Special Article

Guidelines for reading literature reviews

Andrew D. Oxman, MD
Gordon H. Guyatt, MD

One strategy for dealing with the burgeoning medical literature is to rely on reviews of the literature. Although this strategy is efficient, readers may be misled if the review does not meet scientific standards. Therefore, guidelines that will help readers assess the scientific quality of the review are proposed. The guidelines focus on the definition of the question, the comprehensiveness of the search strategy, the methods of choosing and assessing the primary studies, and the methods of combining the results and reaching appropriate conclusions. Application of the guidelines will allow clinicians to spend their valuable reading time on high-quality material and to judge the validity of an author's conclusions.

Une façon efficace de se tenir au courant de la littérature médicale toujours plus abondante c'est de se rabattre sur les revues générales. Mais si celles-ci ne se conforment pas aux normes scientifiques, elles risquent d'induire en erreur. Il est proposé ici des lignes directrices afin d'aider le lecteur à apprécier la qualité scientifique d'une revue générale. Elles s'attachent à déterminer si la question y est bien énoncée, la recherche bibliographique est complète, les travaux retenus sont bien choisis et bien analysés, et les divers résultats sont mis en regard de façon à cerner des conclusions valables. En suivant ces lignes directrices le clinicien utilisera son temps précieux à bon escient.

From the departments of Clinical Epidemiology and Biostatistics and of Medicine, McMaster University, Hamilton, Ont.

Dr. Guyatt is a career scientist of the Ontario Ministry of Health.

Reprint requests to: Dr. Gordon H. Guyatt, McMaster University Health Sciences Centre, 3H7–1200 Main St. W, Hamilton, Ont. L8N 3Z5

Clinicians who are attempting to keep abreast of developments must find ways to deal with the exponentially expanding literature. Efficient strategies for finding and storing relevant studies[1-6] and for discarding invalid or inapplicable studies[7-12] are available. However, processing the literature for an answer to a clinical question remains time consuming, and it is not feasible for clinicians to read all the primary literature for each of the myriad clinical issues that confront them daily.

One solution to this problem is the literature review or overview in which the primary research relevant to a clinical question is examined and summarized. However, reviews, as well as primary studies, must be read selectively and critically. Just as flawed methods in a study of diagnosis or therapy may invalidate the results, an unscientific literature review may come to incorrect conclusions. Authors of reviews do collect and analyse data from primary research, although this is sometimes done subjectively and subconsciously. The fundamental difference between a review and a primary study is the unit of analysis, not the scientific principles that apply.

Five conflicting recommendations for managing mild hypertension, quoted from the literature, are shown below.

- The available data . . . lead this reviewer to conclude that treatment of mild hypertension [90 to 104 mm Hg] to achieve diastolic pressures below 90 mm Hg is the appropriate public health policy based on current evidence.[13]
- Most patients with diastolic blood pressure in the 90 to 104 mm Hg range should be treated unless contraindications to drug therapy exist. . . . In certain patients, vigorous dietary and behavioral modifications may be attempted before instituting or as an adjunct to pharmacologic therapy.[14]
- Non-drug measures are often effective for mild hypertension. The initial choice between thiazides and beta-adrenoceptor blocking drugs often depends on the physician's personal preference. . . . With care, the risks

← For prescribing information see page 713

of antihypertensive therapy are considerably less than the benefits.[15]

- The benefits of drug treatment for patients with mild hypertension [diastolic blood pressure between 90 and 105 mm Hg] remain unproven. Non-drug therapy has also been insufficiently investigated.[16]
- At present, therefore, with the diuretic-based treatments principally studied in the previous trials, treatment of mild-to-moderate hypertension [diastolic blood pressure below 115 mm Hg] is of directly demonstrated value only if the stroke rate is high enough (perhaps due to age or cerebrovascular disease) for halving it to justify the costs and trouble of therapy. . . . Lipid-sparing antihypertensives might have more important effects on MI [myocardial infarction] than on stroke. But, in the trials reviewed, the size of the MI reduction remains uncertain [Rory Collins: unpublished observations, 1987].

If one doesn't have some guidelines for assessing the reviews from which these recommendations are taken, deciding which review to believe is like deciding which toothpaste to use. It is a question of taste rather than a question of science.

One does not have to look far to find other examples of important clinical questions for which recent reviews have come to different conclusions: Should clinicians avoid administering corticosteroids because of concern about clinically important osteoporosis?[17,18] What are the benefits to critically ill patients of catheterizing the right side of the heart?[19,20] Should mild hypokalemia be treated aggressively?[21,22]

Clearly, the expertise of the author is not a sufficient criterion of a review's credibility, since experts reviewing the same topic often come to different conclusions. Nor is the prestige of the journal or textbook in which the review is published a sufficient criterion. Recent surveys of the medical literature have found that the scientific quality of most published reviews, including those in the most highly regarded journals, is poor.[23-27]

In this article we present a reader's guide to assessing research reviews. Similar guidelines have been suggested before, particularly in the psychology and social science literature.[28-30] We focus on how readers of the medical literature can decide whether a review is worth reading and whether its conclusions are to be believed. Our guidelines may also be of use to those planning to write a research review.

Guidelines

We have framed our guidelines as a series of questions (Table I). Before we address each item in detail some general comments are warranted. First, the questions are intended to be used to assess overviews of primary studies on pragmatic questions. Second, the term "primary studies" refers to research reports that contain original information on which the review is based. Third, the intention of the guidelines is to encourage efficient use of the medical literature and a healthy scepticism, not to promote nihilism. Readers who apply these guidelines will find that most published reviews have major scientific flaws.[23-27] Indeed, surveys on the scientific adequacy of medical research reports have found that most primary studies also have major scientific flaws.[25]

There is a need for improvement in the design, implementation and reporting of both reviews and primary studies. None the less, vast amounts of valuable information exist, and to make informed decisions clinicians must use the research available. Although most published reviews do not provide strong support for their conclusions, critical readers can discern useful information and make their own inferences, which may or may not be the same as those of the authors.

Were the questions and methods clearly stated?

When examining a review article readers must decide whether the review addresses a question that is relevant to their clinical practice or interests. They therefore require a clear statement of the questions being addressed.

Table I — Guidelines for assessing research reviews

Were the questions and methods clearly stated?
Were comprehensive search methods used to locate relevant studies?
Were explicit methods used to determine which articles to include in the review?
Was the validity of the primary studies assessed?
Was the assessment of the primary studies reproducible and free from bias?
Was variation in the findings of the relevant studies analysed?
Were the findings of the primary studies combined appropriately?
Were the reviewers' conclusions supported by the data cited?

Table II — Examples of the elements of a causal question

Nature of the question	Population	Exposure/ intervention	Outcome
Etiology	Homosexual men	Human immunodeficiency virus	Acquired immune deficiency syndrome
Diagnosis	Patients with head trauma	Computerized tomography	Hemorrhage
Prognosis	Patients with ulcerative colitis	Ulcerative colitis	Cancer of the colon
Therapy	Patients with Alzheimer's disease	Cholinomimetic agents	Functional status
Prevention	Postmenopausal women	Calcium supplementation	Hip fracture

Any causal question has three key elements: the population, the exposure or intervention and the outcome. Examples of these elements in five key areas of clinical inquiry are presented in Table II. A clear statement of the question requires explicit specification of all three elements if the reader is to quickly decide whether the review is relevant. If there is no clear statement of the questions being addressed at the beginning of the review the reader might as well stop. Fuzzy questions tend to lead to fuzzy answers.

Many reviews address several questions; for example, an article or a chapter in a textbook about acquired immune deficiency syndrome may review what is known about the cause, diagnosis, prognosis, treatment and prevention of the disease. Such reviews may be extremely helpful for readers seeking a broad overview. However, they tend to provide little, if any, support for most of the inferences they make. Typically, an inference is presented as a fact followed by one or more citations. In this case the reader has no basis upon which to judge the strength or validity of the inferences without reading the articles that are cited. Readers seeking answers to specific clinical questions should not rely on reviews that address broad topics and encompass many questions.

In addition, an explicit statement of the methods used for the research review is necessary for the reader to make an informed assessment of the scientific rigour of the review and the strength of the support for the review's inferences. Unfortunately, this information is often lacking. In general, when a review does not state how something was done — for example, how it was decided which primary studies would be included — it is reasonable to assume that it was not done rigorously and that a threat to the validity of the review exists. Readers looking for answers to specific clinical questions should seek reviews that clearly report the methods used. Without knowing the authors' methods the reader cannot distinguish statements based on evidence from those based on the opinions of the authors.

Were comprehensive search methods used to locate relevant studies?

It is surprisingly difficult to locate all the published research in a particular area, even when the area is relatively circumscribed.[31-33] For example, Dickersin and associates[33] found that a MEDLINE search yielded only 29% of the relevant trials on the prevention and treatment of perinatal hyperbilirubinemia.

This problem is exacerbated by the fact that some of the relevant material may not even be published. Furthermore, the unpublished studies may be systematically different from those that have appeared in peer-reviewed journals, not in that their methods are flawed but in that their results are "negative". Research has suggested that of two articles that use the same methods to investigate a question the study yielding positive results is more likely to be published than the one yielding negative results.[33-37] Research conducted by an agency that has an investment in the treatment being studied (such as a pharmaceutical company with a new drug) may not even be submitted for publication if its results are negative. It thus behoves an author to try to determine the extent of the "publication bias" in the area being reviewed.

Authors' search strategies vary widely, and experts are no more likely than nonexperts to be systematic in their search.[38] The more selective or haphazard the authors' search for papers the more likely it is that there will be bias in the review. For example, authors are likely to attend to papers that support their preconceptions.

The reader needs assurance that all the pertinent and important literature has been included in the review. The more comprehensive the authors' search the more likely it is that all the important articles have been found. The reader should look for an explicit statement of the search strategies used. Ideally, such strategies include the use of one or more bibliographic databases (including a specification of the key words and other aspects of the search strategies[39]), a search for reports that cite the important papers found through a database such as the Science Citation Index, perusal of the references of all the relevant papers found and personal communication with investigators or organizations active in the area being reviewed (to make sure important published papers have not been missed and particularly to look for methodologically adequate studies that have not been published).

Were explicit methods used to determine which articles to include in the review?

A comprehensive literature search will yield many articles that may not be directly relevant to the question under investigation or that may be so methodologically weak that they do not contribute valid information. The authors must therefore select those that are appropriate for inclusion in the review. When, as is often the case, this process is unsystematic, opportunities for bias develop. Thus, it is common to find two reviews of the same question in which different primary studies are included and for the choice of studies to contribute to different conclusions. For example, in two methodologically sophisticated and carefully conducted reviews on whether corticosteroids are associated with peptic ulcer the two teams of authors used different criteria for choosing which studies would be included in the review.[40,41] This difference was the main reason for the remarkable result of the two reviews: diametrically opposed conclusions about whether or not the association exists.

The authors should specify how the articles were chosen by referring to the three basic ele-

ments of primary studies: the population, the exposure or intervention and the outcome. For example, in assessing the effect of cholinomimetic agents in patients with dementia the authors could specify the criteria as follows.

- Population: patients with senile dementia in whom causes other than Alzheimer's disease were excluded.
- Intervention: oral administration of cholinomimetic agents.
- Outcome: indicated by measurements of both memory and functional status.

Other methodologic criteria may be used to select primary papers for review. In this example the authors may consider only studies in which patients were selected at random to receive the treatment drug or a placebo and in which both the investigator and the patient were blind to allocation.

Was the validity of the primary studies assessed?

Authors will come to correct conclusions only if they accurately assess the validity of the primary studies on which the review is based. If all the studies have basic flaws their conclusions may be questionable even if their results are comparable. For example, if the literature on extracranial-intracranial bypass surgery for threatened stroke were reviewed before the results of a recent randomized controlled trial[42] were published, a large number of studies with positive results but of suboptimal design and thus open to bias would have been found. The appropriate conclusion would have been that the procedure's effectiveness was still open to question, despite the volume of studies with positive results; indeed, the subsequent trial showed no benefit of surgical over medical therapy.

Methodologic guidelines for studies of etiology,[10,43] diagnosis,[8] prognosis[9] and therapy[11,44] are available. In a study of therapy one is interested in whether the allocation to treatment was random, whether the subjects and investigators were blind to the allocation, and whether all the relevant outcomes were monitored. Important aspects of the design and conduct of each primary study should be critiqued and the standard used in these critiques made explicit. Critiques should be reported in sufficient detail to allow readers to judge the methodologic quality of the primary studies. Although a study-by-study critique can be tedious, presentation of the methodologic assessment in a table may allow a rapid assessment of validity. Readers should be wary of any review that focuses on the results of studies without thoroughly discussing the methods that were used to arrive at the results.

When information about the methods or results has been omitted from a published report the authors of a review can contact the writers of the report to obtain the missing information. A review is strengthened if the authors have discussed the implications of missing information and have attempted to collect the relevant data.

Was the assessment of the primary studies reproducible and free from bias?

Expert assessment of primary research studies generally results in a level of disagreement that is both extraordinary and distressing. For example, correlations measuring agreement about the decision to publish or not publish primary research studies are almost always less than 0.5 and average about 0.3,[28,45,46] a level not much higher than one would expect to achieve by chance.

Not only do assessments lack reproducibility, but also they are often biased. In one study Peters and Ceci[47] resubmitted previously published articles from respected institutions after they substituted the names of the authors and the institutions with fictitious names. Mahoney[35] submitted an article to different referees, varying the results without altering the methods. These studies found that the articles that came from respected institutions and reported positive results were more readily accepted. Furthermore, in Peters and Ceci's study many of the articles were rejected because of "serious methodological flaws", and in Mahoney's study the article was judged as having weaker methods when it described negative results.

It is even possible for authors to disagree on the results of a study. Numerous conflicting reviews have been reported in which an author who favoured a particular treatment classified the primary study as positive, whereas an author who did not favour the treatment classified the study as negative. For example, Miller[48] found five reviews that compared drug therapy plus psychotherapy with drug therapy alone for psychiatric patients. Of the 11 studies cited in two or more of the reviews the results of 6 were interpreted as positive in at least one review and as negative in at least one other.

Problems with reproducibility and bias can affect two stages of the review process: the decision about which papers to include and judgement of the quality of the papers included. Such problems can be minimized if explicit criteria are used. However, many of the criteria will require considerable judgement of the author of a review. In an example we used earlier one of the criteria for inclusion in a review of treatment with cholinomimetic agents for Alzheimer's disease was a definition of the population as patients with senile dementia in whom causes other than Alzheimer's disease were excluded. Is a statement in the text such as "standard methods for diagnosing Alzheimer's disease were used" adequate or does one require details of how other causes of dementia were ruled out?

Explicit criteria offer little advantage if they cannot be reproduced by other authors. Ideally, all

the potential primary studies should be assessed for inclusion by at least two authors, each blind to the other's decision, and the extent of agreement should be recorded. Reproducibility should be quantified with a statistical measure that quantitates agreement above and beyond that which would have occurred by chance, such as an intraclass correlation coefficient[49] or a κ statistic.[50] A similar process should be used to assess the reproducibility of the criteria used to determine the validity of the primary studies.

Even if the criteria for study inclusion or validity can be reproduced there is no guarantee that bias has not intruded. For example, if the authors believe that a new treatment works they may apply inclusion criteria by which studies with negative results are systematically excluded; the validity of such studies that are included may be judged more harshly. What can be done to prevent this sort of bias?

In randomized controlled trials bias is avoided if both the patients and the clinicians are blind to whether the patients are taking the active drug or a placebo. In an assessment of primary studies the major possible sources of bias are related to the authors, their institution and the results. However, one can assess the content and quality of a study through its methods without knowing this information; the relevant sections of the paper can simply be "whited out" so that the reviewers are blind to the authors' institutions and results. Decisions about study inclusion and validity ideally should be made under these conditions. This added precaution will strengthen the review.

Was variation in the findings of the relevant studies analysed?

Authors of reviews are certain to encounter variability in the results of studies addressing the question of interest. Indeed, if all the results of primary research were the same a review article would probably not be necessary. It is the authors' task to try to explain this variability.

Possible sources of variability are the study design, chance and differences in the three basic study components (the population, the exposure or intervention and the outcome).[51] If randomized controlled trials, before-and-after studies and studies with historical controls are all included in a review, and if the randomized controlled trials consistently show results that differ systematically from those of the other studies, the study design probably explains the differences. For example, Sacks and colleagues[52] found that randomized controlled trials consistently show smaller effects than studies that use historical controls.

A second explanation for differences in study results is chance. Even if two investigations use comparable methods and the true size of the effects is identical the play of chance will lead to apparent differences in the size. If the samples are small, chance alone may lead to apparently large differences in the size of the effects. Some trials of acetylsalicylic acid (ASA) in patients with transient ischemic attacks have shown a trend in favour of a placebo, whereas others have shown reductions in risk of up to 50% with ASA.[53] However, the confidence intervals, which represent the upper and lower limits of the size of the effects consistent with the observed results, overlap. Thus, although the apparently discrepant results might suggest hypotheses for testing in subsequent studies, they are all consistent with a reduction in risk of between 15% and 30% with ASA.

In other instances differences in study results may be so large that they cannot be explained by chance. The authors must therefore look to differences in the population, exposure or intervention and outcome. In our example of cholinomimetic agents in patients with Alzheimer's disease the studies with negative results may have included a larger number of severely affected patients than the studies with positive results. One might then assume that the intervention works only in mildly affected patients. However, the intervention may have differed — that is, higher doses or different agents may have been given in the studies with positive results. Finally, the tests used to determine memory and functional status may have been different; some tests are more responsive to changes in patient status. Horwitz[51] has documented many ways in which differences in the methods of randomized controlled trials can lead to differing results.

Readers of a review should be alert to whether these five explanations for differing study results have been considered and should be sceptical when differences are attributed to one explanation without adequate consideration of the others.

Were the findings of the primary studies combined appropriately?

Meta-analysis (the use of several statistical techniques to combine the results of different studies) is becoming increasingly popular, especially as a method of combining results from randomized controlled trials. However, it remains controversial, and clinical readers cannot be expected to judge the merits of a particular statistical technique used by the authors of a meta-analysis. Nevertheless, there are issues that clinical readers can address.

The crudest form of meta-analysis, in which the number of studies with positive results is compared with the number of those with negative results, is not satisfactory. This "vote count" ignores the size of the treatment effects and the sample sizes of each study. The most satisfactory meta-analysis yields two pieces of information: the magnitude of the overall treatment effect and the likelihood that this effect would have occurred by chance if the true effect were zero. The former may

be expressed as a percentage risk reduction, the latter as a p value.

The primary advantage of meta-analysis is that the results of different studies can be combined accurately and reliably to determine the best estimate of the average magnitude of the effects of the exposure or intervention of interest. Before the results are combined, however, one should consider whether it is appropriate to aggregate across the studies. Study designs, or the three basic study elements, may differ sufficiently that a statistical combination of the results does not make sense. Meta-analysis can be used to analyse the variation in study results to generate or test hypotheses about the source of the differences. However, it is on strongest ground when the methods of the primary studies are similar and the differences in the study results can be explained by chance.

Reviews in which the results are not statistically combined should state explicitly the basis for the conclusions and should attempt to explain the conflicting results. Readers should beware of reviews that conclude that there is no effect without having considered the studies' power to detect a clinically important effect. When several studies do not show a significant difference there is a tendency for reviewers who have not used meta-analysis to conclude that there is no effect even when statistical aggregation demonstrates otherwise. Cooper and Rosenthal[54] demonstrated this experimentally by assigning reviewers at random to either use or not use meta-analysis to combine the results of several studies, including some that did not show significant results. Another investigator made the same observation when he polled researchers who had conducted trials of tamoxifen citrate as adjuvant therapy for breast cancer (Rory Collins: personal communication, 1987). Most of the researchers concluded from the available information that tamoxifen did not produce a longer disease-free interval; however, statistical aggregation of all the available results demonstrated a clinically important, statistically significant effect.

It is important to remember that all the other guidelines we have discussed still apply whether or not the authors of a review have used meta-analysis.

Table III — Guidelines for assessing the strength of a causal inference

Is the temporal relation correct? (A positive answer is necessary, but it does not, in itself, confer strength on the inference.)
Is the evidence strong?
Is the association strong?
Is there consistency between studies?
Is there a dose–response relation?
Is there indirect evidence that supports the inference — that is, evidence relating to intermediate outcomes, evidence from studies of different populations (including animals) and evidence from analogous relations (i.e., related exposures or interventions)?
Have the plausible competing hypotheses been ruled out?

Were the reviewers' conclusions supported by the data cited?

Whether or not authors have used meta-analysis, the results of individual primary studies should be reported in sufficient detail that readers are able to critically assess the basis for the authors' conclusions. The method of presenting individual study summaries will depend on the question addressed. For questions of treatment effectiveness and prevention the size of the effects and its confidence interval give the key information. Reviews of diagnostic tests may provide sensitivities, specificities and likelihood ratios (and their confidence intervals).[8] Survival curves may efficiently depict the main results of studies of prognosis.

With questions of etiology and causation for which randomized controlled trials are not available the authors can evaluate the evidence with criteria for causal inference. Variations of these criteria have been presented by several investigators,[10,44,55,56] but common ingredients include the size and consistency of the association between the causal agent and the outcome and the necessity for demonstrating the appropriate temporal relation. Our version of these criteria is presented in Table III. The authors' comments on each of these criteria should, of course, refer directly back to the data in the primary studies cited.

Conclusion

A literature review is a scientific endeavour, and, as with other scientific endeavours, standards are available for conducting the review in such a way that valid conclusions are reached. Just as readers of the clinical literature who are unable to critically appraise the methods of primary studies may arrive at incorrect conclusions, readers who are unable to assess the scientific quality of a review are apt to be misled. We have offered eight guidelines for readers interested in answering a clinical question relevant to their everyday practice. Application of these guidelines will allow readers to quickly discard review articles that are irrelevant or scientifically unsound, to detect potential sources of bias and to be confident of conclusions made from a systematic evaluation of the available research.

We thank Drs. Geoff Norman, David Streiner, David L. Sackett and Brian Hutchison, and Professor Mike Gent for their help in developing the guidelines.

This work was supported in part by the Ontario Ministry of Health.

References

1. Haynes RB, McKibbon KA, Fitzgerald D et al: How to keep up with the medical literature: I. Why try to keep up and how to get started. *Ann Intern Med* 1986; 105: 149–153

2. Idem: How to keep up with the medical literature: II. Deciding which journals to read regularly. Ibid: 309-312
3. Idem: How to keep up with the medical literature: III. Expanding the number of journals you read regularly. Ibid: 474-478
4. Idem: How to keep up with the medical literature: IV. Using the literature to solve clinical problems. Ibid: 636-640
5. Idem: How to keep up with the medical literature: V. Access by personal computer to the medical literature. Ibid: 810-824
6. Idem: How to keep up with the medical literature: VI. How to store and retrieve articles worth keeping. Ibid: 978-984
7. 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: 555-558
8. Idem: How to read clinical journals: II. To learn about a diagnostic test. Ibid: 703-710
9. Idem: How to read clinical journals: III. To learn the clinical course and prognosis of disease. Ibid: 869-872
10. Idem: How to read clinical journals: IV. To determine etiology or causation. Ibid: 985-990
11. Idem: How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. Ibid: 1156-1162
12. Idem: How to read clinical journals: VI. To learn about the quality of clinical care. *Can Med Assoc J* 1984; 130: 377-382
13. Labarthe DR: Mild hypertension: the question of treatment. *Ann Rev Public Health* 1986; 7: 193-215
14. Haber E, Slater EE: High blood pressure. In Rubenstein M, Federman DD (eds): *Scientific American Medicine*, 9th ed, Sci Am, New York, 1986: sect 1, VII1-VII29
15. Risks of antihypertensive therapy [E]. *Lancet* 1986; 2: 1075-1076
16. Sacks HS, Chalmers TC, Berk AA et al: Should mild hypertension be treated? An attempted meta-analysis of the clinical trials. *Mt Sinai J Med* 1985; 52: 265-270
17. Guyatt GH, Webber CE, Mewa AA et al: Determining causation — a case study: adrenocorticosteroids and osteoporosis. *J Chronic Dis* 1984; 37: 343-352
18. Baylink DJ: Glucocorticoid-induced osteoporosis. *N Engl J Med* 1983; 309: 306-308
19. Swan HJC, Ganz W: Hemodynamic measurements in clinical practice: a decade in review. *J Am Coll Cardiol* 1983; 1: 103-113
20. Robin ED: The cult of the Swan-Ganz catheter: overuse and abuse of pulmonary flow catheters. *Ann Intern Med* 1985; 103: 445-449
21. Harrington JT, Isner JM, Kassirer JP: Our national obsession with potassium. *Am J Med* 1982; 73: 155-159
22. Kaplan NM: Our appropriate concern about hypokalemia. *Am J Med* 1984; 77: 1-4
23. Mulrow CD: The medical review article: state of the science. *Ann Intern Med* 1987; 106: 485-488
24. Sacks HS, Berrier J, Reitman D et al: Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450-455
25. Williamson JW, Goldschmidt PG, Colton T: The quality of medical literature. An analysis of validation assessments. In Bailar JC III, Mosteller F (eds): *Medical Uses of Statistics*, NEJM Bks, Waltham, Mass, 1986: 370-391
26. Halvorsen KT: Combining results from independent investigations: meta-analysis in medical research. Ibid: 392-416
27. Oxman AD: *A Methodological Framework for Research Overviews*, MSc thesis, McMaster U, Hamilton, Ont, 1987: 23-31, 98-105
28. Light RJ, Pillemer DB: *Summing Up: the Science of Reviewing Research*, Harvard U Pr, Cambridge, Mass, 1984
29. Jackson GB: Methods for integrative reviews. *Rev Educ Res* 1980; 50: 438-460
30. Cooper HM: *The Integrative Research Review: a Systematic Approach*, Sage, Beverly Hills, Calif, 1984
31. Glass GV, McGaw B, Smith ML: *Meta-Analysis in Social Research*, Sage, Beverly Hills, Calif, 1981
32. Poynard T, Conn HO: The retrieval of randomized clinical trials in liver disease from the medical literature. *Controlled Clin Trials* 1985; 6: 271-279
33. Dickersin K, Hewitt P, Mutch L et al: Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. Ibid: 306-317
34. Simes RJ: Publication bias. The case for an international registry of clinical trials. *J Clin Oncol* 1986; 4: 1529-1541
35. Mahoney MJ: Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* 1977; 1: 161-175
36. Devine EC, Cook TD: Effects of psycho-educational intervention on length of hospital stay: a meta-analytic review of 34 studies. In Light RJ (ed): *Evaluation Studies Review Annual*, 8th ed, Sage, Beverly Hills, Calif, 1983: 417-432
37. Simes RJ: Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987; 6: 11-29
38. Cooper HM: Literature searching strategies of integrative research reviewers: a first survey. *Knowledge* 1986; 8: 372-383
39. Huth EJ: Needed: review articles with more scientific rigor. *Ann Intern Med* 1987; 106: 470-471
40. Messer J, Reitman D, Sacks HS et al: Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med* 1983; 309: 21-24
41. Conn HO, Blitzer BL: Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N Engl J Med* 1976; 294: 473-479
42. EC/IC Bypass Study Group: Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. *N Engl J Med* 1985; 313: 1191-1200
43. Hill AB: *Principles of Medical Statistics*, 9th ed, *Lancet*, London, 1971: 312-320
44. Chalmers TC, Smith H, Blackburn B et al: A method for assessing the quality of a randomized controlled trial. *Controlled Clin Trials* 1981; 2: 31-49
45. Bailar JC III, Patterson K: Journal peer review: the need for a research agenda. In Bailar JC III, Mosteller F (eds): *Medical Uses of Statistics*, NEJM Bks, Waltham, Mass, 1986: 349-369
46. Marsh HW, Ball S: Interjudgmental reliability of reviews for the *Journal of Educational Psychology*. *J Educ Psychol* 1981; 73: 872-880
47. Peters DP, Ceci SJ: Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav Brain Sci* 1982; 5: 187-255
48. Miller TI: *The Effects of Drug Therapy on Psychological Disorders*, PhD dissertation, U of Colorado, Boulder, 1977
49. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428
50. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37-46
51. Horwitz RI: Complexity and contradiction in clinical trial research. *Am J Med* 1987; 82: 498-510
52. Sacks H, Chalmers TC, Smith H: Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72: 233-240
53. Sze PC, Pincus M, Sacks HS et al: Antiplatelet agents in secondary stroke prevention. A meta-analysis of the available randomized control trials [abstr]. *Clin Res* 1986; 34: 385A
54. Cooper HM, Rosenthal R: Statistical versus traditional procedures for summarizing research findings. *Psychol Bull* 1980; 87: 442-449
55. Susser M: Reviews and commentary: the logic of Sir Karl Popper and the practice of epidemiology. *Am J Epidemiol* 1986; 124: 711-718
56. Guyatt GH, Newhouse MT: Are active and passive smoking harmful? Determining causation. *Chest* 1985; 88: 445-451