

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES,
AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.

ABSTRACT

Background In the hierarchy of research designs, the results of randomized, controlled trials are considered to be evidence of the highest grade, whereas observational studies are viewed as having less validity because they reportedly overestimate treatment effects. We used published meta-analyses to identify randomized clinical trials and observational studies that examined the same clinical topics. We then compared the results of the original reports according to the type of research design.

Methods A search of the Medline data base for articles published in five major medical journals from 1991 to 1995 identified meta-analyses of randomized, controlled trials and meta-analyses of either cohort or case-control studies that assessed the same intervention. For each of five topics, summary estimates and 95 percent confidence intervals were calculated on the basis of data from the individual randomized, controlled trials and the individual observational studies.

Results For the five clinical topics and 99 reports evaluated, the average results of the observational studies were remarkably similar to those of the randomized, controlled trials. For example, analysis of 13 randomized, controlled trials of the effectiveness of bacille Calmette-Guérin vaccine in preventing active tuberculosis yielded a relative risk of 0.49 (95 percent confidence interval, 0.34 to 0.70) among vaccinated patients, as compared with an odds ratio of 0.50 (95 percent confidence interval, 0.39 to 0.65) from 10 case-control studies. In addition, the range of the point estimates for the effect of vaccination was wider for the randomized, controlled trials (0.20 to 1.56) than for the observational studies (0.17 to 0.84).

Conclusions The results of well-designed observational studies (with either a cohort or a case-control design) do not systematically overestimate the magnitude of the effects of treatment as compared with those in randomized, controlled trials on the same topic. (N Engl J Med 2000;342:1887-92.)

©2000, Massachusetts Medical Society.

RANDOMIZED, controlled trials were introduced into clinical medicine when streptomycin was evaluated in the treatment of tuberculosis¹ and have become the gold standard for assessing the effectiveness of therapeutic agents.²⁻⁴ The ascendancy of randomized, controlled trials was hastened by a landmark article⁵ comparing published randomized, controlled studies with those that used observational designs. That review of

the literature identified six different therapies evaluated in both randomized, controlled trials (50 studies) and trials with historical controls (56 studies). For each study, subjects in the treatment group were found to have similar rates of the outcome in question regardless of study design, but subjects in the control group in trials with historical controls had worse outcomes than control subjects in randomized, controlled trials. The agent being tested was considered effective in 44 of 56 trials with historical controls (79 percent), but in only 10 of 50 randomized, controlled trials (20 percent). The authors concluded that biases in patient selection may irretrievably weight the outcome of historical controlled trials in favor of new therapies.⁵

Current criticisms of observational studies involve, in addition to trials with historical controls, cohort studies with concurrent selection of control subjects, as well as case-control designs. Advocates of “evidence-based medicine”⁶ classify studies according to “grades of evidence” on the basis of the research design, using internal validity (i.e., the correctness of the results) as the criterion for hierarchical rankings. An example of such rankings is shown in Table 1. The highest grade is reserved for research involving “at least one properly randomized controlled trial,” and the lowest grade is applied to descriptive studies (e.g., case series) and expert opinion; observational studies, both cohort studies and case-control studies, fall at intermediate levels.⁷ Although the quality of studies is sometimes evaluated within each grade, each category is considered methodologically superior to those below it. This hierarchical approach to study design has been promoted widely in individual reports, meta-analyses, consensus statements, and educational materials for clinicians.

Systematic reviews and meta-analyses offer an opportunity to test implicit assumptions about the hierarchy of research designs. If particular associations between exposure and outcome were studied in both randomized, controlled trials and cohort or case-control studies, and if these studies were then included in meta-analyses, the results could be compared according to study design, as was done for trials with

From the Departments of Internal Medicine (J.C., N.S., R.I.H.) and Epidemiology and Public Health (R.I.H.), Yale University School of Medicine, New Haven, Conn.; and the Clinical Epidemiology Unit, West Haven Veterans Affairs Medical Center, West Haven, Conn. (J.C.). Address reprint requests to Dr. Concato at the Department of Internal Medicine, Yale University School of Medicine, 333 Cedar St., SHM IE-61, New Haven, CT 06510, or at john.concato@yale.edu.

TABLE 1. GRADES OF EVIDENCE FOR THE PURPORTED QUALITY OF STUDY DESIGN.*

I	Evidence obtained from at least one properly randomized, controlled trial.
II-1	Evidence obtained from well-designed controlled trials without randomization.
II-2	Evidence obtained from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.
II-3	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence.
III	Opinions of respected authorities, based on clinical experience; descriptive studies and case reports; or reports of expert committees.

*The grades are those of the U.S. Preventive Services Task Force.⁷

historical controls.⁵ In the current study, however, we evaluated only observational studies that used contemporaneous control subjects. The variation in point estimates of associations between exposure and outcome provides data to confirm or refute the assumptions regarding observational studies, as well as the strengths and limitations of a “design hierarchy.”

METHODS

We identified published reports of randomized, controlled trials and reports of observational studies with either a cohort design (i.e., with concurrent selection of controls) or a case-control design that assessed the same clinical topic (clinical intervention and outcome). The articles were selected by first identifying meta-analyses published in five major journals (*Annals of Internal Medicine*, the *British Medical Journal*, the *Journal of American Medical Association*, the *Lancet*, and the *New England Journal of Medicine*) from 1991 to 1995. The meta-analyses were identified by searching Medline for the terms “meta-analysis,” “meta-analyses,” “pooling,” “combining,” “overview,” and “aggregation.” Additional references were found in *Current Contents*, supplemented by searches of printed copies of the relevant journals.

The meta-analyses were then classified, by consensus of two investigators, as including randomized, controlled trials only, observational studies only, or both. Clinical trials were defined as studies that used random assignment of interventions; observational studies had either cohort or case-control designs. Meta-analyses were excluded if they involved cohort studies with historical controls or clinical trials with nonrandom assignment of interventions, or if they did not report results in the format of point estimates (e.g., relative risks or odds ratios) and confidence intervals. In this context, odds ratios and relative risks will be similar in magnitude, because the rates of the outcome events are low. The remaining meta-analyses were then reviewed, and the original studies cited in the bibliographies were retrieved. Although the meta-analyses themselves often used criteria related to quality in the selection of studies, we also evaluated the original reports, using published scoring criteria.⁸⁻¹⁰

We performed two main analyses, the results of which are reported here. In the first analysis, summary estimates and 95 percent confidence intervals were determined for each clinical topic, according to whether the data came from randomized, controlled trials or observational studies. Pooled analyses were performed according to the method of DerSimonian and Laird¹¹ as described by Fleiss.¹² We chose the random-effects model for combining data because it provides more conservative results (wider confidence intervals) than a fixed-effects model. When possible, pooled estimates were computed on the basis of data from the original

reports. If such estimates were not available, however, the data from the published reports of the meta-analyses were used. For example, the meta-analysis of cohort studies involving blood-pressure measurements¹³ used an adjustment for bias due to regression dilution (regression toward the mean) for the calculation of point estimates and 95 percent confidence intervals. A second analysis was used to describe the range of results from the individual studies; such a range is presented for each clinical topic, with the studies grouped according to research design.

RESULTS

The Studies

The search strategy yielded 102 citations of meta-analyses (available from the authors on request), including 6 that examined both randomized, controlled trials and observational studies of the same clinical topic. The remaining 96 meta-analyses included randomized, controlled trials only (72 reports) or observational studies only (24 reports). Among these reports, three additional clinical topics were identified for which each type of study design was assessed separately. The nine clinical topics (a total of 12 meta-analyses) included five that met our eligibility criteria and provided the data for the current analysis. The remaining four topics (data not shown) were excluded because they were the subject of observational studies with only historical controls or because no data were available in the form of point estimates.

The five clinical topics (Table 2) were investigated in 99 original articles with a total of 1,871,681 study subjects. The pooled (summary) point estimates are presented in Table 2, and the range of point estimates from each study, when available, is shown in Figure 1. Among the 99 studies, 6 randomized, controlled trials (6 percent) contributed data to the summary results but did not have a quantitative point estimate for the main association of interest (because no outcomes were observed in one or both of the compared groups).

Bacille Calmette–Guérin Vaccine and Tuberculosis

The effectiveness of bacille Calmette–Guérin vaccine against active tuberculosis was examined in a meta-analysis¹⁴ of 13 randomized trials (with 359,922 subjects), yielding a pooled relative risk of 0.49 (95 percent confidence interval, 0.34 to 0.70), and 10 case-control studies (with 6511 subjects), yielding a pooled odds ratio of 0.50 (95 percent confidence interval, 0.39 to 0.65). The point estimates from the original articles ranged from 0.20 to 1.56 for randomized, controlled trials and from 0.17 to 0.84 for observational studies.

Screening Mammography and Mortality from Breast Cancer

A meta-analysis¹⁵ of eight randomized trials (with 429,043 subjects) of the relation between screening mammography and mortality from breast cancer found a protective effect of screening among women 40 years of age or older, with a pooled relative risk

TABLE 2. TOTAL NUMBER OF SUBJECTS AND SUMMARY ESTIMATES FOR THE EFFECT OF FIVE INTERVENTIONS ACCORDING TO THE TYPE OF RESEARCH DESIGN.

CLINICAL TOPIC	TYPE OF STUDY	META-ANALYSIS*	TOTAL NO. OF SUBJECTS	SUMMARY ESTIMATE (95% CI)†
Bacille Calmette–Guérin vaccine and tuberculosis	13 Randomized, controlled	Colditz et al. ¹⁴	359,922	0.49 (0.34–0.70)
	10 Case–control	Colditz et al. ¹⁴	6,511	0.50 (0.39–0.65)
Mammography and mortality from breast cancer	8 Randomized, controlled	Kerlikowske et al. ¹⁵	429,043	0.79 (0.71–0.88)
	4 Case–control	Kerlikowske et al. ¹⁵	132,456	0.61 (0.49–0.77)
Cholesterol levels and death due to trauma	6 Randomized, controlled	Cummings and Psaty ¹⁶	36,910	1.42 (0.94–2.15)
	14 Cohort	Jacobs et al. ¹⁷	9,377	1.40 (1.14–1.66)
Treatment of hypertension and stroke	14 Randomized, controlled	Collins et al. ¹⁸	36,894	0.58 (0.50–0.67)
	7 Cohort	MacMahon et al. ¹³	405,511	0.62 (0.60–0.65)
Treatment of hypertension and coronary heart disease	14 Randomized, controlled	Collins et al. ¹⁸	36,894	0.86 (0.78–0.96)
	9 Cohort	MacMahon et al. ¹³	418,343	0.77 (0.75–0.80)

*Meta-analyses that included either randomized, controlled trials or observational studies are cited.

†CI denotes confidence interval.

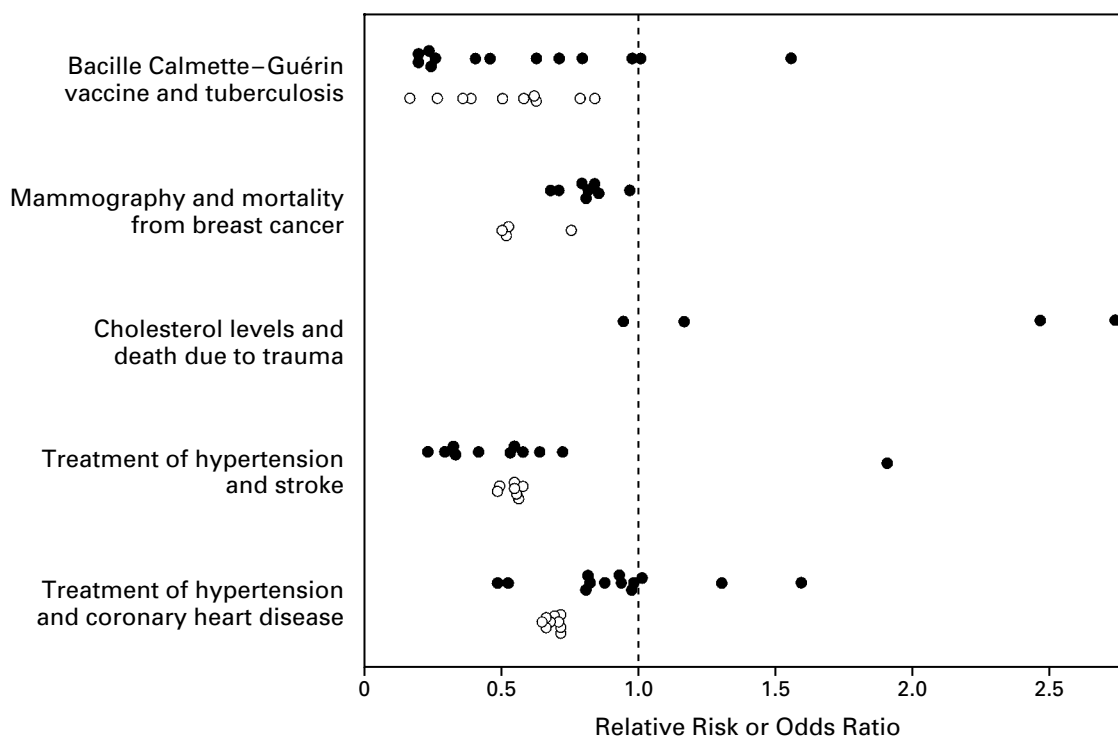


Figure 1. Range of Point Estimates According to Type of Research Design.

The studies evaluated bacille Calmette–Guérin vaccine and active tuberculosis (13 randomized, controlled trials and 10 case–control studies), screening mammography and mortality from breast cancer (8 randomized, controlled trials and 4 case–control studies), cholesterol levels and death due to trauma among men (4 of 6 randomized, controlled trials [2 trials did not provide point estimates]; the results of the 14 cohort studies were not reported individually), treatment of hypertension and stroke among only the men in the studies (11 randomized, controlled trials and 7 cohort studies), and treatment of hypertension and coronary heart disease among only the men in the studies (13 of 14 randomized, controlled trials [1 trial did not provide point estimates] and 9 cohort studies). Solid circles indicate randomized, controlled trials, and open circles observational studies.

of 0.79 (95 percent confidence interval, 0.71 to 0.88); a benefit of screening was also found in four case-control studies (with 132,456 subjects), with a pooled odds ratio of 0.61 (95 percent confidence interval, 0.49 to 0.77). The range of point estimates was 0.68 to 0.97 among randomized, controlled trials and 0.51 to 0.76 among observational studies.

Cholesterol Levels and Death Due to Trauma

A meta-analysis¹⁶ of six randomized, controlled trials (with 36,910 men) reported a pooled relative risk of death due to trauma of 1.42 (95 percent confidence interval, 0.94 to 2.15), indicating an increased risk among subjects taking the drugs that were studied. A separate meta-analysis¹⁷ of 14 cohort studies (9377 subjects) reported a pooled hazard ratio of 1.40 (95 percent confidence interval, 1.14 to 1.66). The range of point estimates from the cohort studies was not reported, precluding a comparison with the range of results from the randomized, controlled trials (the range was 0.25 to 2.74 in four randomized, controlled trials; two trials did not report quantitative results).

Treatment of Hypertension and Stroke

The relation between the treatment of hypertension and a first occurrence of stroke (i.e., the effectiveness of primary prevention) was examined in meta-analyses of 14 randomized, controlled trials¹⁸ and 7 cohort studies.¹³ The pooled results from the randomized, controlled trials (36,894 subjects) yielded a point estimate of the risk of stroke of 0.58 (95 percent confidence interval, 0.50 to 0.67) among patients given antihypertensive treatment; the pooled results from the observational studies (405,511 subjects) yielded an adjusted point estimate of 0.62 (95 percent confidence interval, 0.60 to 0.65). The range of results was 0.24 to 1.91 for randomized, controlled trials (three of the randomized, controlled trials did not provide point estimates) and 0.49 to 0.58 (unadjusted values) for cohort studies.

Treatment of Hypertension and Coronary Heart Disease

A meta-analysis¹⁸ of 14 randomized, controlled trials (36,894 subjects) reported a pooled point estimate of the relative risk of coronary heart disease of 0.86 (95 percent confidence interval, 0.78 to 0.96) among patients treated for hypertension, and a meta-analysis¹³ of 9 cohort studies (418,343 subjects) reported an adjusted, pooled point estimate of 0.77 (95 percent confidence interval, 0.75 to 0.80). The range of results was 0.49 to 1.60 for randomized, controlled trials (one randomized, controlled trial did not report a relative risk) and 0.65 to 0.72 (unadjusted values) for cohort studies.

DISCUSSION

Our results challenge the current consensus about a hierarchy of study designs in clinical research. Con-

trary to prevailing beliefs, the “average results” from well-designed observational studies (with a cohort or case-control design) did not systematically overestimate the magnitude of the associations between exposure and outcome as compared with the results of randomized, controlled trials of the same topic. Rather, the summary results of randomized, controlled trials and observational studies were remarkably similar for each clinical topic we examined (Table 2). Viewed individually, the observational studies had less variability in point estimates (i.e., less heterogeneity of results) than randomized, controlled trials on the same topic (Fig. 1). In fact, only among randomized, controlled trials did some studies report results in a direction opposite that of the pooled point estimate, representing a paradoxical finding (e.g., treatment of hypertension was unexpectedly associated with higher rates of coronary heart disease in several clinical trials).

Although the data we present are a challenge to accepted beliefs, the findings are consistent with three other types of available evidence. For example, previous investigations have shown that observational cohort studies can produce results similar to those of randomized, controlled trials when similar criteria are used to select study subjects. In addition, data from nonmedical research do not support a hierarchy of research designs. Finally, the finding that there is substantial variation in the results of randomized, controlled trials is consistent with prior evidence of contradictory results among randomized, controlled trials.

First, there is evidence that observational studies can be designed with rigorous methods that mimic those of clinical trials and that well-designed observational studies do not consistently overestimate the effectiveness of therapeutic agents. An analysis¹⁹ of 18 randomized and observational studies in health-services research found that treatment effects may differ according to research design, but that “one method does not give a consistently greater effect than the other.” The treatment effects were most similar when the exclusion criteria were similar and when the prognostic factors were accounted for in observational studies.

A specific method used to strengthen observational studies (the “restricted cohort” design⁹) adapts principles of the design of randomized, controlled trials to the design of an observational study as follows: it identifies a “zero time” for determining a patient’s eligibility and base-line features, uses inclusion and exclusion criteria similar to those of clinical trials, adjusts for differences in base-line susceptibility to the outcome, and uses statistical methods (e.g., intention-to-treat analysis) similar to those of randomized, controlled trials. When these procedures were used in a cohort study⁹ evaluating the benefit of beta-blockers after recovery from myocardial infarction, the use of a restricted cohort produced results con-

sistent with corresponding findings from the Beta-Blocker Heart Attack Trial²⁰: the three-year reductions in mortality were 33 percent and 28 percent, respectively.

Second, data in the literature of other scientific disciplines support our contention that research design should not be considered a rigid hierarchy. A comprehensive review of research on various psychological, educational, and behavioral treatments²¹ identified 302 meta-analyses and examined the reports on the basis of several features, including research design. Results were presented from the 74 meta-analyses that included studies with randomized and observational designs. To allow for comparisons among various topics with different outcome variables, effect size was used as a unit-free measure of the effect of the intervention. The observational designs did not consistently overestimate or underestimate the effect of treatment; the mean value of the difference was a trivial 0.05. Thus, these independent data do not support the contention that observational studies overestimate effects as compared with randomized, controlled trials.

Third, a review of more than 200 randomized, controlled trials on 36 clinical topics found numerous examples of conflicting results.²² A more recent example is offered by studies addressing whether therapy with monoclonal antibodies improves outcomes among patients with septic shock (reviewed by Horn²³ and Angus et al.²⁴). In addition, one study²⁵ found that the results of meta-analyses based on randomized, controlled trials were often discordant with those of large, simple trials on the same clinical topic. Regardless of the reasons why randomized, controlled trials produce heterogeneous results, the available evidence indicates that a single randomized trial (or only one observational study) cannot be expected to provide a gold-standard result that applies to all clinical situations.

One possible explanation for the finding that observational studies may be less prone to heterogeneity in results than randomized, controlled trials is that each observational study is more likely to include a broad representation of the population at risk. In addition, there is less opportunity for differences in the management of subjects among observational studies. For example, although there is general agreement that physicians do not use therapeutic agents in a uniform way, an observational study would usually include patients with coexisting illnesses and a wide spectrum of disease severity, and treatment would be tailored to the individual patient. In contrast, each randomized, controlled trial may have a distinct group of patients as a result of specific inclusion and exclusion criteria regarding coexisting illnesses and severity of disease, and the experimental protocol for therapy may not be representative of clinical practice.

The relevance of our findings extends beyond their implications for expert panels such as the U.S. Preventive Services Task Force.⁷ A popular “users’ guide” for clinicians²⁶ warns that “the potential for bias is much greater in cohort and case–control studies than in randomized, controlled trials, [so that] recommendations from overviews combining observational studies will be much weaker.” The studies cited to support that claim^{27,28} (and similar claims²⁹), however, compare randomized, controlled trials with trials using historical controls, unblinded clinical trials, or clinical trials without randomly assigned control subjects — not with the types of cohort and case–control studies included in our investigation. Thus, data based on “weaker” forms of observational studies are often mistakenly used to criticize all observational research.

We examined the possibility that the quality of individual studies could explain our findings. For example, randomized, controlled trials that did not satisfy criteria with respect to quality could be the source of variability in point estimates, or the observational studies might be of uniformly high quality. When standard assessments of quality were applied to the studies, however, no association was found between the number of criteria for high-quality research that a study satisfied and the rank order of its point estimate (data not shown). Thus, although quality scores have been used in some situations to separate high-quality from low-quality randomized, controlled trials,⁸ our results are consistent with other studies³⁰ that did not find an association between summary measures of quality and treatment effects. The issue of how to judge the validity of each study (in terms of the methodologic aspects relevant to each investigation) is beyond the scope of this report. However, judging validity is often not as simple as identifying the type of research design or assessing general characteristics of the study.^{8,26}

The meta-analyses of randomized, controlled trials and observational studies that we evaluated included single reports that combined the two types of research design, as well as separate reports for each category (Table 2). This mix of reports offers reassurance that our findings are not attributable to the methods used in each meta-analysis. (The overall paucity of meta-analyses including both randomized, controlled trials and observational studies of the same research topic is consistent with our premise that observational studies are not considered trustworthy and that they are therefore not included in such investigations.) The validity of our analysis is also supported by another investigation comparing randomized, controlled trials and observational studies of screening mammography that found results similar to ours.³¹

Despite the consistency of our results (involving five clinical topics and 99 separate studies), as well

as the confirmatory evidence available in the literature, we believe that the appropriate role of observational studies may vary in different situations. For example, observational investigations of some kinds of treatments (e.g., surgical operations and other invasive therapies) may be more prone to selection bias than the observational studies of drugs and noninvasive tests that we examined in this study, and “softer” outcomes (e.g., functional status) may be more readily assessed in randomized, controlled trials. In addition, we are aware of the risk that the results of poorly done observational studies may be used inappropriately — for example,³² to promote ineffective alternative therapies.

Randomized, controlled trials will (and should) remain a prominent tool in clinical research, but the results of a single randomized, controlled trial, or of only one observational study, should be interpreted cautiously. If a randomized, controlled trial is later determined to have given wrong answers, evidence both from other trials and from well-designed cohort or case-control studies can and should be used to find the right answers. The popular belief that only randomized, controlled trials produce trustworthy results and that all observational studies are misleading does a disservice to patient care, clinical investigation, and the education of health care professionals.

Dr. Concato is the recipient of a Career Development Award from the Veterans Affairs Health Services Research and Development Service.

REFERENCES

1. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ* 1948;2:769-82.
2. Byar DP, Simon RM, Friedewald WT, et al. Randomized clinical trials: perspectives on some recent ideas. *N Engl J Med* 1976;295:74-80.
3. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 1984;70:767-74.
4. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;52:487-97.
5. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233-40.
6. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
7. Preventive Services Task Force. Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. 2nd ed. Baltimore: Williams & Wilkins, 1996.
8. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
9. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990;89:630-8.
10. Horwitz RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 1979;66:556-64.
11. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
12. Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res* 1993;2:121-45.
13. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990;335:765-74.
14. Colditz GA, Brewer TF, Berkey CS, et al. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. *JAMA* 1994;271:698-702.
15. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography: a meta-analysis. *JAMA* 1995;273:149-54.
16. Cummings P, Psaty BM. The association between cholesterol and death from injury. *Ann Intern Med* 1994;120:848-55.
17. Jacobs D, Blackburn H, Higgins M, et al. Report of the Conference on Low Blood Cholesterol: mortality associations. *Circulation* 1992;86:1046-60.
18. Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease. 2. Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990;335:827-38.
19. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research: interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;319:312-5.
20. A randomized trial of propranolol in patients with acute myocardial infarction. I. Mortality results. *JAMA* 1982;247:1707-14.
21. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *Am Psychol* 1993;48:1181-209.
22. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 1987;82:498-510.
23. Horn KD. Evolving strategies in the treatment of sepsis and systemic inflammatory response syndrome (SIRS). *QJM* 1998;91:265-77.
24. Angus DC, Birmingham MC, Balk RA, et al. E5 murine monoclonal antiendotoxin antibody in gram-negative sepsis: a randomized controlled trial. *JAMA* 2000;283:1723-30.
25. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.
26. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guide to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;274:1800-4. [Erratum, *JAMA* 1996;275:1232.]
27. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-61.
28. Sacks HS, Chalmers TC, Smith H Jr. Sensitivity and specificity of clinical trials: randomized v historical controls. *Arch Intern Med* 1983;143:753-5.
29. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
30. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
31. Demissie K, Mills OF, Rhoads GG. Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J Clin Epidemiol* 1998;51:81-91.
32. Angell M, Kassirer JP. Alternative medicine — the risks of untested and unregulated remedies. *N Engl J Med* 1998;339:839-41.