

## Meta-analysis

### A Quantitative Approach to Research Integration

Stephen B. Thacker, MD, MSc

Meta-analysis is being used with increasing frequency in clinical medicine as an attempt to improve on traditional methods of narrative review by systematically aggregating information and quantifying its impact. Combining data from several studies using meta-analysis can increase statistical power, provide insight into the nature of relationships among variables, and increase generalizability of results more rigorously than less quantitative review methods. Like all review methods, meta-analysis can be limited by sampling bias, inadequate data, and biased outcome interpretation. Still, the advantages noted above make meta-analysis a methodology that warrants testing and empirical evaluation.

(*JAMA* 1988;259:1685-1689)

THE GOAL of an integrative literature review is to summarize the accumulated knowledge concerning a field of interest and to highlight important issues that researchers have left unresolved.<sup>1</sup> Traditionally, the medical literature has been integrated in the narrative form. An expert in a field will review studies, decide which are relevant, and highlight his or her findings, both in terms of results and, to a lesser degree, methodology. Topics for further research may also be proposed. Such narrative reviews have two basic weaknesses.<sup>2,3</sup> First, no systematic approach is prescribed to obtain primary data or to integrate findings; rather, the subjective judgment of the reviewer is used. As a result, no explicit standards exist to assess the quality of a review. Second, the narrative reviewer does not synthesize data quantitatively across literature. Consequently, as the number of studies in any discipline increases, so does the probability that erroneous conclusions will be reached in a narrative review.<sup>4</sup>

From the Center for Environmental Health and Injury Control, Centers for Disease Control, Atlanta.

Reprint requests to Center for Environmental Health and Injury Control, Centers for Disease Control, Atlanta, GA 30333 (Dr Thacker).

Scientific research is founded on integration and replication of results; with the possible exception of a new discovery, a single study rarely makes a dramatic contribution to the advancement of knowledge.<sup>5</sup> In this article I summarize the constraints on reviewers of the medical literature and review alternative methods for synthesizing scientific studies. In particular, I examine meta-analysis, a quantitative

---

See also p 1678.

---

method to combine data, and illustrate with a clinical example its application to the medical literature. Then, I describe the strengths and weakness of meta-analysis and approaches to its evaluation. Finally, I discuss current research issues related to meta-analysis and highlight future research directions.

#### CONSTRAINTS ON LITERATURE REVIEW

The limitations of any approach to literature review can be summarized as follows<sup>6</sup>: (1) sampling bias due to reporting and publication policies; (2) the absence in published studies of specific data desired for review; (3) biased exclusion of studies by the investigator; (4)

the uneven quality of the primary data; and (5) biased outcome interpretation. These concerns are applicable to any form of literature review.

Two types of bias in the published literature must concern a reviewer. First, because authors and journal editors tend to report statistically significant findings, a review limited to published studies will tend to overestimate the effect size. In a survey, for example, 58 investigators indicated that they had conducted 921 randomized controlled trials, and that 96 (21.3%) were unpublished. Positive randomized controlled trials were significantly more likely to be published than negative trials (77% vs 42%,  $P < .001$ ).<sup>7</sup> At the same time, one should not uncritically assume that methods are better in published studies, as the quality of published papers varies dramatically.<sup>8</sup> Second, another form of publication bias, the confirmatory bias, tends to emphasize and believe experiences that support one's views and to ignore or discredit those that do not. Results of a study of 75 journal reviewers asked to referee identical experimental procedures showed poor interrater agreement and a bias against results contrary to their theoretical perspective.<sup>9</sup> Consequently, new or unpopular data tend also to be underreported in the published literature.

Data available from primary research studies may be inadequate for the literature reviewer. The reviewer is often confronted with selective reporting of primary findings, incorrect primary data analysis, and inadequate descriptions of original studies.<sup>10</sup> In a study of psychotherapy outcomes, for example, an effect could not be calculated in 26% of studies because of missing data, a

number comparable with previous reports.<sup>11</sup>

In addition to identifying studies, the investigator must decide which reports to include in a review.<sup>3</sup> One option is to use all available data and thereby maximize the representativeness of the conclusions. Using this approach, however, one will decrease the statistical validity of the data synthesis by including less rigorous studies. Exclusion of studies for methodological reasons, on the other hand, will increase the statistical validity but will decrease the size of the overall pool of data and may sacrifice the ability to generalize from the results.

Variable data quality is probably the most critical limitation for the reviewer. The effect of data quality was seen in a study of quality of life outcomes following coronary bypass graft surgery, when investigators found the estimates of benefit to be 15% less in randomized controlled trials than in trials using matching.<sup>12</sup> Similarly, results of studies in medical care tend to show decreasing odds ratios with increased rigor of studies,<sup>8</sup> although in one large study of psychotherapy, the effect was found to increase with increasing rigor.<sup>11</sup> In quantitative reviews, statistical methods, including stratified analyses and multivariate methods, can be used to measure the impact on the results of varying quality in studies.<sup>8,13,14</sup>

Although these constraints have been recognized previously, the more recent efforts to address concerns about research integration have stimulated new efforts to deal with them.

## QUANTITATIVE APPROACHES TO SUMMARIZING ACROSS STUDIES

During the past several years, there have been several different approaches developed to summarize quantitatively data found in different studies of the same or similar research problems. The simplest approach to the quantitative integration of research is vote counting. With this approach, results of studies under consideration are classified into three categories: (1) statistically significant in one direction, (2) statistically significant in the opposite direction, or (3) no statistically significant difference. Then, the category receiving the most votes is judged to approximate truth.<sup>15</sup> Although simple to use, voting methods do not take into account the magnitude of effect or sample size. In addition, this approach does not address the aforementioned problems inherent in traditional reviews, such as inadequate study methodology and uneven data quality.

In 1971, Light and Smith<sup>15</sup> proposed an alternative to voting methods that

takes advantage of natural aggregations, or clusters, in the population. In this approach, one studies a problem in various clusters, such as neighborhoods or classrooms, and searches for explanations for differences among clusters. If these differences are explainable, the data can be combined and statistical variability can be described.

A third method for combining literature is pooling, a method by which data from multiple studies of a single topic, such as  $\beta$ -blockade after myocardial infarction, are combined in a single analysis.<sup>16</sup> This method is limited by the availability of raw data; the variation in study methods, populations, and outcomes under study; and statistical considerations.<sup>17,18</sup>

In a 1976 study of the efficacy of psychotherapy, Glass<sup>19</sup> coined the term *meta-analysis*, "the statistical analysis of a large collection of results from individual literature, for the purpose of integrating the findings." Alternatively, *meta-analysis* can be defined as any systematic method that uses statistical analyses for combining data from independent studies to obtain a numerical estimate of the overall effect of a particular procedure or variable on a defined outcome.<sup>20</sup>

While there have been several approaches to meta-analysis, the steps can be defined generally as (1) defining the problem and criteria for admission of studies, (2) locating research studies, (3) classifying and coding study characteristics, (4) quantitatively measuring study characteristics on a common scale, (5) aggregating study findings and relating findings to study characteristics (analysis and interpretation), and (6) reporting the results.<sup>21,22</sup>

Problem formulation includes the explicit definition of both outcomes and potentially confounding variables. Carefully done, this step enables the investigator to focus on the relevant measures in the studies under consideration and to specify relevant methods to classify and code study characteristics.

The literature search includes a systematic approach to locating studies.<sup>1</sup> First, one obtains information from the so-called invisible college, ie, the informal exchange of information among colleagues in a particular discipline. Second, one searches indexes (eg, *Index Medicus* and the *Social Science Citation Index*), abstracting services (eg, *International Pharmaceutical Abstracts*), and computerized searches (eg, MEDLINE and TOXLINE) to obtain research articles and sources of both published and unpublished data. Third, references in available studies identify further sources. The retrieval

from academic, private, and government researchers of unreferenced reports, the so-called fugitive literature, as well as unpublished data, further minimizes selective reporting and publication biases.

Several methods are used to measure the results across studies.<sup>3,23</sup> The most commonly used measure in the social sciences is the effect size, an index of both the direction and magnitude of the effect of a procedure under study.<sup>19</sup> Glass and his colleagues<sup>24</sup> developed this method when assessing the efficacy of psychotherapy on the basis of data from controlled studies. One estimate of effect size for quantitative data is the difference between two group means divided by the control group SD:  $(X_e - X_c)/S_c$ , where  $X_e$  is the mean of the experimental or exposed group,  $X_c$  is the mean of the control or unexposed group, and  $S_c$  is the SD of the control group. Effect size expresses differences in SD units so that, for example, if a study has an effect size of 0.2 SD units, the overall effect size is half that of another study that has an effect size of 0.4 SD units. The appropriate measure of effect across literature will vary according both to the nature of the problem being assessed and to the availability of published data.<sup>7,25</sup> Pooling of data from controlled clinical trials, for example, has been more widely used in the medical literature.<sup>16,26</sup>

Effect size for proportions has been calculated in cohort literature as either a difference,  $P_e - P_c$ , or as a ratio,  $P_e/P_c$ .<sup>3</sup> The latter has the advantage of considering the change relative to the control percentage and, in epidemiologic studies, is equivalent analytically to the concept of the risk ratio.

Whatever combination statistic is used, a systematic quantitative procedure to accumulate results across studies should include the following<sup>27</sup>: (1) summary descriptive statistics across studies and the averaging of those statistics; (2) calculation of the variance of a statistic across studies (ie, tests for heterogeneity); (3) correction of the variance by subtracting sampling error; (4) correction in the mean and variance for study artifacts other than sampling, such as measurement error; and (5) comparison of the corrected SD to the mean to assess the size of the potential variation across studies. A growing literature on statistical methods deals with problems in calculating effect size or significance testing as it relates to meta-analysis.<sup>28,29</sup>

## BENEFITS OF META-ANALYSIS

Meta-analysis forces systematic thought about methods, outcomes, cat-

egorizations, populations, and interventions as one accumulates evidence. In addition, it offers a mechanism for estimating the magnitude of effect in terms of a statistically significant effect size or pooled odds ratio. Furthermore, the combination of data from several studies increases generalizability and potentially increases statistical power, thus enabling one to assess more completely the impact of a procedure or variable.<sup>30</sup> Quantitative measures across studies can also give insight into the nature of relationships among variables and provide a mechanism for detecting and exploring apparent contradictions in results. Finally, users of meta-analysis have expressed the hope that this systematic approach would be less subjective and would decrease investigator bias.

## APPLICATIONS OF META-ANALYSIS IN HEALTH

Interest in clinical applications of meta-analysis has risen dramatically in recent years.<sup>31,32</sup> An increasing number of attempts have been made to use meta-analysis outside of mental health or educational settings, including such other settings as chemotherapy in breast cancer,<sup>33</sup> patient education interventions in clinical medicine,<sup>34</sup> spinal manipulation,<sup>35</sup> the effects of exercise on serum lipid levels,<sup>36</sup> and duodenal ulcer therapy.<sup>37</sup> There has also been discussion of the potential applications of meta-analysis to public health.<sup>38</sup> An interesting application of meta-analysis was an effort to quantify the impact on survival and safety of a wide range of surgical and anesthetic innovations.<sup>39</sup> More typical are efforts to draw conclusions from data pooled from a limited number of studies, usually controlled clinical trials.<sup>26,40-47</sup> Pooling techniques have also been applied to data from nonrandomized studies in attempts to address incompletely studied problems and to increase representativeness.<sup>25,48,49</sup>

## A CASE STUDY: ELECTRONIC FETAL MONITORING

In a 1979 review of the efficacy and safety of intrapartum electronic fetal monitoring, Banta and Thacker<sup>50</sup> set out to assess the evidence for the efficacy and safety of the routine use of electronic fetal monitoring. The independent variable was defined as the clinical application of all forms of electronic fetal monitoring to both high- and low-risk pregnant women; the outcomes measured were various measures of maternal and fetal morbidity and mortality, as well as the occurrence of cesarean delivery. Cost issues were also addressed.

A literature search began with the exchange of information with colleagues in obstetrics, pediatrics, epidemiology, technology assessment, and economics. References to published research articles were obtained from MEDLINE and *Index Medicus* and supplemented with references in articles under review. Efforts were also made to obtain unpublished reports and professional meeting abstracts. Although this review was systematic and extensive and comparable evidence from studies was sought, a quantitative analysis across studies was limited to descriptive statistics.

A 1987 meta-analysis of this same issue focused on evidence from randomized controlled trials and the previous literature search supplemented with information from the Oxford Database of Perinatal Trials and from direct correspondence with individual investigators.<sup>51</sup> Variables were codified and, where possible, made comparable. For example, published measures of the Apgar score varied in timing (at 1, 2, and 5 minutes) and classification (abnormal was defined variably to include or exclude a score of 7); authors were asked to provide one-minute Apgar scores where a normal score included 7.

The primary data were then organized into descriptive tables that listed study results for specific outcomes, such as low Apgar score, perinatal mortality, and cesarean delivery, as well as for measures of diagnostic precision, such as sensitivity, specificity, and predictive value (Table 1).<sup>50</sup> The findings of the randomized controlled trials were evaluated for comparability and then pooled (Table 2), and the pooled analyses were stratified by data quality.<sup>51</sup> The results of the pooled analyses were then reported, conclusions were drawn, and recommendations were made.

The 1979 study concluded that the data did not support the routine use of electronic fetal monitoring and recommended additional randomized controlled trials and limitation of electronic fetal monitoring to high-risk pregnancies.<sup>50</sup> The 1987 report included randomized controlled trials already cited in the original study and three additional randomized controlled trials (seven randomized controlled trials from five countries). No known clinical trials were excluded from this report although the largest trial,<sup>52</sup> which included more subjects than the other six in combination, was analyzed separately and compared with the pooled results of the others.

Analyses of different subsets of these studies based on differences in design (eg, use of fetal scalp blood in sampling)

and study quality found minor variations in results, but no changes in the basic findings. Data from the randomized controlled trial that scored highest in an assessment of the quality of study design and implementation, however, indicated that electronic fetal monitoring combined with fetal scalp blood sampling could be used to identify infants at risk of neonatal seizures.<sup>52</sup> That study had been suggested by pooled analyses of earlier randomized controlled trials.<sup>53</sup> The pooled cesarean delivery rate was twofold higher in the group with electronic fetal monitoring. While both of these reports illustrate the advantages of the systematic and comprehensive approach to a literature review, the meta-analytic methods used in the 1987 report illustrate both increased statistical power derived from data pooling and increased information found from stratification of studies. Subsequently available trials reported results consistent with that meta-analysis.<sup>54,55</sup>

## CRITICISMS OF META-ANALYSIS

When meta-analysis was introduced in the psychology literature, it did not meet with universal acceptance. It was variously described as "an exercise in mega-silliness" and "an abuse of research integration."<sup>56,57</sup> In addition to the constraints listed above related to literature review, the meta-analyst is confronted with additional challenges in an effort to synthesize data quantitatively across studies.

Statistical significance testing that is familiar to most clinicians is based on an assumption that data are selected randomly from a well-specified population. Nonrandom selection of studies and multiple tests of the same data, either through repeated publication of partial or entire data sets or through use of more than one outcome for each person, are two ways that this assumption is violated. Nevertheless, standard parametric statistics have been considered to be sufficiently robust to be usable in meta-analyses.<sup>58</sup>

The current use of parametric statistical methods for meta-analysis requires additional theoretical study.<sup>29</sup> Other methodological issues of concern to meta-analysts include bias,<sup>59</sup> variability between studies,<sup>60</sup> and the development of models to measure variability across studies.<sup>61</sup> Additional statistical research should include study of the impact of outliers on the meta-analysis and the potential insight that they could provide into a research question.<sup>28</sup> Statistically valid methods to combine data across studies of varying quality and design, including data from case-control

Table 1.—Accuracy of Electronic Fetal Monitoring Using Apgar Score as Measure of Outcome\*

Investigator, y	No. of Patients	PPV	NPV	Sensitivity, %	Specificity, %
Bissonnette, <sup>69</sup> 1975	714	80	94	57	82
Gabert and Stenchever, <sup>70</sup> 1973	749	66	91	84	80
Schiffrin and Dame, <sup>71</sup> 1972	307	43	93	54	90
Saldana et al., <sup>72</sup> 1976	620	23	86	71	44
Tipton and Shelley, <sup>73</sup> 1971	100	81	93	82	93

\*Abnormality defined as one-minute Apgar score less than 7 (except Gabert—one-minute Apgar score <6). PPV indicates positive predictive value; NPV, negative predictive value.

Table 2.—Pooled Data From Six Controlled Trials Assessing Efficacy of Routine Electronic Fetal Monitoring in Labor

	Pooled Odds Ratio	95% Confidence Interval	Test for Heterogeneity*
Apgar score <7	1.07	0.88-1.30	$\chi^2_6 = 1.28$
<4	0.87	0.56-1.30	$\chi^2_5 = 1.84$
Neonatal seizures	0.39	0.14-1.08	$\chi^2_6 = 3.75$
NICU admissions†	1.01	0.84-1.22	$\chi^2_6 = 12.15$
Perinatal deaths	1.73	0.53-5.64	$\chi^2_6 = 5.51$

\*All tests were not significant. The subscript numbers refer to *df*.

†NICU indicates neonatal intensive care unit.

studies, will enable meta-analysts to maximize the value of their data syntheses.<sup>48</sup>

One serious concern about quantitative reviews of the literature is that although meta-analysis is more explicit, it may be no more objective than a narrative review.<sup>62</sup> Both critics and advocates of meta-analysis are concerned that an unwarranted sense of scientific validity, rather than true understanding, may result from quantification.<sup>63,64</sup> In other words, sophisticated statistics will not improve poor data but could lead to an unwarranted comfort with one's conclusions.<sup>65</sup>

### EVALUATION OF META-ANALYSIS

The evaluation of a literature review, like its conduct, should be systematic and quantitative. Evaluation criteria for meta-analysis include the need for the following: (1) clear identification of the problems under study; (2) active effort to include all available studies; (3) assessment of publication bias; (4) identification of data used; (5) selection and coding based on theoretical framework, not convenience; (6) detailed documentation of coding; (7) use of multiple raters to assess coding, including assessment of interrater reliability; (8) assessment of comparability of the cases, controls, and circumstances in the studies analyzed; (9) consideration of alternative explanations in the discussion; (10) relation of study characteristics to problems under review; (11) careful limitation of generalization to the domain of the literature review; (12) reporting in enough detail to enable replication by a reviewer; and (13) guidelines for future research.<sup>8,66</sup>

### COMMENT

Meta-analysis is an attempt to improve traditional methods of narrative review by systematically aggregating information and quantifying its impact. Meta-analysis was introduced to address the problem of synthesizing the large quantity of information on a particular subject, a problem that has been exacerbated by the large volume of published research in the past 20 years. It is viewed, however, only as a step in the process of developing better tools to quantify information across studies. It should neither be considered the final word in quantitative reviewing, nor should it be dropped in haste because of the problems and criticisms discussed above. Certainly, benefits are to be obtained from systematic and rigorous review of available information, including increases in power and generalizability, better understanding of complex issues, identification of correlations among variables, and identification of gaps to be addressed by appropriate research.

When criticizing meta-analysis, one must distinguish between those problems that are inherent in any literature review and those that are specifically a problem with meta-analysis. For example, data quality, sampling bias, and data retrieval are limitations inherent in any literature review. Similarly, while outcome interpretation may be affected by the various styles of summarizing research findings, biases are not limited to the meta-analyst. On the other hand, one must be wary of inappropriate weight being given to a procedure just because it is quantitative, particularly when used by those who do not under-

stand the limitations of the statistical methods utilized. Finally, critics should empirically test the impact of their criticisms so as to take meta-analysis or its alternative methods of quantitative summarization of research to the next level of usefulness.

It has been suggested that investigators should combine quantitative and qualitative review data to enable practitioners to apply results to individual patients or program problems.<sup>67</sup> In this way, researchers can investigate issues that are important but difficult to quantify. Nonquantitative information, such as expert opinion and anecdotal evidence, does have a significant impact on policy. Finally, one must be concerned that although even the best meta-analysis may represent all available trials and relevant studies, it may not represent clinical practice because of the nature of how and where research is conducted.<sup>68</sup>

Several things can be done to assess meta-analysis and to improve methods of quantitative review. First, one can compare the results of meta-analysis with those of narrative reviews to identify differences in interpretation and conclusions. In one study where a statistical procedure for summarizing research findings was compared with narrative reviews, it was found that the statistical reviewer was more likely to support the hypothesis both in direction and magnitude, although the basic recommendations did not differ between groups.<sup>68</sup> A second important area of research is in statistical methodology. Both theoretical research into the assumptions of alternative methods and empirical research testing of the accuracy and efficiency of these methods need to be undertaken. Third, methods to assess the quality of meta-analysis need to be tested and refined.<sup>66</sup> Finally, in assessing meta-analysis, one must be careful to limit the extrapolation of conclusions to the field of study covered by the literature review. Although this is true of any cumulative review, the boundaries of the review must be carefully delineated and interpretation confined to those boundaries.

In summary, the systematic, quantitative review and organization of the cumulative experience in a subject matter is fundamental to good scientific practice. Meta-analysis is a methodology that warrants testing and empirical evaluation. This is similarly true of alternative approaches to synthesizing information. The need to use available information optimally cannot be avoided by the rational scientist. The particular framework of review—be it meta-analysis or some other approach—should be addressed as an important scientific en-

deavor. The importance of addressing this issue must be underscored in an era where scientific information is increasing exponentially and the potential for application of these findings is unprecedented.

## References

1. Cooper HM: Scientific guidelines for conducting integrative research reviews. *Rev Educ Res* 1982;52:291-302.
2. Jackson GB: Methods for integrative reviews. *Rev Educ Res* 1980;50:438-460.
3. Light RJ, Pillemer DB: *Summing Up the Science of Reviewing Research*. Cambridge, Mass, Harvard University Press, 1984.
4. Levine RL, Hunter JE: Regression methodology: Correlation, meta-analysis, confidence intervals, and reliability. *J Leisure Res* 1983;15:323-343.
5. Fiske DW: The meta-analytic revolution in outcome research. *J Consult Clin Psychol* 1983;51:65-70.
6. Hattie JA, Hansford BC: Meta-analysis: A reflection on problems. *Aust J Psychol* 1984;36:239-254.
7. Chan SS, Sacks HS, Chalmers TC: The epidemiology of unpublished randomized control trials, abstracted. *Clin Res* 1982;30:234A.
8. Chalmers TC, Smith H Jr, Blackburn B, et al: A method for assessing the quality of a randomized control trial. *Controlled Clin Trials* 1981;2:31-49.
9. Mahoney MJ: Publication prejudices: An experimental study of confirmatory biases in the peer review system. *Cognitive Ther Res* 1977;1:161-175.
10. Searles JS: A methodological and empirical critique of psychotherapy outcome meta-analysis. *Behav Res Ther* 1985;23:453-463.
11. Landman JT, Dawes RM: Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. *Am Psychol* 1982;37:504-516.
12. Wortman PM, Yeaton WH: Cumulating quality of life results in controlled trials of coronary artery bypass graft surgery. *Controlled Clin Trials* 1985;6:289-305.
13. Chalmers TC, Celano P, Sacks HS, et al: Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-1361.
14. Yeaton WH, Wortman PM: Evaluation issues in medical research synthesis, in Yeaton WH, Wortman PM (eds): *Issues in Data Synthesis*. San Francisco, Jossey-Bass Inc Publications, 1984, pp 43-56.
15. Light RJ, Smith PV: Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educ Rev* 1971;41:429-471.
16. Yusef S, Peto R, Lewis J, et al: Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Prog Cardiovasc Dis* 1985;27:335-371.
17. Einarson TR, McGhain WF, Bootman JL, et al: Meta-analysis: Quantitative integration of independent research results. *Am J Hosp Pharm* 1985;42:1957-1964.
18. Bryant FB, Wortman PM: Methodological issues in the meta-analysis of quasi-experiments, in Yeaton WH, Wortman PM (eds): *Issues in Data Synthesis*. San Francisco, Jossey-Bass Inc Publications, 1984, pp 15-24.
19. Glass GV: Primary, secondary, and meta-analysis of research. *Educ Researcher* 1976;5:3-8.
20. Leviton LC, Cook TD: What differentiates meta-analysis from other forms of review? *J Pers* 1981;49:231-236.
21. Ottenbacher KJ, Petersen P: Quantitative reviewing of medical literature: An approach to synthesizing research results in clinical procedures. *Clin Pediatr* 1983;28:423-427.
22. Curlette WL, Cannella KS: Going beyond the narrative summarization of research findings: The meta-analysis approach. *Res Nurs Health* 1985;8:293-301.
23. Rosenthal R: Combining results of independent studies. *Psychol Bull* 1978;85:185-193.
24. Glass GV, McGraw B, Smith ML: *Meta-analysis in Social Research*. Beverly Hills, Calif, Sage Publications Inc, 1981.
25. Goldsmith JR, Beeser S: Strategies for pooling data in occupational epidemiological studies. *Ann Acad Med Singapore* 1984;13(suppl 2):297-307.
26. Sacks HS, Chalmers TC, Berk AA, et al: Should mild hypertension be treated? An attempted meta-analysis of the clinical trials. *Mt Sinai J Med* 1985;52:265-270.
27. Hunter JE, Schmidt FL, Jackson GB: *Meta-analysis: Cumulating Research Findings Across Studies*. Beverly Hills, Calif, Sage Publications Inc, 1982.
28. Klitgaard RF: Identifying exceptional performers. *Policy Analysis* 1978;4:529-547.
29. Hedges LV, Olkin I: *Statistical Methods for Meta-analysis*. Orlando, Fla, Academic Press Inc, 1985.
30. Pillemer DB, Light RJ: Synthesizing extremes: How to use research evidence from many studies. *Harvard Educ Rev* 1980;50:176-195.
31. Proceedings of the Workshop on Methodologic Issues in Overviews of Randomized Clinical Trials. *Stat Med* 1987;6:217-409.
32. L'Abbe KA, Detsky AS, O'Rourke K: Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224-233.
33. Himel HN, Liberati A, Gelber RD, et al: Adjuvant chemotherapy for breast cancer: A pooled estimate based on published randomized control trials. *JAMA* 1986;256:1148-1159.
34. Mumford E, Schlesinger HJ, Glass GV: The effects of psychological intervention on recovery from surgery and heart attacks: An analysis of the literature. *Am J Public Health* 1982;72:141-151.
35. Ottenbacher K, Difabio RP: Efficacy of spinal manipulation/mobilization therapy: A meta-analysis. *Spine* 1985;10:833-837.
36. Tran ZV, Weltman A: Differential effects of exercise on serum lipid and lipoprotein levels seen with changes in body weight: A meta-analysis. *JAMA* 1985;254:919-924.
37. Christensen E, Juhl E, Tygstrup N: Treatment of duodenal ulcer: Randomized clinical trials of a decade (1964 to 1974). *Gastroenterology* 1977;73:1170-1178.
38. Louis TA, Fineberg HV, Mosteller F: Findings for public health from meta-analysis. *Annu Rev Public Health* 1985;6:1-20.
39. Gilbert JP, McPeck B, Mosteller F: Progress in surgery and anesthesia: Benefits and risks of innovative therapy, in Bunker JP, Barnes BA, Mosteller F (eds): *Costs, Risks, and Benefits of Surgery*. New York, Oxford University Press Inc, 1977, pp 124-169.
40. Chalmers TC, Matta RJ, Smith H Jr, et al: Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977;297:1091-1096.
41. Stewart AL, Reynolds EOR, Lipscomb AP: Outcome for infants of very low birthweight: Survey of world literature. *Lancet* 1981;1:1038-1041.
42. Baum ML, Anish DS, Chalmers TC, et al: A survey of clinical trials of antibiotic prophylaxis in colon surgery: Evidence against further use of no-treatment controls. *N Engl J Med* 1981;305:795-799.
43. Blackburn BA, Smith H, Chalmers TC: The inadequate evidence for short hospital stay after hernia for varicose vein stripping surgery. *Mt Sinai J Med* 1982;49:383-390.
44. Stamfer MJ, Goldhaber SZ, Yusuf S, et al: Effect of intravenous streptokinase on acute myocardial infarction. *N Engl J Med* 1982;307:1180-1182.
45. Toth PJ, Horwitz RI: Conflicting clinical trials and the uncertainty of treating mild hypertension. *Am J Med* 1983;75:482-488.
46. Thacker SB: Quality of controlled clinical trials: The case of imaging ultrasound in obstetrics: A review. *Br J Obstet Gynaecol* 1986;92:437-444.
47. Colditz GA, Tuden RL, Oster G: Rates of venous thrombosis after general surgery: Combined results of randomized clinical trials. *Lancet* 1986;2:143-146.
48. Eddy DM: The use of confidence profiles to assess tissue-type plasminogen activator, in Califf RM, Wagner GS (eds): *Acute Coronary Care 1987*. Boston, Martinus Nijhoff Publishing Co, 1986, chap 7.
49. Schneider AP II: Breast milk jaundice in the newborn: A real entity. *JAMA* 1986;255:3270-3274.
50. Banta HD, Thacker SB: Assessing the costs and benefits of electronic fetal monitoring. *Obstet Gynecol Surv* 1979;34:627-642.
51. Thacker SB: The efficacy of intrapartum electronic fetal monitoring. *Am J Obstet Gynecol* 1987;156:24-30.
52. MacDonald D, Grant A, Sheridan-Pereira M, et al: The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol* 1985;152:524-539.
53. Chalmers I: Randomized controlled trials of intrapartum monitoring, in Thalhammer O, Baumgarten KV, Pollack A (eds): *Perinatal Medicine*. Stuttgart, West Germany, Georg Thieme Verlag, 1979, pp 260-265.
54. Leveno KJ, Cunningham FG, Nelson S, et al: A prospective comparison of selective and universal electronic fetal monitoring in 34995 pregnancies. *N Engl J Med* 1986;315:615-619.
55. Luthy DA, Shy KK, Van Bell G, et al: A randomized trial of electronic fetal monitoring in premature labor. *Obstet Gynecol* 1987;69:687-695.
56. Eysenck HJ: An exercise in mega-silliness. *Am Psychol* 1978;33:517.
57. Eysenck HJ: Meta-analysis: An abuse of research integration. *J Spec Educ* 1984;18:41-59.
58. Taveggia TC: Resolving research controversy through empirical cumulation. *Sociol Methods Res* 1974;2:395-407.
59. Chalmers TC, Levin H, Sacks HS, et al: Meta-analysis of clinical trials as a scientific discipline: I. Control of bias and comparison with large cooperative trials. *Stat Med* 1987;6:315-325.
60. Chalmers TC, Berrier J, Sacks HS, et al: Meta-analysis of clinical trials as a scientific discipline: II. Replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987;6:733-744.
61. DerSimonian R, Laird N: Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-188.
62. Mintz J: Integrating research evidence: A commentary on meta-analysis. *J Consult Clin Psychol* 1983;51:71-75.
63. Cook TD, Leviton LC: Reviewing the literature: A comparison of traditional methods with meta-analysis. *J Pers* 1980;48:449-472.
64. Cooper HM, Arkin RM: On quantitative reviewing. *J Pers* 1981;49:225-230.
65. Wilson GT, Rachman SJ: Meta-analysis and the evaluation of psychotherapy outcome: Limitations and liabilities. *J Consult Clin Psychol* 1983;51:54-64.
66. Sacks HS, Berrier J, Reitman D, et al: Meta-analysis of randomized controlled trials. *N Engl J Med* 1987;316:450-455.
67. Light RJ, Pillemer DB: Numbers and narrative: Combining their strengths in research reviews. *Harvard Educ Rev* 1982;52:1-26.
68. Cooper HM, Rosenthal R: Statistical versus traditional procedures for summarizing research findings. *Psychol Bull* 1980;87:442-449.
69. Bissonnette JM: Relationship between continuous fetal heart rate patterns and Apgar score in the newborn. *Br J Obstet Gynaecol* 1975;82:24-28.
70. Gabert HA, Stenchever MA: Continuous electronic monitoring of fetal heart rate during labor. *Am J Obstet Gynecol* 1973;115:919-923.
71. Schiffrin BS, Dame L: Fetal heart rate patterns, prediction of Apgar score. *JAMA* 1972;219:1322-1325.
72. Saldana LR, Schulman H, Yang W: Electronic fetal monitoring during labor. *Obstet Gynecol* 1976;47:706-710.
73. Tipton R, Shelley T: An index of fetal welfare in labour. *J Obstet Gynecol Br Commonwealth* 1971;78:702-706.