Special Article

DISCREPANCIES BETWEEN META-ANALYSES AND SUBSEQUENT LARGE RANDOMIZED, CONTROLLED TRIALS

JACQUES LELORIER, M.D., PH.D., GENEVIÈVE GRÉGOIRE, M.D., ABDELTIF BENHADDAD, M.D., JULIE LAPIERRE, M.D., AND FRANÇOIS DERDERIAN, M.SC.

ABSTRACT

Background Meta-analyses are now widely used to provide evidence to support clinical strategies. However, large randomized, controlled trials are considered the gold standard in evaluating the efficacy of clinical interventions.

Methods We compared the results of large randomized, controlled trials (involving 1000 patients or more) that were published in four journals (the *New England Journal of Medicine,* the *Lancet,* the *Annals of Internal Medicine,* and the *Journal of the American Medical Association*) with the results of meta-analyses published earlier on the same topics. Regarding the principal and secondary outcomes, we judged whether the findings of the randomized trials agreed with those of the corresponding meta-analyses, and we determined whether the study results were positive (indicating that treatment improved the outcome) or negative (indicating that the outcome with treatment was the same or worse than without it) at the conventional level of statistical significance ($P<0.05$).

Results We identified 12 large randomized, controlled trials and 19 meta-analyses addressing the same questions. For a total of 40 primary and secondary outcomes, agreement between the meta-analyses and the large clinical trials was only fair (kappa = 0.35; 95 percent confidence interval, 0.06 to 0.64). The positive predictive value of the meta-analyses was 68 percent, and the negative predictive value 67 percent. However, the difference in point estimates between the randomized trials and the meta-analyses was statistically significant for only 5 of the 40 comparisons (12 percent). Furthermore, in each case of disagreement a statistically significant effect of treatment was found by one method, whereas no statistically significant effect was found by the other.

Conclusions The outcomes of the 12 large randomized, controlled trials that we studied were not predicted accurately 35 percent of the time by the meta-analyses published previously on the same topics. (N Engl J Med 1997;337:536-42.)

LARGE randomized, controlled trials are generally considered the gold standard in evaluations of the efficacy of clinical interventions. However, since such trials are not always available, clinicians increasingly rely on meta-analysis to support their choice of clinical strategies. Critics have emphasized the intrinsic weaknesses of meta-analysis.[1-5] Pooled results incorporate the biases of individual studies and embody new sources of bias, mostly because of the selection of studies and the inevitable heterogeneity among them.

Although much has been said about the strengths and weaknesses of meta-analysis, there are limited data systematically comparing the results of meta-analyses of several small trials with those of large randomized, controlled trials. Villar et al.[6] reviewed 30 meta-analyses of various interventions in perinatal medicine from the Cochrane data base. They recalculated the results of each meta-analysis after removing the largest trial from the analysis and then compared the results with those of the large trial that had been removed. They found a kappa of 0.46 to 0.53 and a positive predictive value of 50 to 67 percent. We compared the results of a series of systematically compiled large randomized, controlled trials with those of the relevant meta-analyses that had been published previously.

METHODS

Data Base

We searched the *New England Journal of Medicine*, the *Lancet*, the *Annals of Internal Medicine*, and the *Journal of the American Medical Association* and retrieved all large randomized, controlled trials (those in which 1000 patients or more were studied) that were published between January 1, 1991, and December 31, 1994. All the trials had to have adequate statistical power to detect the desired benefit specified by the authors. Adequate power was defined on the basis of the a priori calculations of power reported by the authors in the Methods sections of their articles.

From the Research Center, Hôtel-Dieu de Montréal Hospital, and the Department of Medicine, Faculty of Medicine, University of Montreal — both in Montreal. Address reprint requests to Dr. LeLorier at the Research Center, Hôtel-Dieu de Montréal, 3850 St. Urbain St., Pavilion Marie de la Ferre, 2nd Fl., Montreal, QC H2W 1T8, Canada.

We then searched for meta-analyses of similar topics that had been published before the large randomized, controlled trial. Our search included the references listed in the randomized trials and computerized searches of Medline without language restrictions. We then compared each trial with the set of meta-analyses corresponding to it and selected only those meta-analyses that coincided with the trial with regard to the similarity of the populations studied, the therapeutic intervention, and at least one outcome. We studied the principal and secondary outcomes.

For each outcome that was studied in both the large randomized, controlled trial and the meta-analysis, we determined whether the results were positive (indicating that treatment resulted in a better outcome) or negative (indicating that treatment resulted in an equal or worse outcome) at the conventional level of statistical significance ($P<0.05$). Two investigators working independently of each other reviewed each trial and its corresponding meta-analyses. Discrepancies were resolved by consensus, with the help of a third investigator. To quantify the effect of interobserver variation, we performed a sensitivity analysis; the statistical calculations were performed with the data obtained by consensus and were repeated with the data that corresponded to the opinion of the dissenting investigator.

Statistical Analysis

Two-by-two tables were used to calculate the degree of agreement between the large randomized, controlled trial and its associated meta-analysis as expressed by the kappa statistic and its 95 percent confidence interval, as well as the sensitivity, specificity, positive predictive value, and negative predictive value. The point estimates in each pair were compared by using a test statistic constructed as the difference in the proportions or means divided by the square root of the sum of the variances.

The odds ratios of the randomized, controlled trial and the meta-analysis were represented graphically. When the result of the meta-analysis was not presented as an odds ratio for a dichotomous outcome, we computed the odds ratio and its 95 percent confidence interval by the fixed-effects Mantel–Haenszel method.[7] When no odds ratio could be computed for a meta-analysis that represented the size of the treatment effect, we transformed the odds ratio in the corresponding randomized, controlled trial into an effect size by treating the proportion for each group as the mean of a distribution of 0's and 1's.[8] These transformations were made only to permit graphic representation and did not affect the P values reported in the corresponding papers. Figure 1 shows the odds ratios computed by the fixed-effects method, and Figure 2 shows the effect sizes obtained by transformation of the odds ratios. P values of less than 0.05 were considered to indicate statistical significance. All the calculations and statistical tests were done with the SAS statistical package (SAS Institute, Cary, N.C.).

RESULTS

We identified 12 large randomized, controlled trials to which 19 meta-analyses corresponded in terms of the populations studied, the therapeutic interventions, and at least one outcome. Since both the primary and the secondary outcomes were considered, a total of 40 outcomes coincided and were included in the analysis.

Table 1 shows the data on which we based our evaluation of the performance of meta-analysis as a predictor of the results of subsequent large randomized, controlled trials. The meta-analysis occupied the role usually assigned to a diagnostic test being assessed, whereas the trial was considered the gold standard. Table 2 shows the results in terms of sensitivity, specificity, and negative and positive predictive values. The results for the consensus opinion are all in a range of values (65 to 70 percent) that corresponds to the values usually obtained in average diagnostic tests. The kappa statistic, which measures agreement beyond that due to chance alone, was 0.35 (95 percent confidence interval, 0.06 to 0.64). Kappa values at or below 0.40 are considered to represent fair-to-slight agreement. Table 2 also shows the results of the sensitivity analysis, which compares the results obtained when the calculations were made on the basis of the consensus between investigators with those obtained when the calculations were based on the dissenting investigator's opinion.

Figures 1 and 2 show the results graphically and include the most pertinent information about each cluster of comparisons. They show that independently of their statistical significance, the point estimates were on the same side of 1.0 in Figure 1 and on the same side of 0 in Figure 2 in 32 of the 40 comparisons (80 percent). No situation was found in which the point estimates were both statistically significant and on opposite sides of 1.0 or 0. All the disagreements thus occurred because one result showed a statistically significant treatment effect, whereas the other indicated that such an effect was lacking. There was a statistically significant difference between the randomized clinical trial and the meta-analysis in 5 of the 40 comparisons (12 percent).

Five positive outcomes from four meta-analyses[10,28,31,37] that used fixed-effects models were followed by negative randomized clinical trials. We had the information needed to redo the statistical analyses with random-effects models for four of these outcomes,[10,28,31] and the results in all four remained statistically significant.

We found very good agreement between the meta-analysis and the randomized clinical trial with regard to the following six clinical matters: the effect of magnesium on overall mortality in patients with myocardial infarction,[12,13] the effect of treatment for hypercholesterolemia on coronary events and mortality from cardiovascular causes among patients with coronary heart disease,[14-16] the effect of vitamin A supplementation on mortality from all causes and mortality from diarrhea among children in developing countries,[18-20] the effect of angiotensin-converting–enzyme inhibitors on the mortality of patients with congestive heart failure,[21,22] the effect of adjuvant therapy on disease-free survival in patients with breast cancer,[32,33] and the value of multiple interventions as compared with single interventions in smoking cessation.[38,39]

Considerable divergence was evident in several other cases. With regard to the effects of late thrombolysis (thrombolysis performed at least six hours after the first symptoms of myocardial infarction)[9-11] and nitroglycerin on mortality in patients with myocardial infarction, the meta-analyses were positive, where-

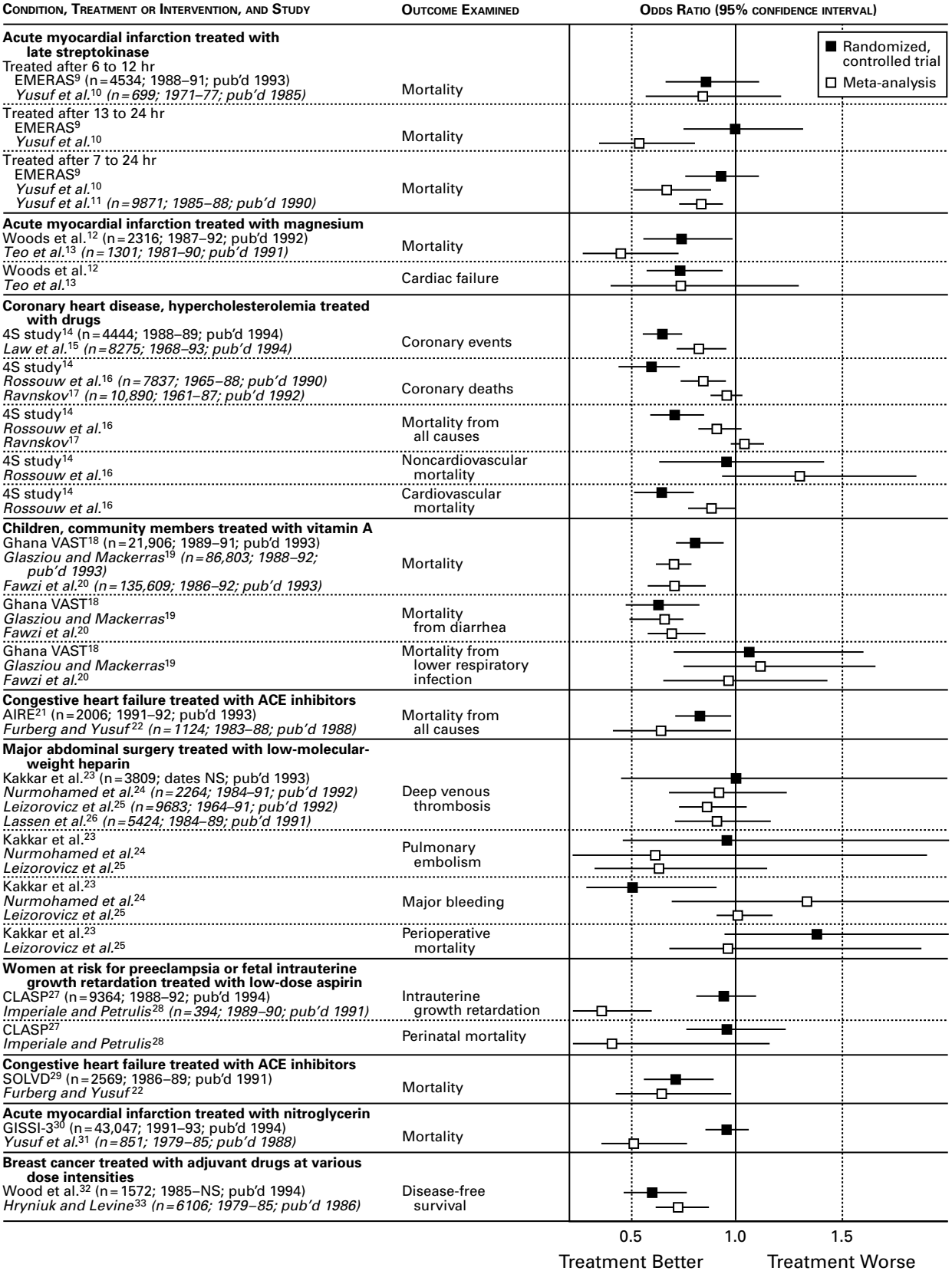

Figure 1. Odds Ratios and 95 Percent Confidence Intervals for Clusters of Studies in Which the Findings of Large Randomized, Controlled Trials Were Compared with the Results of One or More Meta-Analyses on the Same Subject, in Which at Least One Common Outcome Was Studied.

Each randomized trial and its associated meta-analyses are separated from the others by a solid horizontal line. Dashed lines delineate each cluster of trials and meta-analyses in which a single outcome was examined. The solid squares at right are the point estimates (odds ratios) for the randomized trials, and the open squares are the odds ratios for the meta-analyses. The bars on either side of the squares are 95 percent confidence intervals.

The vertical line indicating an odds ratio of 1.0 is the line at which treatment was found to have no effect. Odds ratios to the left of that line (lower than 1.0) indicate that outcome was better with treatment; those to the right (higher than 1.0) indicate that outcome was worse. When the 95 percent confidence interval does not span the "no difference" line at 1.0, the study findings are considered to be significant ($P<0.05$).

Names of randomized, controlled trials are given in roman type, and names of meta-analyses are given in italics. For the randomized, controlled trials, the inclusive dates listed are the years when the first and last patients were enrolled; for the meta-analyses, the dates are the years when the first and last papers were published. Pub'd denotes published, ACE angiotensin-converting enzyme, and NS not specified.

as the results of the subsequent large randomized, controlled trials were on the positive side of 1.0 but were not statistically significant. In these instances statistical power could not have been the issue, because the randomized, controlled trials included more patients than the meta-analyses. With regard to the question of preventing intrauterine growth retardation with low-dose aspirin in women at risk of preeclampsia, a clearly positive meta-analysis[28] with only 394 patients was followed by a very large randomized, controlled trial with 9364 patients that had negative results.[27] Despite a negative meta-analysis,[35] a large randomized, controlled trial[34] showed that sodium reduction decreases diastolic blood pressure, whereas in the case of calcium supplementation the reverse occurred.[34,37]

Since the decision to conduct a large randomized, controlled trial could have been made when clinicians and researchers saw a meta-analysis as inconclusive, we examined whether the meta-analysis had already been published at the time the first patient was randomized in the corresponding clinical trial. Four of the 12 trials[9,21,30,38] had evidently been started and most probably designed after the publication of the corresponding meta-analysis. Of these four trials, two[9,30] (evaluating the merits of thrombolysis and treatment with nitroglycerin) had results that diverged from those of the meta-analysis — that is, a negative randomized, controlled trial did not confirm the findings of a positive meta-analysis.

DISCUSSION

Few will disagree with the use of the large randomized, controlled trial as the gold standard in the evaluation of the efficacy of therapeutic interventions. All the meta-analyses except one that were found by our process of systematic research had been published in major peer-reviewed journals, where they were in a position to influence clinical practice.

The strategy we used to decide whether a given meta-analysis corresponded to a specific randomized, controlled trial raises certain methodologic issues. For the studies to qualify, the population studied, the therapeutic intervention, and at least one outcome had to be similar. In some cases, such similarity could involve judgment and thus be subject to variation between observers. By having two investigators decide independently on the appropriateness of each match, we could quantify the variation and adjust for it. The sensitivity analysis (Table 2) shows that our findings were essentially the same both when the calculations were based on consensus and when they were based on the opinion of the dissenting investigator. Another methodologic issue is raised by the dichotomous classification of the results as positive or negative. The reason for choosing this approach was that the outcome of interest was whether the results of the meta-analysis should be applied to clinical practice. Clinical decisions tend to be dichotomous in that a treatment is said either to work and be recommended or not to work and not to be recommended.

According to our analysis, if there had been no subsequent randomized, controlled trial, the meta-analysis would have led to the adoption of an ineffective treatment in 32 percent of cases (100 percent minus the positive predictive value) and to the rejection of a useful treatment in 33 percent of cases (100 percent minus the negative predictive value). It is important to recognize that these measures of disagreement, which are constructed from the perspective of medical decision making, tend to overstate the degree of statistical discrepancy. This is evident from the fact that in no case was there a divergence in which the randomized clinical trial and the meta-analysis gave statistically significant and opposite answers. Furthermore, wherever the point estimates were located in relation to the "no difference" line, the difference in results between the meta-analysis and the randomized, controlled trial was statistically significant for only 5 of the 40 comparisons (12 percent); this does not appear to be a large percentage, since a divergence in 5 percent of cases would be expected on the basis of chance alone.

In our study, 46 percent of the divergences in results involved a positive meta-analysis followed by a negative randomized, controlled trial. There are sev-

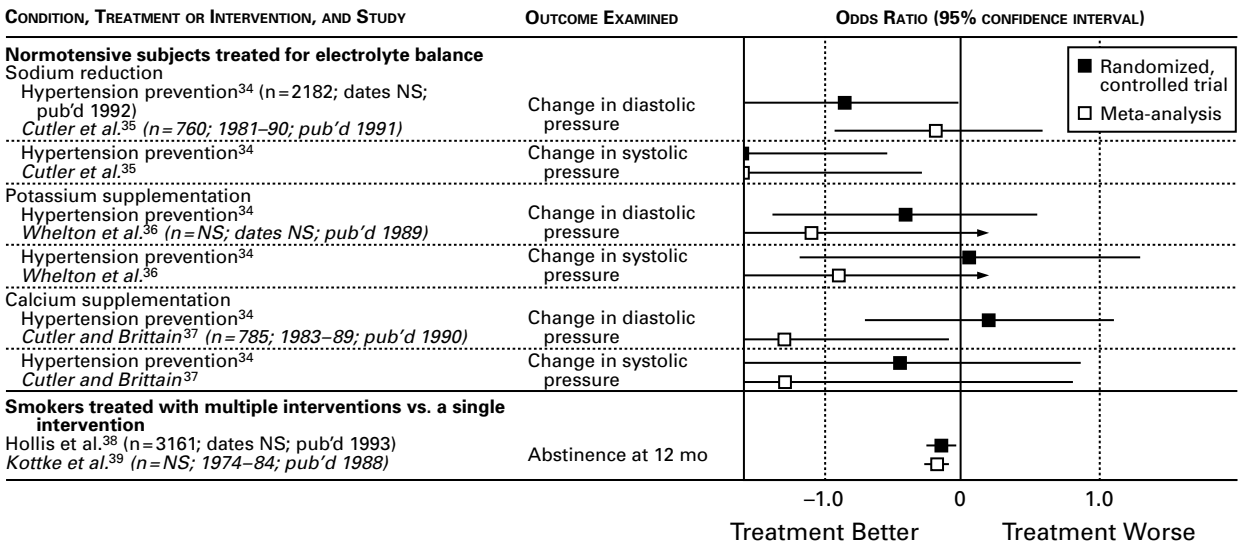

Figure 2. Treatment Effects and 95 Percent Confidence Intervals after Transformation of the Odds Ratios in Clusters of Studies in Which the Findings of Large Randomized, Controlled Trials Were Compared with the Results of One or More Meta-Analyses on the Same Subject, in Which at Least One Common Outcome Was Studied.

Each randomized trial and its associated meta-analysis are separated from the others by a solid horizontal line. Dashed lines delineate each cluster of trials and meta-analyses in which a single outcome was examined. The solid squares at right are the point estimates (effect sizes) for the randomized trials, and the open squares are the effect sizes for the meta-analyses (calculated as described in the Methods section). The bars on either side of the squares are 95 percent confidence intervals. Arrows mean that the differences were not statistically significant but the 95 percent confidence intervals could not be determined.

The vertical line indicating an effect size of 0 is the line at which treatment was found to have no effect. Odds ratios to the left of that line (lower than 0) indicate that outcome was better with treatment; those to the right (higher than 0) indicate that outcome was worse. When the 95 percent confidence interval does not span the "no difference" line at 0, the study findings are considered to be significant ($P<0.05$).

Names of randomized, controlled trials are given in roman type, and names of meta-analyses are given in italics. For the randomized, controlled trials, the inclusive dates listed are the years when the first and last patients were enrolled; for the meta-analyses, the dates are the years when the first and last papers were published. NS denotes not specified, and pub'd published.

eral reasons why a meta-analysis might have positive results that would not be confirmed by a subsequent trial. Publication bias refers to the tendency of investigators to preferentially submit studies with positive results for publication, and the tendency of editors to accept them. A meta-analysis that excluded unpublished studies or did not locate and include them would thus be more likely to have a false positive result. The systematic exclusion of papers written in languages other than English (the "Tower of Babel" bias[40]) can add to the publication bias. In our sample, the use of the fixed-effects model, which narrows the confidence interval, does not appear to account for the statistically positive meta-analyses whose findings were not subsequently confirmed by a randomized trial, since the four studies that could be reanalyzed by the random-effects model remained positive and continued to have statistically significant results when that reanalysis was done.

The remaining 54 percent of identified divergences involved a negative meta-analysis followed by a positive randomized, controlled trial. The heterogeneity of the trials included in the meta-analysis may partially account for divergence of this type, since meta-analysis assumes that such variation is mostly caused by random error, rather than by differences in the characteristics of the selected studies. A properly done meta-analysis involves the a priori determination of strict standards to ensure that the criteria used for the inclusion of patients, the administration of the principal treatment, and the ascertainment of outcome events are similar in all the trials selected. Although according to these strict criteria the protocols of the selected trials look very similar, their application usually yields very different products. The patients enrolled in comparable trials may belong to the same basic population, but even small differences in the criteria for diagnosis, coexisting conditions, severity of disease, and age will produce very different groups of patients. Differences in doses, time to onset, and duration of therapies can also produce substantial disparity among trials that are included in meta-analyses with the intention of evaluating a therapeutic intervention. The choice of concomitant

TABLE 1. AGREEMENT OR DISAGREEMENT BETWEEN RANDOMIZED, CONTROLLED TRIALS AND META-ANALYSES IN 40 CASES IN WHICH THE TWO WERE COMPATIBLE.*

RESULTS OF META-ANALYSIS	RESULTS OF RANDOMIZED, CONTROLLED TRIAL		
	POSITIVE	NEGATIVE	TOTAL
Positive	13	6	19
Negative	7	14	21
Total	20	20	40

*Positive indicates that the outcome of treatment was significantly better ($P<0.05$) than the outcome of no treatment, and negative indicates that the outcome of treatment was worse or the same.

TABLE 2. VARIABLES MEASURING THE ABILITY OF META-ANALYSES TO PREDICT THE RESULTS OF LARGE RANDOMIZED, CONTROLLED TRIALS.

VARIABLE	BASIS OF CALCULATION*	
	CONSENSUS OPINION	DISSENTING OPINION ONLY
Sensitivity — %	65	60
Specificity — %	70	75
Positive predictive value — %	68	71
Negative predictive value — %	67	65
Kappa value (95 percent confidence interval)†	0.35 (0.06–0.64)	0.35 (0.06–0.64)

*The sensitivity analysis compared the results of two calculations, one based on consensus between two investigators and the other based on the opinion of the dissenting observer. The two observers disagreed 5 percent of the time.

†A positive kappa value indicates that there is more than chance agreement. A value of 0.61 or above denotes substantial agreement, a value between 0.41 and 0.60 moderate agreement, and a value of 0.40 or below fair-to-slight agreement.

therapies and the degree of leeway in their administration can also affect the results. Changes in medical practice over time may also account for important differences in concomitant therapies, since the trials included in a given meta-analysis are often conducted over a period of a decade or more.

How should clinicians use meta-analyses, given that systematic comparison with randomized clinical trials shows that they have poor predictive ability? Most will agree that if a large, well-done randomized trial has been conducted, practice guidelines should be strongly influenced by its results. The question arises when the only available evidence is from a series of small randomized, controlled trials. The simplest solution, and currently the most popular one, has been to rely on the results of a meta-analysis. Our findings seem to indicate that summarizing all the information contained in a set of trials into a single odds ratio may greatly oversimplify an extremely complex issue. The popularity of meta-analysis may at least partly come from the fact that it makes life simpler and easier for reviewers as well as readers. However, oversimplification may lead to inappropriate conclusions.

The result of this study would appear to encourage readers to go beyond the point estimates and confidence intervals that represent the aggregate findings of a meta-analysis and, as Cook et al. have suggested,[41] look carefully at the studies that were included and evaluate the consistency of their results. When the results are mostly on the same side of the no-difference line, the meta-analysis merits more confidence. Others may consider following the advice of Horwitz[42] and appraising each trial separately. Although such an approach is admittedly more laborious, it has the advantage of allowing pragmatic clinicians to benefit from the diversity of studies by distinguishing the effects of treatment among them.

We are indebted to Dr. Jean-François Boivin for helpful criticisms and to Ms. Hélène Harnois and Ms. Anita Massicotte for clerical assistance.

REFERENCES

1. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71-9.
2. Horwitz RI. "Large-scale randomized evidence: large, simple trials and overviews of trials": discussion: a clinician's perspective on meta-analyses. J Clin Epidemiol 1995;48:41-4.
3. Bailar JC III. The practice of meta-analysis. J Clin Epidemiol 1995;48: 149-57.
4. Meinert CL. Meta-analysis: science or religion? Control Clin Trials 1989;10:Suppl:257S-263S.
5. Abrahamson JH. Meta-analysis: a review of pros and cons. Public Health Rev 1990/91;18:1-47.
6. Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. Lancet 1995;345:772-6.
7. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-48.
8. McNemar Q. Psychological statistics. 3rd ed. New York: John Wiley, 1962.
9. EMERAS (Estudio Multicéntrico Estreptoquinasa Repúblicas de América del Sur) Collaborative Group. Randomised trial of late thrombolysis in patients with suspected acute myocardial infarction. Lancet 1993; 342:767-72.
10. Yusuf S, Collins R, Peto R, et al. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. Eur Heart J 1985;6:556-85.
11. Yusuf S, Sleight P, Held P, McMahon S. Routine medical management of acute myocardial infarction: lessons from overviews of recent randomized controlled trials. Circulation 1990;82:Suppl II:II-117–II-134.
12. Woods KL, Fletcher S, Roffe C, Haider Y. Intravenous magnesium sulphate in suspected acute myocardial infarction: results of the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2). Lancet 1992;339:1553-8.
13. Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. BMJ 1991;303:1499-503.
14. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 1994;344:1383-9.

15. 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;308:367-72.
16. Rossouw JE, Lewis B, Rifkind BM. The value of lowering cholesterol after myocardial infarction. N Engl J Med 1990;323:1112-9.
17. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992;305:15-9. [Erratum, BMJ 1992;305:505.]
18. Ghana VAST Study Team. Vitamin A supplementation in northern Ghana: effects on clinic attendances, hospital admissions, and child mortality. Lancet 1993;342:7-12. [Erratum, Lancet 1993;342:250.]
19. Glasziou PP, Mackerras DE. Vitamin A supplementation in infectious diseases: a meta-analysis. BMJ 1993;306:366-70.
20. Fawzi WW, Chalmers TC, Herrera MG, Mosteller F. Vitamin A supplementation and child mortality: a meta-analysis. JAMA 1993;269:898-903.
21. The Acute Infarction Ramipril Efficacy (AIRE) Study Investigators. Effect of ramipril on mortality and morbidity of survivors of acute myocardial infarction with clinical evidence of heart failure. Lancet 1993;342: 821-8.
22. Furberg CD, Yusuf S. Effect of drug therapy on survival in chronic congestive heart failure. Am J Cardiol 1988;62:41A-45A.
23. Kakkar VV, Cohen AT, Edmonson RA, et al. Low molecular weight versus standard heparin for prevention of venous thromboembolism after major abdominal surgery: the Thromboprophylaxis Collaborative Group. Lancet 1993;341:259-65.
24. Nurmohamed MT, Rosendaal FR, Buller HR, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. Lancet 1992;340:152-6.
25. Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. BMJ 1992;305:913-20.
26. Lassen MR, Borris LC, Christiansen HM, et al. Clinical trials with low molecular weight heparins in the prevention of postoperative thromboembolic complications: a meta-analysis. Semin Thromb Hemost 1991;17: Suppl 3:284-90.
27. CLASP (Collaborative Low-dose Aspirin Study in Pregnancy) Collaborative Group. CLASP: a randomised trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. Lancet 1994;343:619-29.
28. Imperiale TF, Petrulis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. JAMA 1991;266: 260-4.
29. The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991;325:293-302.
30. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Lancet 1994;343:1115-22.
31. Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. Lancet 1988;1:1088-92.
32. Wood WC, Budman DR, Korzun AH, et al. Dose and dose intensity of adjuvant chemotherapy for stage II, node-positive breast carcinoma. N Engl J Med 1994;330:1253-9. [Erratum, N Engl J Med 1994;331:139.]
33. Hryniuk W, Levine MN. Analysis of dose intensity for adjuvant chemotherapy trials in stage II breast cancer. J Clin Oncol 1986;4:1162-70.
34. The Trials of Hypertension Prevention Collaborative Research Group. The effects of nonpharmacologic interventions on blood pressure of persons with high normal levels: results of the Trials of Hypertension Prevention, phase I. JAMA 1992;267:1213-20. [Erratum, JAMA 1992;267: 2330.]
35. Cutler JA, Follmann D, Elliott P, Suh I. An overview of randomized trials of sodium reduction and blood pressure. Hypertension 1991;17: Suppl:I27-I33.
36. Whelton PK, Thaker GK, Klag MJ, Seidler AJ. Blood pressure effects of potassium supplementation. Circulation 1989;80:Suppl II:II-301. abstract.
37. Cutler JA, Brittain E. Calcium and blood pressure: an epidemiologic perspective. Am J Hypertens 1990;3:137S-146S.
38. Hollis JF, Lichtenstein E, Vogt TM, Stevens JV, Biglan A. Nurse-assisted counseling for smokers in primary care. Ann Intern Med 1993; 118:521-5.
39. Kottke TE, Battista RN, DeFriese GH, Brekke ML. Attributes of successful smoking cessation interventions in medical practice: a meta-analysis of 39 controlled trials. JAMA 1988;259:2883-9.
40. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995;48:159-63.
41. Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. Chest 1995;108:Suppl:227S-230S.
42. Horwitz RI. Complexity and contradiction in clinical trial research. Am J Med 1987;82:498-510.

Correspondence

Meta-Analyses and Large Randomized, Controlled Trials

To the Editor: We were pleased to see that using an independent protocol, LeLorier et al. (Aug. 21 issue)[1] confirmed both our[2] previous estimates of the frequency of discrepancies between large trials and meta-analyses and those of Villar et al.[3] Their selection of 12 large trials from four influential journals may have inflated the frequency of apparent discrepancies. Such journals may tend to publish trials that are likely to change practice, whose results disagree with prior evidence.[4] Still, the estimates of LeLorier et al. are largely similar to prior estimates. However, we are concerned that several of their premises propagate outdated myths.

First, why is the latest single large trial always the gold standard against which all prior evidence (often including several large trials) must be measured? In 6 of the 12 cases discussed, the meta-analysis had more patients than the subsequent gold standard. Second, decision making based solely on which side of 0.05 the P value lies is potentially misleading; an odds ratio of 0.7 (95 percent confidence interval, 0.5 to 0.9; $P = 0.01$), although different in precision, is hardly discrepant with an odds ratio of 0.7 (95 percent confidence interval, 0.3 to 1.8; $P = 0.4$). The measure that LeLorier et al. use may misrepresent the true frequency of disagreement.

Third, even with appropriate measures, discrepancies between meta-analyses and large trials should be expected, given the variable characteristics and treatment responses in different persons, protocols, and populations. Not only are trials in meta-analyses frequently heterogeneous, but also the idea of the homogeneous single trial is often a myth. Discrepancies occur even within trials[5] and between large trials themselves, as studies of magnesium in myocardial infarction exemplify.[6] Meta-analysis has recently been evolving toward evaluating this heterogeneity. It is more constructive to quantify reasons for discrepancies[2] rather than wait for the latest larger and better trial that may nullify past experience. Unfortunately, LeLorier et al. did not explore such reasons systematically.

Fourth, potential biases exist in both meta-analyses and clinical trials. If nothing else, meta-analysis sensitizes us to several of these biases regarding the conduct and reporting of trials.[4] LeLorier and colleagues made use of such scientific advances to make their points. Meta-analysis is not statistical alchemy that makes life easier by distilling one magic number from confounded data; it is a scientific discipline that aims to quantify evidence and to explore bias and diversity in research systematically. We should keep trying to improve clinical trials and meta-analyses, not undermine them.

JOHN P.A. IOANNIDIS, M.D.
National Institute of Allergy and Infectious Diseases
Bethesda, MD 20892

JOSEPH C. CAPPELLERI, PH.D., M.P.H.
Pfizer Central Research
Groton, CT 06340

JOSEPH LAU, M.D.
New England Medical Center Hospitals
Boston, MA 02111

1. 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.
2. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 1996;276:1332-8.
3. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. Lancet 1995;345:772-6.
4. Ioannidis JP, Cappelleri JC, Sacks HS, Lau J. The relationship between study design, results, and reporting of randomized clinical trials of HIV infection. Control Clin Trials 1997;18:431-44.
5. Horwitz RI, Singer BH, Makuch RW, Viscoli CM. Can treatment that is helpful on average be harmful to some patients? A study of the conflicting information needs of clinical inquiry and drug regulation. J Clin Epidemiol 1996;49:395-400.
6. Woods KL. Mega-trials and management of acute myocardial infarction. Lancet 1995;346:611-4.

INSTRUCTIONS FOR LETTERS TO THE EDITOR

Letters to the Editor are considered for publication (subject to editing and abridgment) provided they do not contain material that has been submitted or published elsewhere. Please note the following: •Your letter must be typewritten and triple-spaced. •Its text, not including references, must not exceed 400 words (please include a word count). •It must have no more than five references and one figure or table. •It should not be signed by more than three authors. •Letters referring to a recent *Journal* article must be received within four weeks of its publication. •Please include your full address, telephone number, and fax number (if you have one). •You may send us your letter by post, fax, or electronic mail.

Our address: Letters to the Editor • *New England Journal of Medicine* • 10 Shattuck St. • Boston, MA 02115

Our fax numbers: 617-739-9864 and 617-734-4457

Our e-mail address: letters@nejm.org

We cannot acknowledge receipt of your letter, but we will notify you when we have made a decision about publication. We are unable to provide prepublication proofs. Please enclose a stamped, self-addressed envelope if you want unpublished material returned to you. Financial associations or other possible conflicts of interest must be disclosed. Submission of a letter constitutes permission for the Massachusetts Medical Society, its licensees, and its assignees to use it in the *Journal*'s various editions (print, data base, and optical disk) and in anthologies, revisions, and any other form or medium.

To the Editor: LeLorier et al. assume that the results of randomized, controlled trials correctly represent the true effect of an intervention and that the results of meta-analyses must be judged against this gold standard. This comparison, however, is not valid when there are major methodologic differences between the trials included in the meta-analysis and the subsequent randomized, controlled trial.

For example, the authors compare the results of a meta-analysis and a randomized, controlled trial that examined the efficacy of nitrates in patients with acute myocardial infarction. The meta-analysis, published in 1988,[1] found a benefit in terms of mortality from the use of nitrates (odds ratio, 0.65; 95 percent confidence interval, 0.51 to 0.82), but the randomized, controlled trial, published in 1994,[2] found no benefit (odds ratio, 0.94; 95 percent confidence interval, 0.84 to 1.05). However, both the interventions and the patient populations were markedly different. Patients in the meta-analysis were not treated with thrombolytic agents and were rarely treated with beta-blockers, and the control group had a high mortality rate (20.5 percent).[1] In contrast, patients in the randomized, controlled trial were intensively treated with multiple therapies (72 percent received thrombolytic agents and 31 percent received beta-blockers), and the mortality rate (6.9 percent) in the control group was much lower.[2] Rather than indicating that the meta-analysis is wrong, the findings suggest that nitrates decrease mortality only in patients who are not treated acutely with other therapies.

STEPHEN BENT, M.D.
KARLA KERLIKOWSKE, M.D.
DEBORAH GRADY, M.D., M.P.H.
University of California, San Francisco
San Francisco, CA 94143

1. Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. Lancet 1988;1:1088-92.
2. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Lancet 1994;343:1115-22.

To the Editor: . . . An overall estimate from a meta-analysis can be misleading if there is considerable heterogeneity among the included trials that has not been fully investigated. Similarly, it is misleading to compare the results of a single study with those of a meta-analysis without a careful examination of important characteristics of the patients and interventions included in these trials. Unfortunately, the study by LeLorier and colleagues, by giving the impression that the meta-analyses and the large trials were measuring the same thing, applies a simplistic analysis to a complex issue. These potentially misleading comparisons were seized on in the accompanying editorial (Aug. 21 issue)[1] to assert that a conventional narrative review is more reliable than a well-conducted meta-analysis, without providing any objective evidence to demonstrate the predictive accuracy of such narrative reviews. The reliability of large randomized, controlled trials, systematic reviews, meta-analyses, and narrative or ad hoc reviews and their respective roles in the field of clinical evaluation should be decided on the basis of careful scientific inquiry rather than prejudice.

FU-JIAN SONG, B.MED., M.MED., PH.D.
TREVOR A. SHELDON, M.SC.
University of York
York YO1 5DD, United Kingdom

1. Bailar JC III. The promise and problems of meta-analysis. N Engl J Med 1997;337:559-61.

To the Editor: The study by LeLorier et al. comparing the results of meta-analyses and subsequent large randomized, controlled trials illustrates the importance of exploring the heterogeneity of research evidence, a point noticeably missing from the editorial by Bailar. It would surely have been informative for LeLorier et al. to have explored the heterogeneity evident in Figure 1 of their article, particularly with respect to the methodologic quality, numbers of patients, and the length of follow-up. Instead, the authors chose to summarize the results in terms of predictive ability, a simplistic approach, particularly when correlated outcomes from within the same studies were included.

Both the article and the editorial highlight pitfalls that are only too well known to reviewers in the Cochrane Collaboration.[1] However, LeLorier et al. failed to provide information about how closely the meta-analyses followed Cochrane Collaboration guidelines,[1] among which are identifying unpublished studies, specifying whether data on individual patients or aggregate data were used, and revealing the way in which the quality of the original trial design was evaluated and whether heterogeneity between trial results was investigated. None of these points were mentioned by Bailar. Similarly, the only indication of the rigor of the large, randomized trials selected in the study by LeLorier et al. is provided by the journal in which they were published and the number of patients randomized, rather than by the mention of any previously published standards,[2] despite the description and use of such trials as the gold standard in evaluations of the efficacy of clinical interventions. Although recognizing the key role of rigorous, large, randomized, controlled clinical trials, we must not throw out the baby with the bath water, or fall prey to the biases inherent in conventional narrative review,[3] by dismissing systematic reviews and, when appropriate, meta-analysis.

SABOOR KHAN, F.R.C.S.
PAULA WILLIAMSON, PH.D.
ROBERT SUTTON, F.R.C.S.(GEN.)
University of Liverpool
Liverpool L69 3BX, United Kingdom

1. Oxman A. Preparing and maintaining systematic reviews. Cochrane Collaboration hand book. Oxford, England: Cochrane Collaboration, 1996:Section VI.
2. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. N Engl J Med 1982;306:1332-7.
3. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers CT. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. JAMA 1992;268:240-8.

To the Editor: LeLorier et al. restrictively searched for trials in four high-profile journals that can be very selective about publication. It is possible that the trials identified were submitted or published for the very reason that their effect sizes differed from those in previous meta-analyses, whereas trials closely confirming meta-analyses may have appeared in less prestigious journals. A less biased approach might have been to have conducted a similar analysis in which primary selection was applied to the meta-analyses and all journals were searched for subsequent trials. . . .

In our opinion, the editorial presents the biased viewpoint of a single person (much like a conventional review), illustrated by the statement that "when both the trial and the meta-analysis seem to be of good quality, . . . I tend to believe the results of the trial." On what basis? Support of narrative over systematic reviews is worrying. The problems of traditional review are numerous and have been well documented.[1] The expression of such an opinion in a *Journal* editorial is a step back in this era of evidence-based medicine.

LESLEY A. STEWART, PH.D.
MAHESH K.B. PARMAR, D.PHIL.
JAYNE F. TIERNEY, PH.D.
British Medical Research Council Cancer Trials Office
Cambridge CB2 2BW, United Kingdom

1. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987;106:485-8.

To the Editor: LeLorier et al. make a key assumption that the meta-analyses and the randomized trials were both estimating the same underlying effect. They attempted to adjust for any error in this assumption by performing a sensitivity analysis on the determination of similarity by the reviewers.

We believe that more advanced techniques of meta-analysis that explore specific sources of heterogeneity would provide additional insight into why the meta-analyses and their corresponding large trials did not observe the same outcomes. For example, techniques such as hierarchical Bayes'[1] and regression methods could be used to identify specific points on which the large trials and the individual trials in the meta-analyses differ, and to quantify the associations of these sources of heterogeneity with the observed outcomes. These analyses might therefore generate fruitful new directions for research.

When we acknowledge that meta-analysis is a method for studying studies rather than a shortcut for conducting large, randomized trials, we will begin to find the proper place for meta-analysis in our biostatistical toolbag.

IDA SIM, M.D.
PHILIP LAVORI, PH.D.
Palo Alto Veterans Health Care System
Palo Alto, CA 94304

1. DeMouchel WH, Harris JE. Bayes methods for combining the results of cancer studies in humans and other species. J Am Stat Assoc 1983;78:293-308.

To the Editor: Meta-analysis provides an opportunity to look for reasons for inconsistent results among studies, but LeLorier et al. mention only some hypothetical, generic reasons and overlook clinical information that might have explained the discrepant findings.

The discrepancies may be explained more by clinical heterogeneity and details of the study protocols and less by publication bias and analysis of random as opposed to fixed effects. For example, there was a statistically significant discrepancy between the meta-analysis[1] and the large randomized, controlled trial — the Collaborative Low-Dose Aspirin Study in Pregnancy (CLASP)[2] — involving low-dose aspirin for the prevention of intrauterine growth retardation. Eligibility criteria for the study (women at 12 to 32 weeks of gestation with a sufficient risk of preeclampsia or intrauterine growth retardation according to the responsible clinician) were vastly different from those of the meta-analysis (women with prior preeclampsia, intrauterine growth retardation, or placental infarction; primiparas with either increased blood pressure in response to angiotensin II or abnormal uteroplacental blood flow). This difference is reflected in the base-line risks of intrauterine growth retardation in the control groups: 6.6 percent (95 percent confidence interval, 6.2 to 7.0 percent) for women enrolled in the CLASP trial and 28 percent (range, 18 to 63 percent) for the study groups in the meta-analysis. This difference in risk by more than a factor of 4 exists despite the use of a less stringent definition of intrauterine growth retardation in the CLASP trial. Differences in the base-line risk of preeclampsia further highlight the heterogeneity: 7.6 percent (95 percent confidence interval, 6.8 to 8.1 percent) in the CLASP trial and 33 percent (range, 17 to 52 percent) in the meta-analysis.

The difference in the base-line risks of intrauterine growth retardation and preeclampsia, despite an offsetting difference in the criteria for intrauterine growth retardation, is a plausible explanation for the discrepant results. In reality, the CLASP trial and meta-analysis results are not necessarily discrepant, but they may reflect a variation in the effect of treatment with low-dose aspirin as a function of the risk of intrauterine growth retardation.

Without a careful consideration of clinical homogeneity, the work of LeLorier et al. has the same limitations as meta-analyses that do not carefully consider the clinical aspects of data synthesis.

THOMAS F. IMPERIALE, M.D.
Indiana University School of Medicine
Indianapolis, IN 46202

1. Imperiale TF, Petrulis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. JAMA 1991;266:260-4.
2. CLASP (Collaborative Low-Dose Aspirin Study in Pregnancy) Collaborative Group. CLASP: a randomised trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. Lancet 1994;343:619-29.

The authors reply:

To the Editor: Ioannidis et al., as well as Khan et al. and Stewart et al., suggest that the editors of the influential journals that published the trials we chose may tend to favor the publication of large trials whose results disagree with prior evidence. This is a new variation on publication bias that, unfortunately, cannot be proved. Ioannidis

et al. mention that our work confirms their own results[1] and those of Villar et al.,[2] but we want to respond to their comments.

First, we do not agree with the view that the six meta-analyses with more patients than the large randomized, controlled trials are more credible. Although the inclusion of more patients gives more statistical power, it cannot compensate for methodologic flaws. Second, we still think that the precision of an odds ratio is important, since it determines whether a therapy is adopted or rejected. An odds ratio whose confidence interval overlaps 1 will be considered, at best, to represent a tendency, and the null hypothesis will still stand. Third, we fully agree that the problems of heterogeneity are extremely important, and they are the object of our present work. Fourth, we are certainly in favor of having meta-analysis emphasize the systematic exploration of bias and diversity in research rather than the distillation of a magic odds ratio.

According to Bent et al., the higher base-line mortality rates in the meta-analysis[3] of the efficacy of nitrates in patients with myocardial infarction could explain the discrepancy between its results and those of the subsequent large randomized, controlled trial — the third study of the Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico (GISSI-3).[4] The question is whether these differences alone could move the odds ratio from 0.5 to nearly 1, given that the meta-analysis includes studies with base-line mortality rates that are lower than the one in the GISSI-3 trial. The large randomized, controlled trial would thus have met the homogeneity criteria of the meta-analysis. It is fortunate that the investigators decided to examine the role of nitrates in acute myocardial infarction in the era of thrombolytic agents and beta-blockers by conducting a trial rather than a sequential meta-analysis.

Imperiale proposes that the differences in base-line rates of preeclampsia and intrauterine growth retardation can be used to explain why the positive results of his meta-analysis[5] on the effects of aspirin were not confirmed by the large randomized, controlled trial[6] (the CLASP trial). An alternative explanation would be that among the six studies in the meta-analysis, one was a nonrandomized trial[7] and two were not placebo-controlled.[7,8]

We agree with the proposal of Sim and Lavori for the development of statistical techniques to explore specific sources of heterogeneity and assist in the selection of studies. The choice of the data to be included constitutes the first and most fundamental step in a review and is, in our opinion, much more important than its eventual shape or form.

JACQUES LELORIER, M.D., PH.D.
GENEVIÈVE GRÉGOIRE, M.D.
Hôtel-Dieu de Montréal
Montreal, QC H2W 1T8, Canada

1. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 1996;276:1332-8.
2. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. Lancet 1995;345:772-6.
3. Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. Lancet 1988;1:1088-92.
4. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Lancet 1994;343:1115-22.
5. Imperiale TF, Petrulis AS. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. JAMA 1991;266:260-4.
6. CLASP (Collaborative Low-Dose Aspirin Study in Pregnancy) Collaborative Group. CLASP: a randomised trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. Lancet 1994;343:619-29.
7. Schiff E, Peleg E, Goldenberg M, et al. The use of aspirin to prevent pregnancy-induced hypertension and lower the ratio of thromboxane A_2 to prostacyclin in relatively high risk pregnancies. N Engl J Med 1989;321:351-6.
8. Wallenburg HC, Rotmans N. Prevention of recurrent idiopathic fetal growth retardation by low-dose aspirin and dipyridamole. Am J Obstet Gynecol 1987;157:1230-5.

To the Editor: My objections to meta-analysis are purely pragmatic. It does not work nearly as well as we might want it to work. The problems are so deep and so numerous that the results are simply not reliable. My editorial cites a few relevant references, and I could have cited many more. The work of LeLorier et al. adds to the evidence that meta-analysis simply does not work very well in practice.

Khan et al. seem concerned that neither the meta-analyses nor the randomized, controlled trials were performed to their own standard of excellence. But that is just the point. As it is practiced and as it is reported in our leading journals, meta-analysis is often deeply flawed. Many people cite high-sounding guidelines, and I am sure that all truly want to do a superior analysis, but meta-analysis often fails in ways that seem to be invisible to the analyst. We cannot know whether improved implementation would alter the findings.

Stewart et al. suggest that leading journals may deliberately select and publish randomized, controlled trials that disagree with previously published meta-analyses, and they propose that all journals be searched for randomized, controlled trials. That could be useful, but it would pose a much bigger task than the work of LeLorier et al. and might miss the main point: the results of meta-analyses are often at variance with those of randomized, controlled trials. Certainly, randomized, controlled trials can be done as poorly as meta-analyses, and the analysis conducted by LeLorier et al. is also less than perfect. What we need is a guide through the imperfect world of science.

The advocates of meta-analysis and evidence-based medicine should undertake research that might demonstrate that meta-analyses in the real world — not just in theory — improve health outcomes in patients. Review of the long history of randomized, controlled trials, individually weak for this specific purpose, has led to overwhelming evidence of efficacy. Examples include the development of better vaccines, more effective screening for diseases, and improved treatments for childhood cancer, infections, mental illness, cardiovascular disease, and many others. I am not willing to abandon that history to join those now promoting meta-analysis as the answer, no matter how pretty the underlying theory, until its defects are honestly exposed and corrected. The knowledgeable, thoughtful, traditional review of the original literature remains the closest thing we have to a gold standard for summarizing disparate evidence in medicine.

JOHN C. BAILAR III, M.D., PH.D.
University of Chicago
Chicago, IL 60637-1470

Hormone-Replacement Therapy Compared with Simvastatin for Postmenopausal Women with Hypercholesterolemia

To the Editor: With respect to the article by Darling et al. (Aug. 28 issue)[1] on the effects of hormone-replacement therapy and simvastatin in postmenopausal women with hypercholesterolemia: We carried out a study in which we assessed the effects of the combination of these two therapies. A total of 71 women with hypercholesterolemia (mean [±SD] age, 53±4 years) were recruited. The initial phase of the study consisted of three months of hormone-replacement therapy and a cholesterol-lowering diet, which resulted in low-density lipoprotein (LDL) cholesterol levels of less than 160 mg per deciliter in 15 of the women (21 percent). Those with persistently high LDL cholesterol levels entered the second phase, in which simvastatin (10 mg per day) was added. A total of 34 women completed nine months of combined therapy (Table 1).

Our results are consistent with those of Darling et al. — namely, simvastatin was more effective than hormone-replacement therapy. However, we believe that the conclusion suggested by Darling et al. is somewhat inaccurate. Although some women, especially those with mild hypercholesterolemia, will benefit from hormone-replacement therapy as a single cholesterol-lowering treatment, most women with hypercholesterolemia will need a specific, more potent hypercholesterolemic drug. Most studies of the effects of hormone-replacement therapy on the lipid profile have shown only a 5 to 15 percent reduction in persistently high LDL levels,[2] in contrast with the 24 percent reduction reported by Darling et al. Combined treatment with hormone-replacement therapy and low-dose simvastatin proved to be extremely effective in our study, with no serious adverse reactions. Moreover, this combined therapeutic regimen may have a synergistic antiatherogenic effect.

AMOS PINES, M.D.
YORAM LEVO, M.D.
DANIEL AYALON, M.D.
Ichilov Hospital
Tel Aviv 64239, Israel

1. Darling GM, Johns JA, McCloud PI, Davis SR. Estrogen and progestin compared with simvastatin for hypercholesterolemia in postmenopausal women. N Engl J Med 1997;337:595-601.
2. Bush TL, Barrett-Connor E, Cowan LD, et al. Cardiovascular mortality and noncontraceptive use of estrogen in women: results from the Lipid Research Clinics Program Follow-up Study. Circulation 1987;57:1102-9.

TABLE 1. MEAN (±SD) CHOLESTEROL LEVELS IN WOMEN WITH HYPERCHOLESTEROLEMIA TREATED FOR THREE MONTHS WITH HORMONE-REPLACEMENT THERAPY AND DIET AND FOR NINE MONTHS WITH SIMVASTATIN PLUS HORMONE-REPLACEMENT THERAPY AND DIET.*

VARIABLE	BASE LINE	HRT AND DIET	SIMVASTATIN PLUS HRT AND DIET	
Time during regimen (mo)	—	3	1	9
Cholesterol (mg/dl)				
Total	290±29	281±26	213±30†	214±22†
LDL	204±31	187±26‡	122±30†	124±22†
HDL	53±12	62±16‡	60±18‡	63±14‡
Triglycerides (mg/dl)	144±63	170±65	147±68	147±67

*HRT denotes hormone-replacement therapy, LDL low-density lipoprotein, and HDL high-density lipoprotein.

†$P<0.001$ for the comparisons with base line and HRT plus diet.

‡$P<0.05$ for the comparison with base line.

To the Editor: The report by Darling and colleagues ascribes a 27 percent reduction in Lp(a) lipoprotein levels to postmenopausal hormone therapy with combined estrogen and progestin. Unfortunately, the authors do not consider the effects of dietary changes on lipoproteins. According to their summary of the protocol, "At enrollment, all the women were given . . . dietary advice by a trained nurse. Although they were encouraged to continue the recommended diet for the remainder of the study, there was no formal assessment of their compliance." The usual Australian diet is notoriously high in fat, saturated fat, and cholesterol, and lipid profiles may be expected to improve as a result of relatively modest changes.

Although not ideal, brief instruction on appropriate dietary changes can result in significant improvements in lipid profiles. Rhodes et al.[1] reported serum lipid changes in adults with a mean serum cholesterol level of 260 mg per deciliter who were instructed by a nurse or physician on the Step 1 diet of the National Cholesterol Education Program.[2] Dietary fat decreased from 37 percent of energy at base line to 31 percent, and cholesterol intake dropped by approximately 25 percent. Mean reductions of 7 percent in the total serum cholesterol level and 10 percent in the LDL cholesterol level occurred after 12 weeks.[1]

The dramatic reduction in LDL cholesterol levels with only 10 mg of simvastatin per day (36 percent, vs. an expected reduction of approximately 25 percent) may well be attributable to the combined effect of dietary modification and pharmacologic treatment. For any trial that purports to demonstrate the effect of a therapy on lipoprotein levels, it is an absolute necessity to document dietary composition at base line and during treatment.

LINDA CASHIN-HEMPHILL, M.D.
Boston Heart Foundation
Cambridge, MA 02142

LAURA I. VAILAS, PH.D.
University of Houston
Houston, TX 77204

1. Rhodes KS, Bookstein LC, Aaronson LS, Mercer NM, Orringer CE. Intensive nutrition counseling enhances outcomes of National Cholesterol Education Program dietary therapy. J Am Diet Assoc 1996;96:1003-10.
2. Second report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). Circulation 1994;89:1333-445.

The authors reply:

To the Editor: Dr. Pines and colleagues are concerned that our article may promote postmenopausal hormone therapy as the sole pharmacotherapy for postmenopausal women with hypercholesterolemia. In fact, we are trying to promote its individualized use as first-line pharmacotherapy in such women, and we fully concur with the notion that many women will require the addition of conventional lipid-lowering agents to hormone therapy in order to achieve the level of LDL cholesterol currently recommended by the expert panel of the National Cholesterol Education Program.[1] However, it is notable that studies specifically designed to investigate the effect of oral postmenopausal hormone therapy in women with hyperlipidemia consistently show a greater reduction in LDL cholesterol levels (12 to 24 percent)[2-4] than that documented in studies of women with normal lipid levels (10 to 15 percent).

We have also been interested in clarifying the effect of the concurrent use of postmenopausal hormone therapy and simvastatin on lipoprotein. Our unpublished data support the conclusions of Dr. Pines and colleagues and even suggest that the effect of the two therapies may be additive. Thus, the stepwise introduction of diet, followed by individualized hormone therapy, followed by a statin may prove to be the preferred way of managing hypercholesterolemia in postmenopausal women.

We appreciate the comments of Drs. Cashin-Hemphill and Vailas regarding the potential effect of relatively modest dietary modifications on the lipoprotein profile in patients with hypercholesterolemia. The crossover design of our study was used to compare two therapies (postmenopausal hormone therapy and simvastatin) with each other, not to determine the effect of each therapy alone. We were therefore careful not to include P values for the change from base-line values for each therapy alone but to include only P values for the comparison between the two therapies. The mean percentage change from base line with 95 percent confidence intervals was retained to assist in clinical interpretation of the results. Given the comments of Drs. Cashin-Hemphill and Vailas, it might have been more prudent to label the two therapies "postmenopausal hormone therapy plus dietary advice" and "simvastatin plus dietary advice." Even so, the comparison remains valid.

As we stated in our article, "The groups did not differ significantly in any variables at base line or at the end of the washout period (week 0 and week 16)." The dietary advice was given only at week 0. One would expect that any effect of dietary modification that was independent of the pharmacologic therapy would have been detectable at week 16, when all pharmacotherapy had been "washed out." No such effect was observed.

GISELLE M. DARLING, M.D.
SUSAN DAVIS, M.D., PH.D.
Jean Hailes Foundation
Melbourne 3168, Australia

JENNIFER A. JOHNS, M.D.
Austin and Repatriation Medical Centre
Melbourne 3084, Australia

1. Summary of the second report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel II). JAMA 1993;269:3015-23.
2. Tikkanen MJ, Nikkila EA, Vartiainen E. Natural oestrogen as an effective treatment for type-II hyperlipoproteinaemia in postmenopausal women. Lancet 1978;2:490-1.
3. Tonstad S, Ose L, Gørbitz C, Djøseland O, Bard JM, Fruchart JC. Efficacy of sequential hormone replacement therapy in the treatment of hypercholesterolaemia among postmenopausal women. J Intern Med 1995; 238:39-47.
4. Perrone G, Stefanutti C, Galoppi P, et al. Effect of oral and transdermal hormone replacement therapy on lipid profile and Lp(a) level in menopausal women with hypercholesterolemia. Int J Fertil Menopausal Stud 1996; 41:509-15.

Obesity

To the Editor: In their article on obesity, Rosenbaum et al. (Aug. 7 issue)[1] characterize sibutramine, currently undergoing regulatory review as a drug for the treatment of obesity, as having both catecholaminergic and serotonergic agonist effects. Sibutramine, in fact, is not an agonist at catecholamine or serotonin receptors but instead acts by inhibiting the reuptake of serotonin and norepinephrine at central synapses. Thus, sibutramine's mode of action is similar to that of other monoamine-reuptake inhibitors such as venlafaxine (an inhibitor of serotonin and norepinephrine reuptake) and fluoxetine (a selective serotonin-reuptake inhibitor). The mode of action of drugs that alter appetite by enhancing central monoamine activity has several implications. For instance, primary pulmonary hypertension has been associated with certain appetite-suppressant drugs[2] (e.g., fenfluramine), as well as other drugs (e.g., cocaine), that act by causing monoamine release.[3] Rosenbaum et al. characterize the implicated anorectic drugs as reuptake inhibitors. In fact, monoamine-reuptake inhibitors that do not cause monoamine release, such as tricyclic antidepressant drugs, selective serotonin-reuptake inhibitors, and serotonin- and norepinephrine-reuptake inhibitors, have not been associated with an increased risk of primary pulmonary hypertension or neurotoxicity[4] or with the cardiac valvulopathy reported by Connolly et al.[5] Given the diversity of the satiety-enhancing drugs currently being developed, attention to the details of the mechanism of action may be even more important in the future.

TIMOTHY B. SEATON, M.D.
CARL M. MENDEL, M.D.
STEVEN P. WEINSTEIN, M.D., PH.D.
Knoll Pharmaceutical
Parsippany, NJ 07054

1. Rosenbaum M, Leibel RL, Hirsch J. Obesity. N Engl J Med 1997;337: 396-407.
2. Abenheim L, Moride Y, Brenot F, et al. Appetite-suppressant drugs and the risk of primary pulmonary hypertension. N Engl J Med 1996;335:609-16.
3. Lane R, Baldwin D. Selective serotonin reuptake inhibitor–induced serotonin syndrome: review. J Clin Psychopharmacol 1997;17:208-21.
4. McCann UD, Seiden LS, Rubin LJ, Ricaurte GA. Brain serotonin neurotoxicity and primary pulmonary hypertension from fenfluramine and dexfenfluramine: a systematic review of the evidence. JAMA 1997;278:666-72.
5. Connolly HM, Crary JL, McGoon MD, et al. Valvular heart disease associated with fenfluramine–phentermine. N Engl J Med 1997;337:581-8.

To the Editor: The excellent review of obesity contains a contradiction. The authors say that insulin causes weight gain because it increases the expression of neuropeptide Y

messenger RNA, with subsequent central appetite stimulation. They later say that insulin reduces food intake by inhibiting the expression of neuropeptide Y, and in another section, insulin is said to increase the expression of leptin in adipose tissue. Can the authors clarify the role of insulin in the stimulation or inhibition of food intake?

ELLIOTT EISENBUD, M.D.
6403 Coyle Ave.
Carmichael, CA 95608

The authors reply:

To the Editor: We appreciate the letter from Seaton et al. concerning the mode of action of sibutramine. We wished to note only that sibutramine has agonist actions in both serotonin and catecholamine systems. The full delineation of the exact mode of action of centrally active anorexiant drugs will undoubtedly receive increased scrutiny because some of these drugs may be neurotoxic,[1] cause primary pulmonary hypertension, and cause valvular heart disease.[2] Since our review was published, the Food and Drug Administration (FDA) and Wyeth–Ayerst Laboratories, acting on the concern about these possible adverse effects, have removed fenfluramines from the market.

We thank Dr. Eisenbud for his careful reading of the text. On page 401 (left-hand column), the statement, "The expression of neuropeptide Y mRNA is increased by insulin and glucocorticoids," is incorrect. It should read, "The expression of neuropeptide Y mRNA is increased by androgens and glucocorticoids and decreased by leptin, insulin, and estrogen."

There are two other errors. First, on page 397 (right-hand column), the statement suggesting that lesions of the median forebrain bundle are equivalent to those of the ventromedial hypothalamus is incorrect. Lesions of the median forebrain bundle are more likely to induce anorexia in a manner similar to that of lesions of the lateral hypothalamus. Second, on page 403, in the third sentence under the heading "Drug Therapy," *fenfluramine* should be *dexfenfluramine*. Dexfenfluramine was approved by the FDA in 1996, but fenfluramine was approved in 1973.

JULES HIRSCH, M.D.
RUDOLPH L. LEIBEL, M.D.
MICHAEL ROSENBAUM, M.D.
Rockefeller University
New York, NY 10021

1. McCann UD, Seiden LS, Rubin LJ, Ricaurte GA. Brain serotonin neurotoxicity and primary pulmonary hypertension from fenfluramine and dexfenfluramine: a systematic review of the evidence. JAMA 1997;278:666-72.
2. Connolly HM, Crary JL, McGoon MD, et al. Valvular heart disease associated with fenfluramine–phentermine. N Engl J Med 1997;335:581-8.

Transient Renal Failure Due to Simultaneous Ibuprofen and Aminoglycoside Therapy in Children with Cystic Fibrosis

To the Editor: Ibuprofen can retard the decline in pulmonary function in children with cystic fibrosis.[1] Aminoglycosides are often given to treat pulmonary infections in children with cystic fibrosis. Both ibuprofen and aminoglycosides are nephrotoxic. We have seen four children with cystic fibrosis who had transient renal failure during exacerbations of their lung disease that we believe was caused by the intravenous administration of an aminoglycoside while maintenance treatment with ibuprofen was continued.

The first patient was a 16-year-old girl with severe lung disease in whom nausea and vomiting developed six days after admission and the commencement of therapy with intravenous tobramycin for exacerbation of her lung disease. She received furosemide for peripheral edema, and oliguria developed the next day. Her serum creatinine concentration rose from 0.5 mg per deciliter (48 μmol per liter) at the time of admission to 2.8 mg per deciliter (249 μmol per liter) six days later. The ibuprofen and tobramycin were discontinued, and the serum creatinine concentration was 0.8 mg per deciliter (74 μmol per liter) two days later. The maximal serum tobramycin concentration was 27 mg per milliliter. The patient died nine days later of lung disease; minimal tubulointerstitial nephritis was seen at autopsy.

The second patient was a 10-year-old girl with moderately severe lung disease in whom nausea, vomiting, and abdominal cramps developed two days after admission and the initiation of therapy with intravenous gentamicin for exacerbation of her chronic lung disease. She was found to have a supratherapeutic peak serum gentamicin concentration of 16 mg per milliliter, and her serum creatinine concentration had increased from 0.7 mg per deciliter (64 μmol per liter) at the time of admission to 2.4 mg per deciliter (211 μmol per liter). The ibuprofen and gentamicin were discontinued, and her serum creatinine concentration was 0.9 mg per deciliter (82 μmol per liter) five days later.

The other two patients were twin 23-month-old brothers who were hospitalized simultaneously for exacerbations of chronic lung disease. Both patients were treated with intravenous gentamicin. The ibuprofen they were taking before admission was inadvertently continued, despite the existence of a policy of stopping ibuprofen during hospitalization if aminoglycoside therapy was given. This policy was instituted as a result of the first two cases. One twin began vomiting four days after admission. Ibuprofen was discontinued nine days later, but lethargy, increased vomiting, and periorbital edema occurred the next day, followed by generalized edema and oliguria. The child's serum creatinine concentration rose from 0.2 mg per deciliter (20 μmol per liter; measured 3 months earlier) to 5.2 mg per deciliter (460 μmol per liter) 16 days after admission. He received peritoneal dialysis for eight days, by which time urine output was normal and his serum creatinine concentration was 0.4 mg per deciliter (32 μmol per liter). The other brother had a transient asymptomatic increase in the serum creatinine concentration, from 0.6 mg per deciliter (49 μmol per liter) at admission to 1.5 mg per deciliter (134 μmol per liter) 18 days later. In both cases, subsequent audiologic testing had normal results, although the more severely affected brother had transient ataxia.

Our observations suggest that the combination of intravenous aminoglycoside and ibuprofen can cause acute renal insufficiency. We have not seen this complication with ibuprofen alone or with ibuprofen and nebulized tobramycin.[2] Aminoglycoside-induced toxicity is potentiated by

extracellular volume depletion,[3] and ibuprofen and other nonsteroidal antiinflammatory drugs interfere with the intrarenal production of prostaglandin E_2 and prostacyclin, which cause renal vasodilatation in the presence of reduced circulating volume.[4] Our findings suggest that ibuprofen should be stopped during intravenous aminoglycoside therapy.

THOMAS A. KOVESI, M.D.
RONALD SWARTZ, B.SC.PHM.
NONI MACDONALD, M.D.
Children's Hospital of Eastern Ontario
Ottawa, ON K1H 8L1, Canada

1. Konstan MW, Byard PJ, Hoppel CL, Davis PB. Effect of high-dose ibuprofen in patients with cystic fibrosis. N Engl J Med 1995;332:848-54.
2. Ramsey RW, Dorkin HL, Eisenberg JD, et al. Efficacy of aerosolized tobramycin in patients with cystic fibrosis. N Engl J Med 1993;328:1740-6.
3. Bennett WM. Aminoglycoside nephrotoxicity. Nephron 1983;35:73-7.
4. Marasco WA, Gikas PW, Azziz-Baumgartner R, Hyzy R, Eldredge CJ, Stross J. Ibuprofen-associated renal dysfunction. Arch Intern Med 1987; 147:2107-16.

Case 21-1997: Paraneoplastic Cerebellar Degeneration and Hodgkin's Disease

To the Editor: Dr. Eder, in his discussion of the paraneoplastic cerebellar degeneration associated with Hodgkin's disease (July 10 issue),[1] quoted the 1976 paper by Trotter and colleagues[2] but did not mention that the antibodies they described are now recognized as distinct antibodies associated with Hodgkin's disease. These antibodies have been designated anti-Tr antibodies.[3] They have a characteristic dotted staining pattern, suggestive of immunoreactivity of the dendritic spines of Purkinje cells, and are specific for Hodgkin's disease.

TAHSEEN MOZAFFAR, M.B., B.S.
Aga Khan University College of Medicine
Karachi 74800, Pakistan

1. Case Records of the Massachusetts General Hospital (Case 21-1997). N Engl J Med 1997;337:115-22.
2. Trotter JL, Hendin BA, Osterland CK. Cerebellar degeneration with Hodgkin disease: an immunological study. Arch Neurol 1976;33:660-1.
3. Graus F, Dalmau J, Valldeoriola F, et al. Immunological characterization of a neuronal antibody (anti-Tr) associated with paraneoplastic cerebellar degeneration and Hodgkin's disease. J Neuroimmunol 1997;74: 55-61.

Professionalism

To the Editor: Karen Ignagni, the president of the American Association of Health Plans, stated in the October 9 issue, "The editor of the *New England Journal of Medicine* is entitled to be a critic of managed care, but it is profoundly disturbing to see such an important and presumably dispassionate publication used as a sounding board for these critical views."[1] That managed care promotes the slogan "Putting patients first" says it all. That is the unspoken assumption of the medical profession. It has been the foundation of the physician–patient relationship since Hippocrates. This special relationship is based on honor, and honor need not be spoken. Doctors do not require a slogan for trust. When trust is gone, it cannot be restored, like a soul that has left the body. Trust that is honored cannot be captured by a managed-care contract. Grasp a butterfly with hot tongs and the butterfly dies.

What physician, including the editor, can be dispassionate about the current destruction of our medical family at the hands of profit-hungry CEOs? Putting profit first is their unspoken assumption. As a practicing neurosurgeon for 28 years, I recommend passion when compassion for the patient is first, foremost, and central. As a young resident, I remember Wilder Penfield's words: "Keep the businessman out of medicine." I challenge you, as editor, to continue to speak in strong terms about the heart of the matter. In medicine we are witnesses to, and to some extent accomplices in, the social revolution aimed at converting people into integers. As Osler said, we can have both, science and faith, if only we keep them separate. There is plenty of room for dispassionate science in the *Journal*, but honor, trust, and dignity are matters of faith, not science. Once we were knights, duty bound to protect each and every patient, regardless of monetary concerns.

We physicians have allowed the current gross decline in our once noble profession. We have been passive passengers, docile slaves obedient to the gag clause. We cannot rely on government or the profit-oriented insurance industry to correct the medical train wreck in progress. We must take over the engine. We must lift ourselves up, with passion.

CHARLES W. NEEDHAM, M.D.
5 Elmcrest Terr.
Norwalk, CT 06850

1. Ignagni K. Putting patients first? N Engl J Med 1997;337:1084.

To the Editor: You point out how the fee-for-service system participates in the support of research and education.[1] You omit, however, an important aspect of that support: the countless hours that voluntary clinical faculty contribute as ward attendings, clinic attendings, and preceptors in their own offices. These physicians are culled from any area where there is a medical school (and often where there is not). They give willingly and generously of their time and expertise, believing that being a physician includes passing along their hard-won knowledge and skills to the next generation. Many of my colleagues and I have been preceptors and attending physicians year after year, for countless young medical students and physicians. We do this silently and without fanfare. Most of the general public does not even know about it, as witnessed by the reactions of my new patients (the old ones are used to it), who are surprised and usually delighted to see students in the office.

PEPI GRANAT, M.D.
7800 Red Rd.
South Miami, FL 33143

1. Kassirer JP. Putting patients first? N Engl J Med 1997;337:1086.