Observational *Versus* Experimental Studies: What's the Evidence for a Hierarchy?

John Concato

Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, and the Clinical Epidemiology Research Center, West Haven Veterans Affairs Medical Center, West Haven, Connecticut 06516

Summary: The tenets of evidence-based medicine include an emphasis on hierarchies of research design (i.e., study architecture). Often, a single randomized, controlled trial is considered to provide "truth," whereas results from any observational study are viewed with suspicion. This paper describes information that contradicts and discourages such a rigid approach to evaluating the quality of research design. Unless a more balanced strategy evolves, new claims of methodological authority may be just as problematic as the traditional claims of medical authority that have been criticized by proponents of evidence-based medicine. **Key Words:** Cohort studies, case control studies, clinical trials, evidence-based medicine, bias.

INTRODUCTION

Evidence-based medicine classifies studies into grades of evidence based on research architecture.[1,2] This hierarchical approach to study design has been promoted widely in individual reports, meta-analyses, consensus statements, and educational materials for clinicians. For example, a prominent publication[3] reserved the highest grade for "at least one properly randomized, controlled trial," and the lowest grade for descriptive studies (e.g., case series) and expert opinion. Observational studies, including cohort and case-control, fall into intermediate levels (Table 1). Although the quality of studies is sometimes evaluated within each grade, each category is considered methodologically superior to level(s) below it.

The ascendancy of randomized, controlled trials (experimental studies) to become the "gold standard" strategy for assessing the effectiveness of therapeutic agents[4–6] was based in part on a landmark paper[7] comparing published articles that used randomized and historical control trial designs. The corresponding results found that the agent being tested was considered effective in 44 of 56 (79%) historical controlled trials, but only 10 of 50 (20%) randomized, controlled trials. The authors concluded "biases in patient selection may irretrievably weight the outcome of historical controlled trials in favor of new therapies."[7]

Although the cited article[7] compared randomized, controlled trials to historical controlled trials only, contemporary criticisms of observational studies also include cohort studies with concurrent (nonhistorical) selection of control subjects as well as case-control designs.[8] A possibility exists, however, that data based on "weaker" forms of observational studies can be used mistakenly to criticize all observational research. The premise of this paper is that evidence-based medicine has contributed to the development of a rigid hierarchy of research design that underestimates the limitations of randomized, controlled trials, and overstates the limitations of observational studies.

WHY USE A HIERARCHY OF RESEARCH DESIGN?

A hierarchy of types of research design would be desirable for providing a "checklist" to evaluate clinical studies, but the complexity of medical research suggests that such approaches are overly simplistic. Although randomization protects against certain types of bias that can threaten the validity of a study (i.e., obtaining the correct answer to the question posed, among the study participants involved), a corresponding randomized, controlled trials protocol may restrict the sample of patients selected, the intervention delivered, or the outcome(s) mea-

Address correspondence and reprint requests to John Concato, M.D., M.P.H., Yale University School of Medicine, Department of Internal Medicine, Sterling Hall of Medicine IE-61, New Haven, CT 06510. E-mail: john.concato@yale.edu.

TABLE 1. *"Grades of Evidence" Rating the Purported Quality of Study Design*[3]

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.

sured, impairing the so-called generalizability of a study (i.e., the extent to which it applies to patients in the "real world"). For example, a randomized, controlled trial may exclude older patients, it may administer therapy in a manner that is difficult to replicate in actual practice, or it may use short-term or surrogate endpoints. In addition, numerous problems can occur when randomized, controlled trials are conducted improperly. Conversely, if properly-conducted observational studies can overcome threats to validity (using strategies discussed later in this paper), and if such studies incorporate more relevant clinical features, then corresponding results would likely be very generalizable to practicing clinicians. Yet, the conventional wisdom suggests that observational studies consistently provide biased results compared with randomized, controlled trials, regardless of the type of observational study or how well it was conducted. The remainder of this paper will focus on these issues.

EVIDENCE AGAINST A RIGID HIERARCHY

A recent study recognized that systematic reviews and meta-analyses offered an opportunity to test the implicit assumptions of grades (or levels) of evidence and similar hierarchies of research design.[9] We identified particular exposure–outcome associations that were studied with both randomized, controlled trials as well as cohort or case-control studies. The major distinctions of our approach (compared with prior research), however, were that we evaluated observational studies that used concurrent (not historical) control subjects, and we focused on summary results rather than individual study findings. The variation in point estimates of exposure–outcome associations provided data to confirm or refute the assumptions regarding observational studies, as well as the strengths and limitations of a "design hierarchy."

Our methods involved identifying meta-analyses published in five major journals (*Annals of Internal Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *Lancet*, and *New England Journal of Medicine*) from 1991 to 1995, using searches of MEDLINE, with the terms "meta-analysis," "meta-analyses," "pooling," "combining," "overview," and "aggregation." Additional references were found in *Current Contents*, supplemented by manual searches of the relevant journals. The meta-analyses identified via this process were then classified by consensus as including clinical trials only, observational studies only, or both. Clinical trials were defined as studies that used randomized interventions; observational studies included cohort or case-control designs. Meta-analyses were excluded if they were based on cohort studies with historical control subjects, or clinical trials with nonrandom assignment of interventions, or if they did not report results in the format of a point estimate (e.g., relative risk, odds ratio) and confidence intervals. The remaining meta-analyses were then reviewed, and the original studies cited in the bibliographies were retrieved.

The search strategy yielded 102 citations for meta-analyses, mainly involving (as expected) randomized, controlled trials only. Data for five clinical topics[10–15] met our eligibility criteria and provided sufficient data for analysis, involving 99 original articles and 1,871,681 total study subjects. The summary (pooled) point estimates are presented in Table 2, and the ranges of the point estimates are displayed in Figure 1. For example, the relationship between treatment of hypertension and the first occurrence of stroke (i.e., primary prevention) was examined in meta-analyses of 14 randomized, controlled trials[15] and seven cohort studies.[10] The pooled results from randomized, controlled trials ($N = 36,894$) found a point estimate of 0.58 (95% confidence interval 0.50-0.67); the pooled results from observational studies ($N = 405,511$) found an adjusted point estimate of 0.62 (95% confidence interval 0.60–0.65). Results for other associations (Table 2) were also similar, based on data from randomized, controlled trials and observational studies. In another example, the effectiveness of bacillus Calmette-Guerin (BCG) vaccine against tuberculosis was examined in a meta-analysis[11] that included 13 randomized trials ($N = 359,922$ subjects) with a pooled relative risk of 0.49 (95% confidence interval 0.34-0.70), and 10 case-control studies ($N = 6511$ subjects) with a pooled odds ratio of 0.50 (95% confidence interval 0.39–0.65).

The results of our investigation contradict the idea of a "fixed" hierarchy of study design in clinical research.

TABLE 2. *Total Number of Subjects and Summary Estimates for the Impact of Five Interventions ("Clinical Topics") Based on Type of Research Design*

Clinical Topic	Study Type	Total Subjects	Summary Estimate (95% CI)	Reference No.*
Treatment of hypertension and stroke	14 RCT	36,894	0.58 (0.50–0.67)	15
	7 cohort	405,511	0.62 (0.60–0.65)	10
Treatment of hypertension and CHD	14 RCT	36,894	0.86 (0.78–0.96)	15
	9 cohort	418,343	0.77 (0.75–0.80)	10
Bacillus Calmette-Guerin vaccine and tuberculosis	13 RCT	359,922	0.49 (0.34–0.70)	11
	10 case-control	6511	0.50 (0.39–0.65)	11
Mammography and breast cancer mortality	8 RCT	429,043	0.79 (0.71–0.88)	12
	4 case-control	132,456	0.61 (0.49–0.77)	12
Treatment of hyperlipidemia and traumatic death	6 RCT	36,910	1.42 (0.94–2.15)	13
	14 cohort	9377	1.40 (1.14–1.66)	14

*Citation for meta-analysis that included corresponding randomized, controlled trials or observational studies.
CHD = coronary heart disease; CI = confidence interval; RCT = randomized, controlled trial.

Importantly, another publication[16] addressing the same general question found "little evidence that estimates of treatment effects in observational studies reported after 1984 are either consistently larger than or qualitatively different from those obtained in randomized, controlled trials." In addition, an evaluation[17] of the literature on screening mammography found similar results to ours on that particular topic. Thus, contrary to prevailing beliefs, average results from well-designed observational (cohort and case-control) studies did not systematically overestimate the magnitude of exposure-outcome associations reported in randomized, controlled trials. Rather, the summary results from randomized, controlled trials and observational studies were remarkably similar for each clinical question addressed.

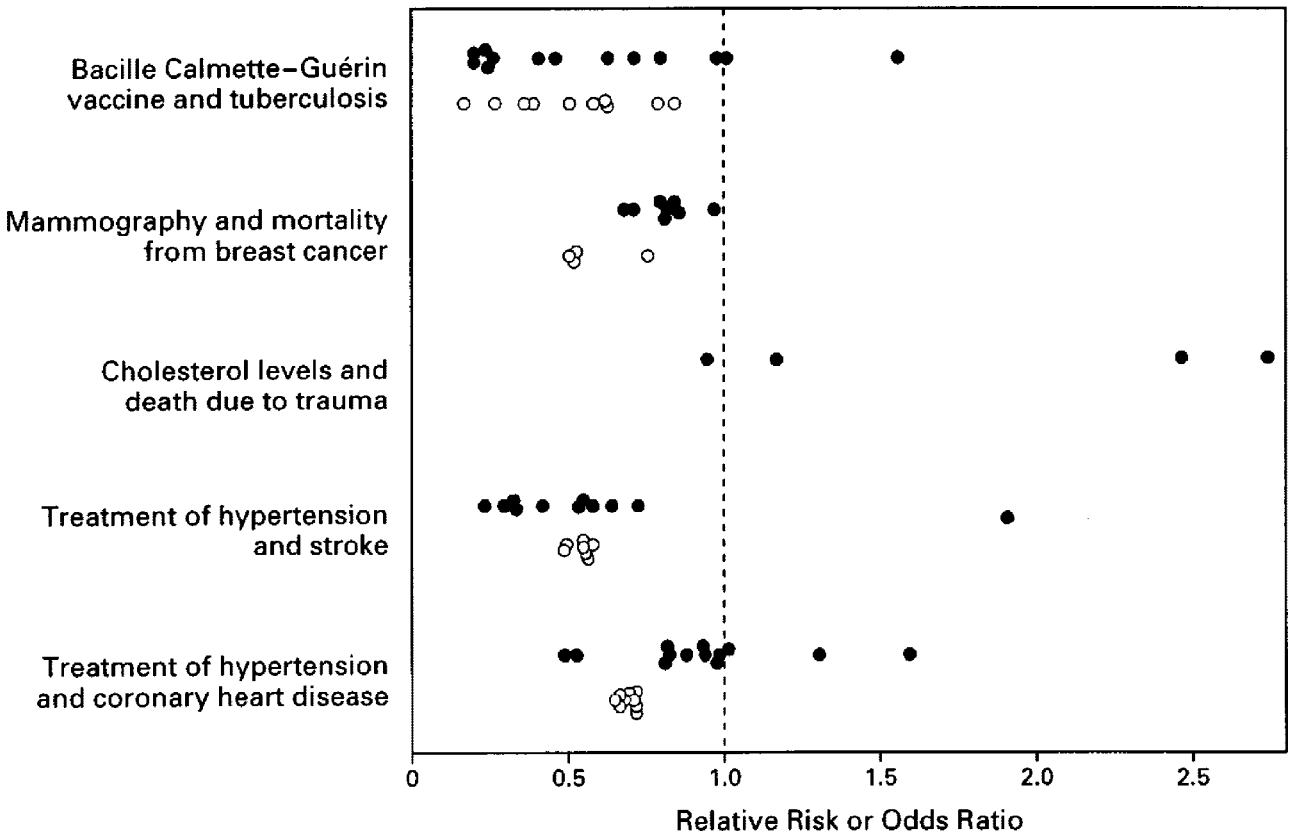

FIG. 1. Range of relative risks or odds ratios, based on the following types of research design: bacillus Calmette-Guerin vaccine and tuberculosis (13 randomized, controlled trials and 10 case-control studies), screening mammography and breast cancer mortality (eight randomized, controlled trials and four case-control studies), treatment of hyperlipidemia and traumatic death among men (four randomized, controlled trials and 14 cohort studies), treatment of hypertension and stroke among men (11 randomized, controlled trials and seven cohort studies), treatment of hypertension and coronary heart disease among men (13 randomized, controlled trials and nine cohort studies). Filled circles, randomized, controlled trials; open circles, observational studies. (Reproduced with permission.)

Another finding, also contrary to current perceptions, was that observational studies individually demonstrated less variability (heterogeneity) in point estimates compared to the variability in point estimates observed in randomized, controlled trials on the same topic (FIG. 1). Indeed, only among randomized, controlled trials did individual studies report results that were opposite to the direction of the pooled point estimate, representing a "paradoxical" finding (e.g., treatment of hypertension was associated with higher rates of coronary heart disease in several clinical trials).

One possible explanation for the finding that observational studies were less prone to heterogeneity in results (compared with randomized, controlled trials) is that each observational study is more likely to include a broad representation of the at-risk population. In addition, less opportunity exists for differences in the management of subjects "across" observational studies. For example, although general agreement exists that physicians do not use therapeutic agents in a uniform way, an observational study would generally include patients with a wider spectrum of severity (regarding the disease of interest), more comorbid ailments, and treatments that were tailored for each individual patient. In contrast, randomized, controlled trials may have distinct groups of patients based on specific inclusion and exclusion criteria, and the experimental protocol for therapy may not be representative of clinical practice. Therefore, randomized, controlled trials often have limited generalizability.

ADDITIONAL EVIDENCE AGAINST A RIGID HIERARCHY

At the time of our previous study,[9] investigations had already shown that observational cohort studies often produce results similar to those of randomized, controlled trials, when using similar criteria to assemble study participants and suitable methodological precautions. For example, an analysis of 18 randomized and nonrandomized studies in health services research found that treatment effects may differ based on research design but that "one method does not give a consistently greater effect than the other."[18] In that assessment, results were found to be most similar when exclusion criteria across studies were comparable, and when prognostic factors were accounted for in observational studies. In addition, a specific strategy used to strengthen observational studies (called a "restrictive cohort" design[19]) adapts principles of randomized, controlled trials to 1) identify a zero-time for determining patient eligibility and baseline prognostic risk, 2) use inclusion and exclusion criteria similar to clinical trials, 3) adjust for differences in baseline susceptibility for the outcome, and 4) use similar statistical strategies (e.g., intention-to-treat) as in randomized, controlled trials. When these procedures were used in a cohort study[19] evaluating the benefit of beta blockers after recovery from myocardial infarction, the restricted cohort produced results consistent with corresponding findings from the Beta-Blocker Heart Attack Trial.[20]

A second line of evidence supporting our contention that research design should not be considered a rigid hierarchy is also available in the literature of other scientific disciplines that carry out subject-based intervention trials. Examples include a comprehensive review of psychological, educational, and behavioral treatment research[21]; the findings from this review did not support a contention that observational studies overestimate effects relative to randomized, controlled trials.

Further evidence against a rigid hierarchy is based on results from the trials themselves. For example, a review of more than 200 randomized, controlled trials found numerous individual trials that were supportive, equivocal, or nonsupportive for each of 36 clinical topics.[22] Several publications have discussed various aspects of randomized, controlled trials in neurology.[23–28] Recent publications indicate that randomized, controlled trials continue to generate conflicting results, e.g., addressing the question of whether therapy with monoclonal antibodies improve outcomes among patients with septic shock.[29,30] In addition, results of "large, simple" randomized, controlled trials contribute to the evidence of contradictory results from randomized, controlled trials; one report found that results of meta-analyses based on randomized, controlled trials were often discordant with findings from large, simple trials on the same clinical topic.[31] Regardless of the reasons that individual 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 for all clinical situations.

EXAMPLES FROM THE LITERATURE AND IMPLICATIONS FOR CLINICAL CARE

Vitamin E and coronary heart disease

The Heart Outcomes Prevention Evaluation (HOPE) study,[32] a randomized, controlled trial, was cited as helping to "restrain earlier observational claims that vitamin E lowers the risk of cardiovascular disease."[33] A review of this topic illustrates the methodological issues involved. Several observational studies[34–36] found a "positive" association; in contrast, the HOPE study suggested that vitamin E has no effect on cardiovascular outcomes. Yet, a thorough examination of randomized, controlled trials on this topic provides a more complete assessment. Although two randomized, controlled trials[37,38] also found no effect on mortality, two other randomized, controlled trials[39,40] found decreased mortality associated

TABLE 3. *Foci for Comparison of Observational and Experimental Study Designs: Example of Vitamin E and Coronary Disease*

Patients	● Primary *versus* secondary prevention ● Presence or absence of comorbidity
Exposure	● Dietary intake *versus* supplements ● Dose and duration ● With or without co-therapy
Outcome	● Overall *versus* cause-specific mortality ● Morbidity ● Duration of follow-up ● Single *versus* combined endpoint

with vitamin E. Thus, data from clinical trials are themselves contradictory, and selecting one randomized, controlled trial as a gold standard to criticize observational studies is overly simplistic.

This clinical topic was used to support the statement that "...society expects us to evaluate new healthcare interventions by the most scientifically sound and rigorous methods available. Although observational studies often are cheaper, quicker, and less difficult to carry out, we should not lose sight of one simple fact: ignorance calls for careful experimentation. This means high-quality randomized, controlled trials, not observations that reflect personal choices and beliefs."[33] An alternative, more rigorous, and less dogmatic approach would be to compare published studies based on components of their research design, whether randomized or observational (Table 3), and not make a priori judgments regarding a single randomized, controlled trial constituting a gold standard.

Hormone replacement therapy and coronary heart disease

Another example of this controversy involves hormone replacement therapy disease for postmenopausal women. In summary, observational studies (such as the Nurses Health Study[41]) suggested a protective benefit of hormones; whereas randomized, controlled trials (including the Women's Health Initiative[42] and the Heart and Estrogen/Progestin Replacement Study[43]) pointed to no benefit, or even harm. Rather than assume the randomized, controlled trials inherently reveal "truth," potential explorations for the discordant findings could be explored. First, it should be noted that results of randomized, controlled trials and observational studies are remarkably consistent for most outcomes in studies of hormone replacement therapy, including stroke, breast cancer, colorectal cancer, hip fracture, and pulmonary embolism. The outcome of coronary artery disease has received most attention, and has been described as an anomaly.[44]

An assessment of this topic described plausible methodological and biological explanations for the differences in findings.[44] For example, available data indicate that women with higher socioeconomic status are more likely to be hormone replacement therapy users and less likely to have coronary artery disease, suggesting that the observational studies were vulnerable to "healthy user bias" (or "confounding") in this context. (Confounding, as a general term, occurs when a third variable, socioeconomic status in this situation, is related to both the exposure [hormone therapy] and outcome [coronary artery disease] variables for the association of interest. The exposure variable [hormone therapy] would then be described as a "marker" for the confounding variable, rather than actually causing the outcome.) In addition, the randomized, controlled trials themselves have been criticized for having bias.[45]

Another issue involves incomplete capture of early clinical events.[44] Observational studies typically enroll participants who have been taking hormone replacement therapy for some time, whereas randomized clinical trials initiate therapy in nonusers. Accordingly, clinical events that occur soon after initiating the medication would be captured by randomized, controlled trials, but typical observational studies assess what is likely to happen when patients remain on therapy for an extended period of time (patients initiating therapy recently would account for a very small proportion of the overall population). Other explanations for discordant results involve differences in protocols among observational studies and randomized, controlled trials. For example, daily combinations of estrogen and progestin were administered in Women's Health Initiative[42] and Heart and Estrogen/Progestin Replacement Study,[43] compared with estrogen alone or combined regimens for 10-14 days per month in observational studies such as the Nurses Health Study.[41]

These differences are not "fatal flaws" of observational studies, unless a rigid opinion is adopted that designates randomized, controlled trials as infallible. Most of the issues raised involve either methodological differences without a definite "winner" (e.g., examining early *vs* late clinical events), or true biological differences (e.g., in patients or protocols). Regarding the issue of confounding (e.g., healthy user bias, as described previously), methods are available[19] to measure and adjust for such variables.

A MORE BALANCED VIEW OF OBSERVATIONAL AND EXPERIMENTAL EVIDENCE

Given that randomized, controlled trials have not and often cannot be done for many clinical interventions, much of the clinical care provided in neurology (and all other specialties in medicine) would necessarily be considered unsubstantiated, if observational studies are discounted from consideration. The available evidence suggests, however, that observational studies can be

conducted with sufficient rigor to replicate the results of randomized, controlled trials. The key issue is designing appropriate observational studies, usually with suitable (observational) cohort or case-control architecture; a methodological task for investigators to complete and reviewers to evaluate.

Despite the consistency of our results[9] (involving five clinical topics and 99 separate studies), as well as confirmatory evidence available in the literature,[16–18] we believe that the role of observational studies may vary in different situations. For example, different exposures (e.g., surgical operations and other invasive therapies) may be more prone to selection bias in observational investigations than the drugs and noninvasive tests examined in our report,[9] and "softer" outcomes (e.g., functional status) may be assessed more readily in randomized, controlled trials. In addition, we emphasized the potential risk associated with poorly done observational studies; for example, to promote ineffective "alternative" therapies.[46]

Finally, a point of emphasis involves the general belief that randomization is necessary to balance known and (especially) unknown potential factors that can cause biased estimates of treatment effects through confounding. Given that unknown factors, by definition, would not be recognized by clinicians, a bias in assigning treatment would not occur according to those factors. Although such factors could be associated with outcome, they would not be associated with exposure, and therefore would not be confounding variables and would not affect the validity of results.

Randomized, controlled trials will (and should) remain a prominent tool in clinical research, but the results of a single randomized, controlled trial, or only one observational study, should be interpreted cautiously. If a randomized, controlled trial is later determined to be "wrong" in its conclusions, evidence from both other trials and well designed cohort or case-control studies can and should be used to establish the "right" answers.

CONCLUSION

The issues raised in this paper are not intended to diminish the important role that randomized, controlled trials play in clinical medicine (e.g., for evaluating interventions or for satisfying regulatory criteria). Yet, the popular belief that randomized, controlled trials inherently produce gold standard results, and that all observational studies are inferior, does a disservice to patient care, clinical investigation, and education of health care professionals. We should recognize the potential problem we face, that "the justification for why studies are included or excluded from the evidence base can rest on competing claims of methodologic authority that look little different from the traditional claims of medical authority that proponents of evidence-based medicine have criticized...interpretive decisions by old pre-evidence-based medicine experts may be replaced by interpretive decisions from a new group of experts with evidence-based medicine credentials..."[47] A more balanced and scientifically justified approach is to evaluate the strengths and limitations of well done experimental and observational studies, recognizing the attributes of each type of design.

REFERENCES

1. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 268:2420–2425, 1992.
2. 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* [Erratum 275:1232, 1996] 274:1800–1804, 1995.
3. U. S. Preventive Services Task Force. Guide to clinical preventive services, Ed 2, p 862. Baltimore: U. S. Preventive Services Task Force, 1996.
4. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellengerg JH et al. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 295:74–80, 1976.
5. Feinstein AR. Current problems and future challenges in randomized clinical trials. *Circulation* 70:767–774, 1984.
6. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 52:487–497, 1999.
7. Sacks H, Chalmers TC, Smith HJ. Randomized versus historical controls for clinical trials. *Am J Med* 72:233–240, 1982.
8. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM, p 106. London: Churchill Livingstone, 2000.
9. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 342:1887–1892, 2000.
10. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J et al. Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 335:765–774, 1990.
11. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Feinberg HV et al. Efficacy of bacillus Calmette-Guerin vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 271:698–702, 1994.
12. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA* 273: 149–154, 1995.
13. Cummings P, Psaty BM. The association between cholesterol and death from injury. *Ann Intern Med* 120:848–855, 1994.
14. Jacobs D, Blackburn H, Higgins M, Reed D, Iso H, McMillan G et al. Report of the conference on low blood cholesterol: mortality associations. *Circulation* 86:1046–1060, 1992.
15. Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 335:827–838, 1990.
16. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 342:1878–1886, 2000.
17. 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* 51:81–91, 1998.
18. 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* 319:312–315, 1999.

19. Horwitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 89:630–638, 1990.
20. Beta-blocker heart attack trail research group. A randomized trial of propanolol in patients with acute myocardial infarction. *JAMA* 247:1707-1714, 1982.
21. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. *Am Psychol* 48:1181–1209, 1993.
22. Horwitz RI. Complexity and contradiction in clinical trial research. *Am J Med* 82:498–510, 1987.
23. Bates D. Practical problems in the organisation of clinical trials in multiple sclerosis. *Neuroepidemiology* 6:6–16, 1987.
24. McKhann GM. The trials of clinical trials. *Arch Neurol* 46:611–614, 1989.
25. Loeb C, Gandolfo C. Methodological problems of clinical trials in multi-infarct dementia. *Neuroepidemiology* 9:223–227, 1990.
26. Riggs JE, Hobbs GR. Clinical trials, outcomes, and statistics: how better can be worse. *Neurology* 51:1234–1235, 1998.
27. Chadwick D, Privitera M. Placebo-controlled studies in neurology: where do they stop? *Neurology* 52:682–685, 1999.
28. Glauser TA. Integrating clinical trial data into clinical practice. *Neurology* 58:S6–S12, 2002.
29. Horn KD. Evolving strategies in the treatment of sepsis and systemic inflammatory response syndrome (SIRS). *QJM* 91:265–277, 1998.
30. Angus DC, Birmingham MC, Balk RA, Scannon PJ, Collins D, Kruse JA et al. E5 murine monoclonal antiendotoxin antibody in gram-negative sepsis. *JAMA* 283:1723–1730, 2000.
31. LeLorier L, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large, randomized, controlled trials. *N Engl J Med* 337:536–542, 1997.
32. Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P. Vitamin E supplementation and cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 342:154–160, 2000.
33. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 342:1907–1909, 2000.
34. Stampfer MJ, Hennekens CH, Manson JE, Colditz GA, Rosner B, Willett WC. Vitamin E consumption and the risk of coronary disease in women. *N Engl J Med* 328:1444–1449, 1993.
35. Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary disease in men. *N Engl J Med* 328:1450–1456, 1993.
36. Hodis HN, Mack WJ, LaBree L, Cashin-Hemphill L, Sevanian A, Johnson R et al. Serial coronary angiographic evidence that antioxidant vitamin intake reduces progression of coronary artery atherosclerosis. *JAMA* 273:1849–1854, 1995.
37. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 330:1029–1035, 1994.
38. GISSI-Prevenzione Investigators (Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico). Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. *Lancet* 354:447-455, 1999.
39. Blot WJ, Li JY, Taylor PR, Guo W, Dawsey S, Wang GQ et al. Nutrition intervention trials in Linxian, China: supplementation with specific vitamin/mineral combinations, cancer incidence, and disease-specific mortality in the general population. *J Natl Cancer Inst* 85:1483–1492, 1993.
40. Stephens NG, Parsons A, Schofield PM, Kelly F, Cheeseman K, Michinson MJ. Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study. *Lancet* 347:781–786, 1996.
41. Grodstein F, Manson JE, Colditz GA, Willett WC, Speizer FE, Stampfer MJ. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med* 133:933–941, 2000.
42. Writing Group for the Women's Health Initiative Investigators. Risk and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 288:321–333, 2002.
43. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 280:605–613, 1998.
44. Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal therapy. *N Engl J Med* 348:645–650, 2003.
45. Garbe E, Suissa S. Hormone replacement therapy and acute coronary outcomes: methodological issues between randomized and observational studies. *Hum Reprod* 19:8–13, 2004.
46. Angell M, Kassirer JP. Alternative medicine—the risks of untested and unregulated remedies. *N Engl J Med* 339:839–841, 1998.
47. Goodman SN, The mammography dilemma: a crisis for evidence-based medicine? *Ann Intern Med* 137:363–364, 2002.