The Medical Literature

Users' Guides to the Medical Literature

VI. How to Use an Overview

Andrew D. Oxman, MD, MSc; Deborah J. Cook, MD, MSc; Gordon H. Guyatt, MD, MSc;
for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

A 55-year-old man had his serum cholesterol level measured at a shopping mall 2 months ago. His cholesterol level was elevated and he comes to you, his primary care physician, for advice. He does not smoke, is not obese, and does not have hypertension, diabetes mellitus, or any first-order relatives with premature coronary heart disease (CHD). You repeat his cholesterol test and schedule a follow-up appointment. The test confirms an elevated cholesterol level (7.9 mmol/L [305 mg/dL]), but before deciding on a treatment recommendation, you elect to find out just how big a reduction in the risk of CHD this patient could expect from a cholesterol-lowering diet or drug therapy.

THE SEARCH

There are a number of cholesterol-lowering trials, and instead of trying to find and review all of the original studies yourself, you use Grateful Med to find a recent overview. On the first subject line you select hypercholesterolemia or cholesterol from the list of Medical Subject Headings (MeSH) used to index articles. On the second subject line you use the MeSH term coronary disease, which you explode to capture articles that are indexed with more specific terms that come under coronary disease, such as myocardial infarction. You limit your search to English-language articles, and to find a quantitative review, you use the term meta-analysis on the line for publication type. Titles and abstracts suggest two of the nine references from this search are definitely on target, and you decide to examine both.[1,2]

INTRODUCTION

Systematic overviews of the medical literature that summarize scientific evidence (in contrast to unsystematic narrative reviews that mix together opinions and evidence) are becoming increasingly prevalent. These overviews address questions of treatment, causation, diagnosis, or prognosis. In each case, the rules for deciding whether the overviews are credible, and for interpreting their results, are similar. In this article, we provide guidelines for distinguishing a good overview from a bad one and for using the results. In doing so, we will ask the same key questions that we have suggested for original reports of research[3]: Are the results valid? If they are, what are the results, and will they be helpful in my patient care (Table 1)?

Authors sometimes use the terms "systematic review," "overview," and "meta-analysis" interchangeably. We use overview as a term for any summary of the medical literature and meta-analysis as a term for reviews that use quantitative methods to summarize the results. Investigators must make a host of decisions in preparing an overview, including determining the focus; identifying, selecting, and critically appraising the relevant studies (which we will call the "primary studies"); collecting and synthesizing (either quantitatively or nonquantitatively) the relevant information; and drawing conclusions. Avoiding errors in both meta-analyses and other overviews requires a systematic approach, and enabling users to assess the validity of an overview's results requires explicit reporting of the methods. A number of authors have recently examined issues pertaining to the validity of overviews.[4-7] In this article we will emphasize key points from the perspective of a clinician needing to make a decision about patient care.

You can use the first two validity guides in Table 1 to quickly screen out most published review articles.[7] The discrepancies between the results of systematic meta-analyses and the recommendations made by clinical experts in nonsystematic review articles[8] reflects the limited validity of most published review articles. Archie Cochrane pointed out the need for more systematic overviews when he wrote: "It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials [RCTs]."[9] The Cochrane Collaboration, an international effort to prepare, maintain, and disseminate systematic reviews of the effects of health care, has evolved in response to this challenge.[10,11] As the Collaboration develops, you will find more and more systematic reviews of RCTs addressing important issues in patient management.

ARE THE RESULTS OF THE OVERVIEW VALID?

Primary Guides

Did the Overview Address a Focused Clinical Question?—Unless an overview clearly states the question it addresses, you can only guess whether it is pertinent to your patient care. Most clinical questions can be formulated in terms of a simple relationship between the patient, some exposure (to a treatment, a diagnostic test, a potentially harmful agent, and the like), and one or more outcomes of interest. If the main question that an overview addresses is not clear from the title or abstract, it is probably a good idea to move on to the next article.

From the Departments of Clinical Epidemiology and Biostatistics (Drs Oxman, Cook, and Guyatt), Medicine (Drs Cook and Guyatt), and Family Medicine (Dr Oxman), McMaster University, Hamilton, Ontario.

A complete list of members (with affiliations) of the Evidence-Based Medicine Working Group appears in the first article of this series (*JAMA*. 1993;270:2093-2095). The following members contributed to this article: Eric Bass, MD, MPH; Patrick Brill-Edwards, MD; George Browman, MD, MSc; Allan Detsky, MD, PhD; Michael Farkouh, MD; Hertzel Gerstein, MD, MSc; Ted Haines, MD, MSc; Brian Haynes, MD, MSc; Robert Hayward, MD, MPH; Anne Holbrook, MD, PharmD, MSc; Roman Jaeschke, MD, MSc; Elizabeth Juniper, MCSP, MSc; Andreas Laupacis, MD, MSc; Hui Lee, MD, MSc; Mitchell Levine, MD, MSc; Virginia Moyer, MD, MPH; David Naylor, MD, DPhil; Jim Nishikawa, MD; Ameen Patel, MD; John Philbrick, MD; Scott Richardson, MD; Stephane Sauve, MD, MSc; David Sackett, MD, MSc; Jack Sinclair, MD; Brian Strom, MD, MPH; K. S. Trout, FRCE; Sean Tunis, MD, MSc; Stephen Walter, PhD; John Williams, Jr, MD, MHS; and Mark Wilson, MD, MPH.

Reprint requests to Room 2C12, McMaster University Health Sciences Centre, 1200 Main St W, Hamilton, Ontario, Canada L8N 3Z5 (Dr Guyatt).

Users' Guides to the Medical Literature section editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

Table 1.—Users' Guides for How to Use Review Articles

Are the results of the study valid?
Primary guides:
Did the overview address a focused clinical question?
Were the criteria used to select articles for inclusion appropriate?
Secondary guides:
Is it unlikely that important, relevant studies were missed?
Was the validity of the included studies appraised?
Were assessments of studies reproducible?
Were the results similar from study to study?
What are the results?
What are the overall results of the review?
How precise were the results?
Will the results help me in caring for my patients?
Can the results be applied to my patient care?
Were all clinically important outcomes considered?
Are the benefits worth the harms and costs?

Table 2.—Guides for Selecting Articles That Are Most Likely to Provide Valid Results*

Therapy	• Was the assignment of patients to treatments randomized? • Were all of the patients who entered the trial properly accounted for and attributed at its conclusion?
Diagnosis	• Was there an independent, blind comparison with a reference standard? • Did the patient sample include an appropriate spectrum of the sort of patients to whom the diagnostic test will be applied in clinical practice?
Harm	• Were there clearly identified comparison groups that were similar with respect to important determinants of outcome, other than the one of interest? • Were outcomes and exposures measured in the same way in the groups being compared?
Prognosis	• Was there a representative and well-defined sample of patients at a similar point in the course of disease? • Was follow-up sufficiently long and complete?

*From Oxman et al.[3]

Many overviews address a number of questions. For example, a review article or a chapter from a textbook might include sections on the etiology, diagnosis, prognosis, treatment, and prevention of asthma. While such broad reviews can provide a useful introduction to an area, they usually offer limited support for their conclusions. Typically, you will find only a declarative statement followed by one or more citations. You must then study the references in order to judge the validity of the authors' conclusions.

Were the Criteria Used to Select Articles for Inclusion Appropriate?—To determine if the investigators reviewed the appropriate research, the reader needs to know the criteria they used to select research. These criteria should specify the patients, exposures, and outcomes of interest. They should also specify the methodologic standards used to select studies, and these standards should be similar to the primary validity criteria we have described for original reports of research[3] (Table 2).

In looking at the effectiveness of lowering cholesterol on CHD, investigators might restrict themselves to studies of patients who did not have clinically manifest CHD at the beginning of the study (primary prevention), to studies of patients who already had symptomatic CHD (secondary prevention), or include both. They might include only trials of diet therapy, only trials of drug therapy, or both. They might consider several different outcomes, such as nonfatal CHD, CHD mortality, and total mortality. With respect to methodologic criteria, they might consider only RCTs or include observational studies.

Differences in the patients, exposures, and outcomes can lead to different results among overviews that appear to address the same clinical question.[12] The clinician must be sure the criteria used to select the studies correspond to the clinical question that led her to the article in the first place. The impact of cholesterol-lowering strategies, for instance, differs in studies of primary vs secondary prevention.[1,2]

If the authors state their inclusion criteria, it is less likely they will (as they are wont to do) preferentially cite studies that support their own prior conclusion. Bias in choosing articles to cite is a problem for both overviews and original reports of research (in which the discussion section often includes comparisons with the results of other studies). Gøtzsche, for example, reviewed citations in reports of trials of new nonsteroidal anti-inflammatory drugs in rheumatoid arthritis.[13] Among 77 articles where the authors could have referenced other trials with and without outcomes favoring the new drug, nearly 60% (44) cited a higher proportion of the trials with favorable outcomes. In 22 reports of controlled trials of cholesterol lowering, Ravnskov[14] found a similar bias toward citing positive studies.

Secondary Guides

Is It Unlikely That Important Relevant Studies Were Missed?—It is important that authors conduct a thorough search for studies that meet their inclusion criteria. This should include the use of bibliographic databases, such as MEDLINE and EMBASE, checking the reference lists of the articles they retrieved, and personal contact with experts in the area. Unless the authors tell us what they did to locate relevant studies, it is difficult to know how likely it is that relevant studies were missed.

There are two important reasons why a review's authors should use personal contacts. The first is so they can identify published studies that might have been missed (including studies that are in press or not yet indexed or referenced). The second is so they can identify unpublished studies. Although the inclusion of unpublished studies is controversial,[15] their omission increases the chances of "publication bias"—a higher likelihood for studies with positive results to be published[16-19] and the attendant risk for the review to overestimate efficacy or adverse effects.

If investigators include unpublished studies in an overview, they should obtain full written reports and appraise the validity of both published and unpublished studies; they may also use statistical techniques to explore the possibility of publication bias.[20] Overviews based on a small number of small studies with weakly positive effects are the most susceptible to publication bias.

Was the Validity of the Included Studies Appraised?—Even if a review article includes only RCTs, it is important to know whether they were of good quality. Unfortunately, peer review does not guarantee the validity of published research.[21] For exactly the same reason that the guides for using original reports of research begin by asking if the results are valid, it is essential to consider the validity of research included in overviews.

Differences in study methods might explain important differences among the results.[22,23] For example, less rigorous studies tend to overestimate the effectiveness of therapeutic and preventive interventions.[24] Even if the results of different studies are consistent, it is still important to know how valid the studies are. Consistent results are less compelling if they come from weak studies than if they come from strong studies.

There is no one correct way to assess validity. Some investigators use long checklists to evaluate methodologic quality, while others focus on three or four key aspects of the study. You will remember that in our previous articles about therapy, diagnosis, and prognosis in the Users' Guides series, we asked the question, "Is the study valid?" and presented criteria to help you answer these questions. When considering whether to believe the results of an overview, you should check whether the authors examined criteria similar to those we have presented in deciding on the credibility of their primary studies (Table 2).

Were Assessments of Studies Reproducible?—As we have seen, authors of review articles must decide which studies to include, how valid they are, and which data to extract from them. Each of these decisions requires judgment by

Table 3.—Assessments of Overviews From the Clinical Scenario*

Criterion	Davey Smith et al,[1] 1993	Silberberg and Henry,[2] 1991
Are the results of the study valid?		
Did the overview address a focused clinical question?	Yes: to examine effects of cholesterol lowering on mortality in relationship to baseline risk of CHD death	Yes: to examine effects of drug treatment to lower cholesterol in primary and secondary prevention of CHD events
Were the criteria used to select articles for inclusion appropriate?	Yes, although inclusion of trials of estrogen and surgery can be questioned: single-factor (dietary interventions, lipid-lowering drugs [including estrogen] or surgery) RCTs of cholesterol lowering with ≥6 mo follow-up and at least 1 death—35 trials, 57 124 patients	Yes, although exclusion of nondrug trials could be questioned: single-factor RCTs of drug treatments (excluding trials of estrogen and thyroxine)—9 trials, 26 609 patients
Is it unlikely that important relevant studies were missed?	Yes: MEDLINE, previous overviews, and personal contact with investigators were used to identify studies	Can't tell: MEDLINE and previous overviews were used to identify studies; investigators were not contacted, non–English-language publications and unpublished data were not included
Was the validity of the included studies appraised?	No	No
Were assessments of studies reproducible?	Can't tell	Yes: data were extracted independently by two reviewers
Were the results similar from study to study?	Probably not (test of homogeneity not reported): baseline risk of CHD death and percent reduction in cholesterol levels hypothesized as explanation for variation in effect of treatment	Probably not (test of homogeneity not reported), but pooled ORs for primary and secondary prevention studies respectively were similar: baseline risk hypothesized as explanation for variation in absolute risk reduction
What are the results?		
What are the overall results of the review? How precise were the results?	For total mortality, the OR (and 95% CI) was 0.74 (0.60-0.92 for high-risk groups [>50 deaths/1000 person-years in the control group]), 0.96 (0.84-1.09) for medium-risk groups (10-50 deaths/1000 person-years), and 1.22 (1.06-1.42) in low-risk groups (<10 deaths/1000 person-years)	For CHD death, the OR (and 95% CI) was 0.85 (0.64-1.14) in primary prevention and 0.84 (0.75-0.95) in secondary prevention studies; the NNT to prevent one death from CHD was 675 and 38 in the primary and secondary trials, respectively

*CHD indicates coronary heart disease; RCTs, randomized controlled trials; OR, odds ratio; CI, confidence interval; and NNT, number needed to treat.

the reviewers and each is subject to both mistakes (random errors) and bias (systematic errors). Having two or more people participate in each decision guards against errors, and if there is good agreement among the reviewers, the clinician can have more confidence in the results of the overview.

Were the Results Similar From Study to Study?—Despite restrictive inclusion criteria, most systematic overviews document important differences in patients, exposures, outcome measures, and research methods from study to study. Readers must decide when these factors are so different that it no longer makes sense to combine the study results.

One criterion for deciding to combine results quantitatively is whether the studies seem to be measuring the same underlying magnitude of effect. In meta-analyses, investigators can test the extent to which differences among the results of individual studies are greater than you would expect if all studies were measuring the same underlying effect and the observed differences were due only to chance. The statistical analyses that are used to do this are called "tests of homogeneity."

The more significant the test of homogeneity, the less likely it is that the observed differences in the size of the effect are due to chance alone. Both the "average" effect and the confidence interval (CI) around the average effect need to be interpreted cautiously when there is "statistically significant" heterogeneity (a low probability of the differences in results from study to study being due to chance alone, indicating that differences in patients, exposures, outcomes, or study design are responsible for the varying treatment effect).

Unfortunately, a nonsignificant test does not necessarily rule out important heterogeneity. Hence, clinically important differences between study results still dictate caution in interpreting the overall findings, despite a nonsignificant test of homogeneity. However, even when there are large differences between the results of different studies, a summary measure from all of the best available studies may provide the best estimate of the impact of the intervention or exposure.[25-27]

Neither of the two overviews identified in the scenario reported a test of homogeneity. However, both of them included graphic and tabular displays of the results of the primary studies that suggest differences in study results that are likely to be both clinically important and statistically significant. Both of the overviews suggest possible explanations for the observed heterogeneity (Table 3).

WHAT ARE THE RESULTS?

What Are the Overall Results of the Overview?—In clinical research, investigators collect data from individual patients. Because of the limited capacity of the human mind to handle large amounts of data, investigators use statistical methods to summarize and analyze them. In overviews, investigators collect data from individual studies. These data must also be summarized, and increasingly, investigators are using quantitative methods to do so.

Simply comparing the number of positive studies with the number of negative studies is not an adequate way to summarize the results. With this sort of "vote counting," large and small studies are given equal weights, and (unlikely as it may seem) one investigator may interpret a study as positive, while another investigator interprets the same study as negative.[28] For example, a clinically important effect that is not statistically significant could be interpreted as positive in light of clinical importance and negative in light of statistical significance. There is a tendency to overlook small but clinically important effects if studies with statistically nonsignificant (but potentially clinically important) results are counted as negative.[29] Moreover, a reader cannot tell anything about the magnitude of an effect from a vote count even when studies are appropriately classified using additional categories for studies with a positive or negative trend.

Typically, meta-analysts weight studies according to their size, with larger studies receiving more weight. Thus, the overall results represent a weighted average of the results of the individual studies. Occasionally studies are also given more or less weight depending on their quality, or poorer quality studies might be given a weight of zero (excluded) either in the primary analysis or in a "sensitivity analysis" to see if this makes an important difference in the overall results.

Table 4.—Odds Ratio, Relative Risk, Risk Reduction, and Number Needed to Treat

Treatment or Exposure	Adverse Outcome*	
	Positive	Negative
Positive	*A*	*B*
Negative	*C*	*D*

*When the outcome is undesirable, a relative risk (RR) or odds ratio (OR) of <1.0 represents a beneficial treatment or exposure, with zero representing 100% effectiveness. An absolute risk reduction (ARR) of <0 represents a benefit, and 100% effectiveness would be equivalent to the risk observed in the control group. The OR can also be expressed as $(A/C)+(B/D)$ (ie, the odds of a case having been exposed relative to the odds of a control having been exposed), and both of these expressions are equivalent to $(A \cdot D \cdot D)+(B \cdot C)$. From the two expressions, if *A* is small relative to *B* and *C* is small relative to *D*, the OR and the RR are approximately the same.
Thus,

OR=$(A/B)/(C/D)$
RR=$[A/(A+B)]/[C/(C+D)]$
RR reduction=1−RR
ARR=$[A/(A+B)]-[C/(C+D)]$
Number needed to treat=1/ARR

You should look to the overall results of an overview the same way you look to the results of primary studies. In our articles concerning therapy, we described the relative risk and the absolute risk reduction, and how they could be interpreted.[30] In the articles about diagnostic tests, we discussed likelihood ratios.[31] In overviews of treatment and etiologic and prognostic factors, you will often see the ratio of the odds of an adverse outcome occurring in those exposed (to a treatment or risk factor) to the odds of an adverse outcome in those not exposed. This odds ratio, illustrated in Table 4, has desirable statistical properties when combining results across studies. Whatever method of analysis the investigators used, you should look for a summary measure (such as the number needed to treat[32]) that clearly conveys the practical importance of the result.

Sometimes the outcome measures that are used in different studies are similar but not exactly the same. For example, different trials might measure functional status using different instruments. If the patients and the interventions are reasonably similar, it might still be worthwhile to estimate the average effect of the intervention on functional status. One way of doing this is to summarize the results of each study as an "effect size."[33] The effect size is the difference in outcomes between the intervention and control groups divided by the standard deviation (SD). The effect size summarizes the results of each study in terms of the number of SDs of difference between the intervention and control groups. Investigators can then calculate a weighted average of effect sizes from studies that measured an outcome in different ways.

You are likely to find it difficult to interpret the clinical importance of an effect size (if the weighted average effect is one half of an SD, is this effect clinically trivial, or is it large?). Once again, you should look for a presentation of the results that conveys their practical importance (for example, by translating the summary effect size back into natural units).[34] For instance, if clinicians have become familiar with the significance of differences in walk test scores in patients with chronic lung disease, the effect size of a treatment on a number of measures of functional status (such as the walk test and stair climbing) can be converted back into differences in walk test scores.

Although it is generally desirable to have a quantitative summary of the results of a review, it is not always appropriate. For example, there may be unexplained heterogeneity in study results or the studies may be of such poor quality that the overall results would be uninterpretable. In these cases investigators should still present tables or graphs that summarize the results of the primary studies, and their conclusions should be cautious.

How Precise Were the Results?— In the same way that it is possible to estimate the average effect across studies, it is possible to estimate a CI around that estimate; ie, a range of values with a specified probability (typically 95%) of including the true effect. A previous article in this series provides a guide for understanding CIs.[30]

WILL THE RESULTS HELP ME IN CARING FOR MY PATIENTS?

Can the Results Be Applied to My Patient Care?—One of the advantages of an overview is that since it includes many studies, the results come from a very diverse range of patients. If the results are consistent across studies, they apply to this wide variety of patients. Even so, the clinician may still be left with doubts about the applicability of the results. Perhaps the patient is older than any of those included in the individual trials summarized by the overview. If studies using different members of a class of drug have been combined, one might question whether one of the drugs has a larger effect than the others.

These questions raise the issue of subgroup analysis. Detailed guides for deciding whether to believe subgroup analyses are available.[26,27] One of the most important guides is that conclusions that are drawn on the basis of between-study comparisons (comparing patients in one study with patients in another) should be viewed skeptically. For example, meta-analysis of the effectiveness of β-blockers after myocardial infarction found a statistically significant and clinically important difference in effect between trials of β-blockers with and without intrinsic sympathomimetic activity.[35] This resulted in clinical recommendations that only β-blockers without intrinsic sympathomimetic activity should be used. However, the addition of two subsequent trials eliminated this difference in the overall summary.[25] In fact, a large number of subgroup analyses exploring differences in either patients or the β-blocker regimen used suggest that any apparent differences are probably due to chance.[25]

Other criteria that make a hypothesized difference in subgroups more credible include a big difference in treatment effect; a highly statistically significant difference in treatment effect (the lower the *P* value on the comparison of the different effect sizes in the subgroups, the more credible the difference); a hypothesis that was made before the study began and was one of only a few hypotheses that were tested; consistency across studies; and indirect evidence in support of the difference ("biological plausibility"). If these criteria are not met, the results of a subgroup analysis are less likely to be trustworthy and you should assume that the overall effect across all patients and all treatments, rather than the subgroup effect, applies to the patient at hand and to the treatment under consideration.

Were All Clinically Important Outcomes Considered?—While it is a good idea to look for focused review articles because they are more likely to provide valid results, this does not mean that you should ignore outcomes that are not included in a review. For example, the potential benefits and harms of hormone replacement therapy include reduced risk of fractures and CHD and increased risk of breast cancer and endometrial cancer. Focused reviews of the evidence for individual outcomes are more likely to provide valid results, but a clinical decision requires considering all of them.

Are the Benefits Worth the Harms and Costs?—Finally, either explicitly or implicitly, when making a clinical de-

cision the expected benefits must be weighed against the potential harms and costs. While this is most obvious for deciding whether to use a therapeutic or preventive intervention, providing patients with information about causes of disease or prognosis can also have both benefits and harms. For example, informing a woman about potentially teratogenic exposures might result in her reducing her risk of exposure (with potential benefits), and also cause anxiety or loss of work. Informing an asymptomatic woman with newly detected cancer about her prognosis might help her to plan better, but also label her, cause anxiety, or increase the period during which she is "sick."

A valid review article provides the best possible basis for quantifying the expected outcomes, but these outcomes still must be considered in the context of your patient's values and concerns about the expected outcomes of a decision. In the next articles in this series we will address this issue in the context of decision analysis and clinical practice guidelines.

RESOLUTION OF THE SCENARIO

The two meta-analyses identified in the clinical scenario at the beginning of this article had different objectives, which resulted in differences in the trials included and the analyses. Nonetheless, the results of the two overviews support similar conclusions. Both meta-analyses meet most validity criteria (Table 3), with some limitations. Neither one included explicit assessments of the validity of the primary studies, one does not state whether more than one of the reviewers assessed the studies independently,[1] and there appears (in graphic and tabular displays of the results) to be clinically important heterogeneity among the studies included in both of the overviews. Of more concern than any bias in the overviews, however, is how we should interpret the data on noncardiac and total mortality, an issue that is also the focus of other published meta-analyses of these data,[14,36,37] including one that was published but not yet indexed at the time of the scenario.[38] Despite the uncertainty, it is still possible to draw some reasonable conclusions.

The benefit we can expect from interventions to lower cholesterol depends on the baseline risk of death from CHD and, possibly, whether we are considering dietary or drug interventions.[1,2,38,39] The higher the risk of dying of CHD, the greater the likelihood of benefit. Drug, but not dietary therapy, may be associated with a greater likelihood of death from causes other than CHD.[1,38] The patient described in the scenario has a risk of dying of CHD of approximately 1.0% over the next decade. You would have to treat approximately 1000 such patients for 10 years with a dietary intervention to save a life. If you were to treat such patients with drug therapy, it is not certain you would reduce total mortality. The results suggest that drug therapy should be restricted to those at high risk, such as individuals with known coronary artery disease, and whether diet therapy is worthwhile in low-risk individuals (such as the patient in the scenario) is uncertain.

References

1. Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ.* 1993;306:1367-1373.
2. Silberberg JS, Henry DA. The benefits of reducing cholesterol levels: the need to distinguish primary from secondary prevention, I: a meta-analysis of cholesterol-lowering trials. *Med J Aust.* 1991; 155:665-666, 669-670.
3. Oxman AD, Sackett DL, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, I: how to get started. *JAMA.* 1993;270:2093-2095.
4. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med.* 1987;107: 224-233.
5. Sacks HS, Berries J, Reitman D, Arcona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med.* 1987;316:450-455.
6. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J.* 1988;138:697-703.
7. Mulrow CD. The medical review article: state of the science. *Ann Intern Med.* 1987;106:485-488.
8. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *JAMA.* 1992;268:240-248.
9. Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. In: *Medicines for the Year 2000.* London, England: Office of Health Economics; 1979:2-12.
10. *The Cochrane Collaboration.* Oxford, England: UK Cochrane Centre, National Health Service Research and Development Programme; 1994. Brochure.
11. Enkin MW, Keirse MJNC, Renfrew MJ, Neilson JP, eds. Cochrane Pregnancy and Childbirth Database [derived from the Cochrane Database of Systematic Reviews], Cochrane Updates on Disk, Update Software. Oxford, England: UK Cochrane Centre; 1993: disk issue 2.
12. Chalmers TC, Berrier J, Sacks HS, et al. Meta-analysis of clinical trials as a scientific discipline, II: replicate variability and comparison of studies that agree and disagree. *Stat Med.* 1987;6:733-744.
13. Gøtzsche PC. Reference bias in reports of drug trials. *BMJ.* 1987;295:654-656.
14. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ.* 1992;305:15-19.
15. Cook DJ, Guyatt GH, Ryan G, et al. Should unpublished data be included in meta-analyses? current convictions and controversies. *JAMA.* 1993; 269:2749-2753.
16. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA.* 1990; 263:1385-1389.
17. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA.* 1992;267:374-378.
18. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867-872.
19. Dickersin K, Min Y-I. NIH clinical trials and publication bias. *Online J Curr Clin Trials [serial online].* April 28, 1993;50.
20. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc.* 1988;151:445-463.
21. Williamson JW, Goldschmidt PG, Colton T. The quality of medical literature: analysis of validation assessments. In: Bailar JC, Mosteller F, eds. *Medical Uses of Statistics.* Waltham, Mass: NEJM Books; 1986:370-391.
22. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med.* 1987;82:498-510.
23. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe A. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol.* 1992;45:255-265.
24. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled trials. *N Engl J Med.* 1983;309:1358-1361.
25. Peto R. Why do we need systematic overviews of randomized trials? *Stat Med.* 1987;6:223-240.
26. Yusuf S, Wiftes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA.* 1991;266:93-98.
27. Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. *Ann Intern Med.* 1992;116:78-84.
28. Glass GV, McGaw B, Smith ML. *Meta-analysis in Social Research.* Newbury Park, Calif: Sage; 1981:18-20.
29. Cooper RM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. *Psychol Bull.* 1980;87:442-449.
30. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients? *JAMA.* 1994;271:59-63.
31. Jaeschke R, Guyatt GH, Sackett DL, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, III: how to use an article about a diagnostic test, B: what are the results and will they help me in caring for my patients? *JAMA.* 1994;271:703-707.
32. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med.* 1988;318: 1728-1733.
33. Rosenthal R. *Meta-analytic Procedures for Social Research.* 2nd ed. Newbury Park, Calif: Sage; 1991.
34. Smith K, Cook DJ, Guyatt GH, Mudhaven J, Oxman AD. Respiratory muscle training in chronic airflow limitation: a meta-analysis. *Am Rev Respir Dis.* 1992;145:533-539.
35. 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;27:335-371.
36. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *BMJ.* 1990;301:309-314.
37. Davey Smith G, Pekkanen J. Should there be a moratorium on the use of cholesterol lowering drugs? *BMJ.* 1992;304:431-434.
38. Law MR, Thompson SG, Wald NJ. Assessing possible hazards of reducing serum cholesterol. *BMJ.* 1994;308:373-379.
39. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ.* 1994;308:367-373.