a question with serious clinical implications or when the study receives sensational coverage in the media (generating queries from concerned patients).

Compared with review articles, original research reports are often seen as lending themselves more readily to critical appraisal. Potentially relevant articles reporting original research can be screened methodologically to select studies likely to yield valid results. For example, you can choose not to read articles on the effectiveness of interventions unless they report randomized, controlled trials. Having read an article reporting original research, you can decide for yourself whether the study results are valid and applicable to your patients, using explicit and widely accepted criteria.[9,10]

Representativeness. As a guide to clinical action, however, original research reports have significant limitations. A single study is rarely definitive. More often, truth emerges from the accumulation of

evidence provided by several studies addressing a particular issue. The results of a single study can be misleading if they differ from those obtained in other studies examining the same question. The frequency with which randomized trials on the same topic yield contradictory results is highlighted by Horwitz'[11] identification of 36 topics with conflicting randomized trial results in a "nonexhaustive" search of cardiology and gastroenterology literature.

Table 1. **Published guidelines for assessing research reviews**

MULROW[17]
1. Was the specific purpose of the review stated?
2. Were sources and methods of the citation search identified?
3. Were explicit guidelines provided that determined the material included in, and excluded from, the review?
4. Was a methodologic validity assessment of material in the review performed?
5. Was the information systematically integrated with explication of data limitations and inconsistencies?
6. Was the information weighted or pooled?
7. Was a summary of pertinent findings provided?
8. Were specific directions for new research initiatives proposed?
OXMAN AND GUYATT[18]
1. Were the questions and method clearly stated?
2. Were comprehensive search methods used to locate relevant studies?
3. Were explicit methods used to determine which articles to include in the review?
4. Was the validity of the primary studies assessed?
5. Was the assessment of the primary studies reproducible and free from bias?
6. Was variation in the findings of the relevant studies analyzed?
7. Were the findings of the primary studies combined appropriately?
8. Were the reviewers' conclusions supported by the data cited?

The issue of representativeness becomes very important when we use research literature as a source of information to resolve clinical uncertainty or to inform clinical policy development. Unless we conduct a comprehensive search, using a variety of search strategies, we can be misled; the results of the studies in our sample might not represent the results of all relevant studies.

Bias. Similarly, although the authors of reports of original research often discuss their findings in light of previous related research, they cannot be counted upon to present a comprehensive and balanced review of the relevant literature. The existence of bias in citation of previous studies was demonstrated empirically by Gotzsche[12] in a study of the reference lists of reports of double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. He found an overrepresentation of references to trials with a positive outcome for the "new" drug in 67% of articles in which such bias could have occurred.

Statistical power. Reports of original research can also be misleading when, despite a methodologically strong design, they have inadequate sample sizes to detect (or rule out) clinically important effects or associations. Freiman and colleagues[13] illustrated how frequently this situation occurs. They analyzed 71 "negative" randomized, controlled trials (defined by $P > 0.05$), more than half of which were published in the *New England Journal of Medicine*, *The Lancet*, or the *Journal of the American Medical Association*. Fifty-seven (80%) of the trials had a 50% or greater chance of missing a true 25% reduction in risk of an unfavourable outcome. Thirty-one trials (44%) had a 50% or greater chance of missing a risk reduction of 50%. Viewing such negative trials in isolation can lead to the erroneous conclusion that an intervention is ineffective or that no association exists between a putative causal factor and a condition of interest, whereas an overview of all relevant studies might lead to an opposite (and correct) conclusion. For example, in an overview by Yusuf et al[14] of randomized trials of long-term β-receptor blockade following myocardial infarction, 19 of 23 individual trials failed to show a statistically significant reduction in mortality. However, the aggregated results of all trials showed that β-receptor blockade was associated with a highly statistically significant 23% reduction in risk of death (two-sided $P < 0.0001$).

Meta-analysis. This type of review, which aggregates the results of several

trials, is called a "meta-analysis." Such reviews are appearing with increasing (but still low) frequency in the health care literature. They provide a quantitative synthesis of the results of all relevant primary studies that meet predetermined inclusion criteria in order to produce an overall estimate of, for example, the effect of an intervention, the sensitivity and specificity of a diagnostic test, or the strength of an association between a medical condition and a possible prognostic or etiologic factor. Such research overviews have advantages over both original research reports and traditional narrative reviews as sources of information to resolve clinical dilemmas.

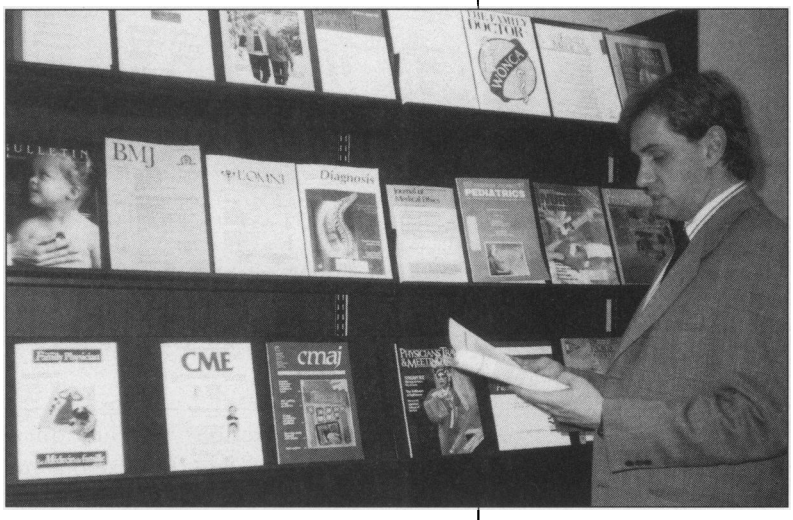
Properly done, meta-analysis eliminates much of the subjectivity and potential for bias from the process of reviewing primary research. Rigorous methodologic standards designed to minimize bias and ensure validity in the conduct of quantitative research overviews have been published.[15] By quantitatively combining results from more than one study, meta-analysis increases statistical power to detect clinically important effects or associations (reduces type II or β error) and increases the precision (narrows the confidence interval) of estimates of the magnitude of an intervention's effect, the sensitivity and specificity of a diagnostic test, or the strength of an association.

Meta-analysis is most helpful when a large number of studies (especially small studies) employing the same (preferably strong) methodology have variable, conflicting, or statistically nonsignificant results. Offsetting the obvious strength of meta-analysis is the fact that only a handful of primary care issues have been addressed to date with this method. Since 1989, meta-analysis has been a medical subject heading (MeSH) in the MEDLINE database and can be used as a search term in computerized literature searches.

At their best, review articles are valuable guides to action because they represent a comprehensive, unbiased assessment and synthesis of the primary research relevant to a clinical question. At a practical level, perhaps the most important advantage of review articles is that someone else does the laborious work of identifying, critically appraising, and synthesizing the relevant primary research.

Unfortunately, the quality of review articles is often poor. (Of course the same can be said for the quality of original research.)[16] Because the methods are rarely stated, review articles are somewhat harder to appraise critically than original research. However, criteria for assessing review articles have recently been developed; Mulrow[17] and Oxman and Guyatt[18] have published similar sets of guidelines for assessing the scientific quality of review articles (*Table 1*).

An instrument for assessing the scientific quality of review articles,[19] based on Oxman and Guyatt's criteria, has been shown to have satisfactory reliability (inter-assessor agreement) when used by experts in research methodology, clinicians with research training, and research assistants. Articles that satisfy the instrument's guidelines can be expected to provide valid answers to the questions addressed by the reviews. Unfortunately, only a small fraction of current reviews meet these standards. Until the overall quality of review articles improves substantially, a set of less stringent and more flexible guidelines are required to assist physicians to discard reviews of mediocre or inferior quality quickly after scanning. A set of such guidelines are presented below and summarized in *Table 2*.



Assessing potentially relevant reviews: *Critical appraisal can identify applicable studies in the vast medical literature.*

Examining a review article

Is the question being addressed clearly defined? Look for a clear statement of the target population, the intervention or exposure of interest, and the outcome(s) (eg, "Does acetysalicylic acid reduce mortality in patients with myocardial infarction?"). When questions are formulated in this fashion, you can make a decision quickly and confidently about the relevance to your clinical practice.

Table 2. **Framework for quality assessment of review articles**

POSITIVE FEATURES
Focuses on a specific question.
Clearly defines the question(s) being addressed.
Methods of conducting search for relevant primary studies are described.
Primary studies are critically appraised, preferably in relation to explicit methodologic criteria.
When results of primary studies are being presented, research design and population studied are described.
Quantitative data from primary studies are summarized, preferably with confidence intervals or *P* values.
NEGATIVE FEATURES
Addresses a broad or general topic.
Author obviously biased.
No references or a scanty list of references.
Summary statements regarding important issues are merely followed by one or more references (or no references) without further description of the studies or their results.
Magnitude of effect is not discussed.

Does the review focus on a specific question or address a broad, general topic? The author of a review that addresses a specific question (eg, continuous electronic fetal monitoring in low-risk obstetric patients) can – but might not – identify, critically review, summarize, and synthesize all the relevant primary research bearing on that question. This thoroughness is clearly unattainable in a review addressing a general topic (eg, low-risk obstetrics) that could contain within it numerous questions related to etiology, pathophysiology, epidemiology, diagnosis, and treatment. Each question has, of necessity, to be dealt with briefly and usually superficially. (This is not to deny the usefulness of broad overviews as an introduction or orientation to an area of knowledge. However, the specific "answers" such an overview provides should not be accepted uncritically as valid).

Is the author obviously biased? Unrestrained author bias sometimes becomes apparent in the first few sentences of a review. Pass over such articles unless you enjoy polemic writing for its entertainment value.

Are the methods used to conduct the search for relevant original research described? Search methods used to identify original research are rarely described. As a reader, you have no way of knowing (unless the author tells you) whether he or she simply pulled a haphazard accumulation of articles from a file drawer or conducted a systematic, comprehensive search. When search methods are described, you can judge for yourself how likely it is that all (or most) relevant primary studies have been included in the review.

Are references absent or scanty? Lack of references is acceptable under certain circumstances, such as articles describing an established technique or procedure (eg, aspiration of the knee joint) and descriptions of "new" diseases. The main problem with unreferenced or scantily referenced review articles is that you are completely at the mercy of the author. You have no way of verifying the author's statements. You cannot know whether they are based on conventional wisdom, personal clinical experience, "common sense," or scientific evidence. This situation is particularly common in commercial (throwaway) journals.

Are primary studies critically appraised? Critical appraisal preferably should use explicit methodologic criteria. Features other than basic research design have important effects on the validity of the results of primary studies. Explicit methodologic criteria are available for evaluating the scientific validity of various types of studies.[1-8] When authors of review articles assess primary studies against such

criteria, our ability to make rational decisions about the clinical application of the findings is enhanced.

When results of primary studies are being presented, are the research design and population studied described? Research design and subjects preferably should be described in a summary table; summary statements regarding important issues should not merely be followed by one or more citations (or no references) without further description of the studies or their results. "Inactivated whole virus vaccine reduces the risk of influenza in the elderly by 50% to 70%."[2-5] What are we to make of statements such as this? We cannot know (unless the titles offer a clue) whether the studies cited provide strong or weak evidence (were they case control, cohort, or randomized, controlled studies?) or whether the study populations were similar to our patient population (were they healthy elderly or chronically ill, institutionalized elderly?). On the other hand, when such information is presented, we can judge for ourselves the strength of the evidence and how well the results apply to our patients.

Are quantitative data from primary studies summarized? Data preferably should be summarized in a table, and preferably with *P* values, or better still, confidence limits. Nonquantitative statements about an association (eg, "Passive exposure to cigarette smoke is associated with an increased risk of lung cancer") or the effect of an intervention (eg, "Propranolol is useful for benign familial tremor") are of limited clinical usefulness. In order to talk sensibly with our patients and to make rational clinical decisions, we need an estimate of the magnitude of an effect or the strength of an association and an indication of the precision of that estimate, such as a *P* value or, preferably, confidence limits (eg, "Mortality was 10% in the treatment group and 20% in the control group. The 95% confidence interval for the difference in mortality between treated and control subjects was 7% to 15%"). Presentation of data in a table or figure helps to ensure a systematic and uniform approach to summarizing information and facilitates comparisons across studies.

Putting guidelines into practice

These guidelines can be used in a variety of ways in regular journal reading. For example, you could select a cluster of criteria against which to assess potentially relevant reviews. *Table 3* shows how such criteria can be applied. The criteria are arranged in order of the ease with which they can be applied. Failure to meet any one of the criteria disqualifies an article. The first criterion can be applied by examining the title or, at most, the introduction; the second by examining the reference list; and the third by reading the author's introduction. Application of the last two criteria usually requires scanning the tables and text. Additional criteria could be added: for example, a requirement for a description of the methods used to identify the primary studies included in the review. This would greatly enhance the quality and trustworthiness of the reviews you ultimately read but, given the sorry state of the current review literature, would have the effect of excluding all but a few articles on clinically relevant topics.

For literature searches designed to answer questions that arise in clinical practice, the guidelines can be used to identify the best reviews that are (readily) available. Where relevant review articles are plentiful, stringent criteria are appropriate. Where little is available, you may have to loosen your selection criteria (or, alternatively, search for articles reporting high-quality original research).

The advent of "structured abstracts" of review articles, pioneered by the *Annals of Internal Medicine*,[20,21] promises (if their use becomes widespread) to simplify the critical appraisal of review articles. Structured abstracts summarize information about review articles under the following headings: objective(s), data sources, study selection, data extraction, data synthesis, and conclusions. Detailed instructions for preparing structured abstracts have been developed.[20] Structured abstracts are not truncated in the MEDLINE database, even when they exceed the usual 250-word limit.[22] Because structured abstracts highlight the key methodologic issues related to the conduct and reporting of research overviews, improvement in the overall quality of review articles could be expected to follow if structured abstracts are adopted

Table 3. **Deciding whether to read a "relevant" review article:** *The article should be rejected if the following are true.*

- It addresses a broad topic
- It has no references or scanty references
- The author is obviously biased
- It has no description or minimal description of primary studies
- The magnitude of effect is not discussed

widely by journal editors. However, a structured abstract is no more a guarantee of high quality than a methods section. Readers still need to be critical.

Review articles have a vital role in family physicians' efforts to keep up-to-date through regular journal reading and to resolve clinical dilemmas by examining the accumulated health care literature. Critical appraisal of review articles based on criteria of scientific quality is both desirable and feasible. ■

Requests for reprints to: *Dr B.G. Hutchison, First Place Family Medical Centre, 350 King St E, Suite 106, Hamilton, ON L8N 3Y3*

Acknowledgment

I thank Dr Andy Oxman of the Departments of Family Medicine and Clinical Epidemiology and Biostatistics at McMaster University for his significant contributions to this work through discussion of key issues and comments on earlier drafts.

References

1. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. I. Why to read them and how to start reading them critically. *Can Med Assoc J* 1981;124:555-8.
2. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. II. To learn about a diagnostic test. *Can Med Assoc J* 1981;124:703-10.
3. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. III. To learn the clinical course and prognosis of disease. *Can Med Assoc J* 1981;124:869-72.
4. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. IV. To determine etiology or causation. *Can Med Assoc J* 1981;124:985-90.
5. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. V. To distinguish useful from useless or even harmful therapy. *Can Med Assoc J* 1981;124:1156-62.
6. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. VI. To learn about the quality of clinical care. *Can Med Assoc J* 1984;130:377-81.
7. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. VII. To understand an economic evaluation (part A). *Can Med Assoc J* 1984;130:1428-34.
8. Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. VII. To understand an economic evaluation (part B). *Can Med Assoc J* 1984;130:1542-9.
9. Sackett DL, Haynes RB, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine.* Boston: Little Brown and Co, 1985.
10. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials.* 2nd ed. Baltimore: Williams & Wilkins, 1988.
11. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987; 82:498-510.
12. Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
13. Freiman JA, Chalmers TA, Smith H Jr, Kuebler RB. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial. Survey of 71 "negative" trials. In: Bailar JC, Mosteller F, editors. *Medical uses of statistics.* Waltham, Mass: New England Journal of Medicine Books, 1986:289-304.
14. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985;27:335-71.
15. Sacks HR, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analysis of randomized controlled trials. *N Engl J Med* 1987;316:450-5.
16. Williamson JW, Goldschmidt PC, Colton T. The quality of medical literature: an analysis of validation assessments. In: Bailar JC, Mosteller F, editors. *Medical uses of statistics.* Waltham, Mass: New England Journal of Medicine Books, 1986:370-91.
17. Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:485-8.
18. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;138:697-703.
19. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991;44:91-8.
20. Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles. *Ann Intern Med* 1988;108:613-5.
21. Haynes RB, Mulrow CD, Huth EJ, Altman GD, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990;113:69-76.
22. Levy N. Structured abstracts. *NLM Tech Bull* 1988;235:10.

• • •