J Clin Epidemiol Vol. 45, No. 8, pp. 885–892, 1992
Printed in Great Britain.

0895-4356/92 $5.00 + 0.00

BIAS IN META-ANALYTIC RESEARCH

DAVID T. FELSON*

Boston University Arthritis Center, Departments of Medicine, Boston City Hospital and University Hospital, Boston, MA 02118, U.S.A.

(Received in revised form 16 January 1992)

Abstract—With the proliferation of meta-analyses in the medical literature have come conflicting studies. In addition, observance of guidelines for the performance of meta-analyses has been spotty. Bias may explain conflicting studies and differentiate carefully performed meta-analyses from others. Meta-analysts may fail to anticipate biases which threaten their study's validity. The three stages at which bias can be injected into a meta-analysis are finding studies, selection of the identified studies for the meta-analysis and extraction of data from the selected studies. This manuscript reviews specific types of bias which are common at each of these stages.

Bias Meta-analysis Epidemiologic methods

Meta-analysis is "a structured and systematic integration of information from different studies of a given problem" [1]. It is a technique which arose in evaluation research and psychology research and has gained wide popularity as a method of summarizing information about medical treatments and about the relationships of risk factors to disease. While meta-analyses were rare in the medical literature until the early 1980s, these studies have proliferated in the last 10 years. Meta-analyses have been influential in clarifying the value of therapies of myocardial infarction [2], breast cancer [3, 4] and other diseases [5].

With the profusion of meta-analyses has come the publication of conflicting studies. Furthermore, according to Sacks *et al.* [6], many meta-analysts have not observed accepted guidelines for the performance of these studies. It is the goal of this manuscript to comprehensively list and define biases in retrieving studies and extracting data from them that can occur in a meta-analysis that threaten its internal validity. These biases often cause conflicting meta-analysis results. Except for publication bias (see below), they have been neglected in medical journals, even though many have been well documented in social science literature. By recording a taxonomy of biases, it is hoped that the prevalence and importance of each can be systematically evaluated. We will suggest remedies, when available, for each type of bias.

As noted by Ingram Olkin, a statistician at Stanford, "doing a meta-analysis is easy, doing one well is hard." One of the central sources of difficulty in performing a meta-analysis is avoiding bias. This is true whether meta-analysis is defined in an orthodox way as the quantitative synthesis of studies on a particular issue or more broadly as the systematic weighing of studies evaluating the quantity and quality of information on a subject.

According to the **Dictionary of Epidemiology** [7], bias consists of "any trend in the collection, analysis, interpretation, publication or review of data that can lead to conclusions that are systemically different from the truth". In meta-analysis, a type of observational study, bias is a

*All correspondence to be addressed to: Dr D. Felson, A203, Boston University School of Medicine, 80 East Concord Street, Boston, MA 02118, U.S.A.

Table 1. Biases in meta-analysis

I. Biases in finding all studies (sampling bias)
(A) Publication bias-published vs unpublished studies
(B) Finding published studies using computerized database searches
(1) Indexing bias
(2) Search bias
(C) Finding published studies using bibliographic reviews or reference lists
(1) Reference bias
(2) Multiple publications bias
(3) Multiply used subjects bias
II. From studies identified to studies chosen
(A) Inclusion criteria bias
(B) Selector bias
III. Bias in obtaining accurate data from selected studies
(A) Bias by meta-analyst
(1) Extractor bias
(2) Bias in scoring study quality
(B) Bias whereby published report is not accurate in presenting study result
(1) Reporting bias
(2) Recording error (bias)

similar concept to that in epidemiology in that it distorts the valid comparison of two groups, usually a treatment–control comparison. Here, we shall restrict biases to those dealing with the selection of studies for a meta-analysis and the accurate recording of the study's results and will not focus on biases in analysis or interpretation of data.

The validity of a meta-analysis depends on complete sampling of all the studies performed on a particular topic. Validity can be preserved if a representative sampling of studies is obtained, but any incomplete sampling is a potentially biased one. Ideally, after all studies are sampled, the data from each are accurately summarized. The major areas of bias which arise in meta-analysis involve three steps (a) retrieval (or finding) of all studies, (b) selection of retrieved studies and (c) accurate extraction of the study data. The specific biases which can occur at each of these stages are listed in Table 1.

SAMPLING BIAS

Publication bias

As shown in Fig. 1, the failure to capture all studies performed on a topic, **retrieval bias**, can occur at any one of several stages. A completed study may be published or remain unpublished. **Publication bias** is the tendency of studies which report statistically significant results to be published. Reported initially in the social sciences [8], publication bias has been well documented in clinical trials in medicine, including oncology trials. For example, for combination chemotherapy in ovarian cancer, Simes [9] (Table 2) discovered that pooled results from published

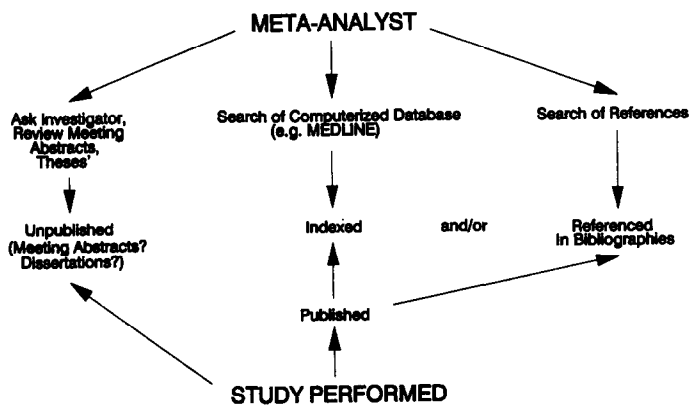

Fig. 1. The steps involved when a meta-analyst attempts to find all studies. Failure to find all studies (retrieval bias) can occur in 3 different ways corresponding to the 3 columns of this figure. Publication bias occurs when unpublished studies are missed (left column). Index bias or searcher bias can lead to failure to capture all indexed studies in a database (middle column). References bias is the overrepresentation of some published studies in published reference lists (right column).

Table 2. Publication bias in cancer trials*

	Published trials	Registered trials
Ovarian cancer combination chemotherapy		
No. of studies	16	13
Pooled p-value	0.02	0.24
Pooled median survival ratio	1.16	1.06
95% CI	1.06–1.27	0.97–1.15
Myeloma combination chemotherapy		
No. of studies	6	10
Pooled p-value	0.04	0.05
Pooled median survival ratio	1.26	1.14
95% CI	0.93–1.70	0.94–1.34

*From Simes, 1986 [7].

trials suggested significant efficacy, while data from prospectively registered trials (both published and unpublished) showed no significant advantage of combination chemotherapy over single agent treatment. For myeloma, Simes found that published trial data yielded a more favorable estimate of efficacy (survival ratio = 1.26) than data from registered trials (1.14). Unpublished studies are not only more likely to be statistically negative, they also have, on average, smaller sample sizes than published studies [10]. In studies approved by the Oxford Research Ethics Committee several years earlier. Easterbrook *et al.* [11] discovered a publication bias of observational studies with statistically significant findings (OR = 3.97, 95%CI 1.47–9.76). Surprisingly, they detected no publication bias for randomized trials with significant results (OR = 0.84).

In the social sciences, dissertations are often a rich source of unpublished studies. In medicine, unpublished studies are frequently presented at scientific meetings and published in abstract form. Unfortunately, abstracts do not usually provide sufficient information for data extraction in a meta-analysis, but at least they permit the existence of the unpublished study to be confirmed.

The desire by authors or editors to conform with or challenge currently held belief can motivate publication. A variation of publication bias is **conformity publication bias**, in which confirmatory studies are published, while studies that contradict currently held beliefs, whether null or not, are not published. Conformity bias may arise because reviewers and editors are unwilling to counter prevailing wisdom. The reverse of conformity bias is possible when new studies contradict a currently held belief and are published because they are newsworthy, whether null or not. Two recent examples of this are published null studies. First, Bradley *et al.* [12] published the results of a trial of osteoarthritis treatment which showed no difference between ibuprofen, an antiinflammatory, and acetaminophen, an analgesic. The recommended treatment has been antiinflammatories. This study's publication would have been less likely and it would have been less interesting had it confirmed the accepted notion that antiinflammatories are more effective than analgesics in this disease. In another example, Epstein *et al.* [13], in an observational study of rheumatoid arthritis patients, found that gold treated patients did no better than patients not treated with gold. This study probably would not have been published or submitted for publication if the authors had instead confirmed the widely-held belief in gold's efficacy. The point is that publication bias may be a function not just of statistical significance but of the ebb and flow of editorial and consensus opinion.

Publication bias can originate from three sources: the authors, the sponsor of the study, and the editor or reviewers of the journal to which the paper is submitted. Studies from the psychology [14] and medical [11] literature have documented that authors are less likely to submit manuscripts if they are statistically null, suggesting that the most important source of publication bias is the author. On the role of editors, Dickersin [15] cites statements by editors encouraging submission of manuscripts with positive results. Finally, the sponsor of the study may play an important role in generating publication bias, especially if it is a pharmaceutical company funded study. If a drug company sponsors a trial of its drug which turns out to be null, the company will likely discourage its publication [11].

There are several ways to assess the magnitude of publication bias in a meta-analysis. Light and Pillemer [16] recommend a funnel plot (See Fig. 2). In such a plot, the effect size of studies is plotted against study sample size. If there were no publication bias, the plot would resemble an inverted funnel with a wide dispersion of results among studies of small sample size and a narrower range of study results for large studies. If one portion of the funnel is missing, then publication bias is likely. Usually, this consists of an absence of published negative small studies. Publication bias can also exist when there are no studies published whose effect size is 0 but rather, on the one hand, positive studies and, on the other, studies that show

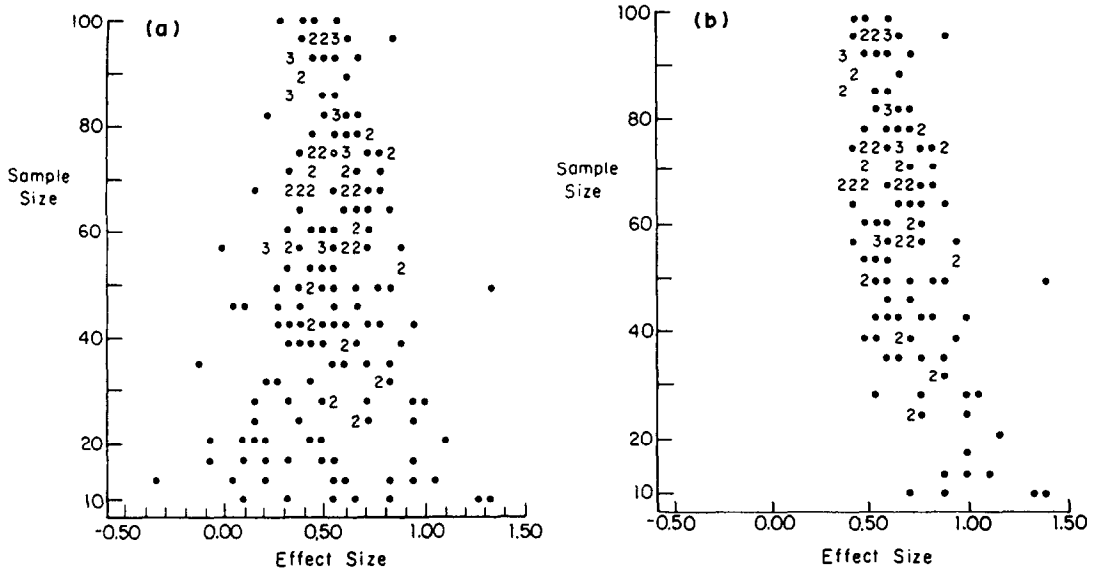

Fig. 2. Funnel plots to assess publication bias. (a, left) Shows full funnel with no publication bias. (b, right) Shows publication bias in which null small studies were not published. [Reprinted by permission of the publishers from: Light RJ, Pillemer DD. **Summing up: The Science of Reviewing Research.** Cambridge, MA: Harvard University Press; Copyright © 1954 by the President and Fellows of Harvard College.]

significant negative effect sizes. In this case the funnel will have a vertical gap or hole. Others have suggested methods for computing the number of unpublished studies that would be needed to negate a meta-analyzed conclusion [14]. If the number of such studies is very large, as when the pooled effect size is far from the null or component studies ae large, then publication bias is unlikely to have an effect on the meta-analyzed conclusion.

In addition to graphing study results to evaluate the chance of publication bias or computing a "fail-safe" number of studies, there are a few ways to avoid publication bias.

First, results of large studies most closely approximate the average result of all studies, published and unpublished. Furthermore, large studies, even with null results, are almost always published [10]. Therefore, the meta-analyst can test the pooled results of large studies to see if they approach the overall pooled result.

In addition, the impact of publication bias may be diminished if a meta-analysis focuses on a question tangential to the theme of the individual studies it evaluates. Examples include meta-analytic investigations of trial methods (e.g. which outcomes are sensitive to change) and meta-analyses which compare effect sizes of two different therapies, each studied in their own set of trials. The latter example builds in publication bias and assumes that it is similar for each drug.

To avoid publication bias, a meta-analyst can also attempt to obtain data from unpublished studies, an endeavor recommended [11, 17]. Nonetheless, finding those studies can be very difficult and incorporating data from unpublished studies introduces several unresolvable quandaries [18]; how does the meta-analyst ensure that those studies are evaluated the same as published studies? Is data likely to be as accurate as in published studies?

Ultimately, one important solution to publication bias may be the establishment of clinical trial registries which include published and unpublished studies.

Retrieval bias in published studies

In addition to publication bias, there are types of bias involved in finding published studies (see Fig. 1). If the study is published, it is indexed and/or referenced in bibliographic reviews or reference lists. Since MEDLINE began indexing articles in 1966, studies published earlier will not be in the database.

One way to avoid bias in finding published articles is to perform a rigorous computerized database search. The capture rate of such a search depends on whether articles were indexed correctly and consistently. Even expert library-trained reviewers have failed to capture a large percentage of identified studies on a subject (see Table 3), suggesting either indexing errors or indexing variability. Indexing bias is defined as biased indexing of published studies. Indexer

Table 3. The capture rate of MEDLARS searches compared with known trials from registers or manual searches

Source	Expert (E) or amateur (A)	Subject	Proportion of studies captured (sensitivity)
Dickersin *et al.*, 1985 [13]	E	Neonatal Hyperbilirubinemia	28/88 (32%)
	A	Neonatal Hyperbilirubinemia	17/88 (19%)
Dickersin *et al.*, 1985 [13]	E	Intraventricular Hemorrhage	19/29 (66%)
	A	Intraventricular Hemorrhage	11/29 (38%)
Poynard and Conn, 1985 [14]	A	Clinical trials in hepatobiliary disease	107/208 (51%)*
Bernstein, 1988 [15]	E	Clinical trials in hepatobiliary disease	155/195 (80%)†

*When MEDLARS search was combined with references obtained from MEDLARS-identified trials, 125 trials were found. This was compared to manual search and references secondarily obtained which found 244 trials (51% capture rate).

†Multiple search strategies boosted capture at the expense of search precision. 9643 studies were found by MEDLARS and had to be reviewed manually to capture 155 trials.

consistency averages 45–50% [15] and may be even lower than this for index terms about trials such as "clinical trial", or "epidemiological method". The quality and consistency of indexing depends on whether articles contain clear descriptions of methods and study content. Since indexing bias is not under the meta-analyst's control, it should be recognized as a potential problem that can be neutralized by using multiple overlapping sources of study retrieval.

In short, many published articles are not discovered even after an expert search. This all assumes that the meta-analyst embarking on a database search chooses appropriate index terms and approaches the search with a systematic strategy. The casual database search is likely to miss an even larger percentage of studies than an expert search (see Table 3) [19–22]. This may result in **search bias**, another type of sampling bias which is a bias in captured studies resulting from an inadequate or incomplete search. The cure for search bias is a careful, informed search strategy.

If the data base search fails to capture studies, the meta-analyst than depends on reference lists from articles or on personal knowledge of studies. The likelihood of being referenced may depend on the prominence of the study or whether literature reviewers regard its results favorably. This may introduce considerable **reference bias**, a tendency of certain studies to be cited while others are not. The meta-analyst cannot necessarily avoid reference bias except to be aware of it and use overlapping search strategies. In fact, relying mostly on references published in other articles or in reviews of literature may build sampling bias into a meta-analysis.

Another factor contributing to a biased retrieval of studies may be **multiple publications bias**. This occurs when studies whose results are published in a series of articles are more likely to be sampled than those published only once. Multiple publications can induce meta-analyst confusion when the publications do not have the same first author or when one publication does not refer to the prior one. We discovered one example of a clinical trial in rheumatoid arthritis in which the two articles reporting its results had no authors in common [23, 24]! In these circumstances, meta-analysts may mistakenly, but understandably, assume that more than one study was performed, resulting in a double counting of the study. Vigilance in recognizing one study published repeatedly will lessen the likelihood of including several publications from the same study in meta-analysis.

Different types of errors may occur in finding studies as noted by Glass *et al.* [8]: "Locating studies is the stage at which the most serious form of bias enters a meta-analysis, since it is difficult to assess the impact of a potential bias. The best protection against this source of bias is a thorough description of the procedures used to locate the studies that were found so that the reader can make an intelligent assessment of the representativeness and completeness of the data base for a meta-analysis".

An additional bias, **multiply used subjects bias** can occur when the same subjects are reported in two separate studies when they were actually a part of only one study, a phenomenon which occurred in a meta-analysis of the therapy of lupus nephritis [5].

SELECTION BIAS

Once studies are captured by the search procedure, a meta-analyst then chooses among studies for the meta-analysis. At this juncture, two other types of bias are possible. One is **inclusion criteria bias**. This bias can occur when the investigator creates a set of inclusion criteria based upon a preliminary review of the literature. However, these criteria could purposely exclude some important studies which the meta-analyst knows of, and this would, in turn, produce bias. Inclusion criteria bias is difficult to avoid since a good knowledge of a topic is a prerequisite to development of inclusion criteria (e.g. outcome measures which are widely accepted and occur with reasonable frequency should be chosen as should interventions or exposures which are relevant to clinical practice). Ultimately, the meta-analyst must honestly attempt to create inclusion criteria most relevant to the issue being studied, irrespective of known results of clinical studies. Inclusion criteria bias has not been well described and is difficult to quantitate. It is probably common and likely accounts for conflicting meta-analyses, especially if the number of included studies is small and/or a few studies are large and influential. Inclusion criteria bias could have been responsible for the conflicting results of meta-analyses investigating steroids and gastrointestinal bleeding (see below).

A related tyupe of bias is **selector bias**. In selector bias, inclusion criteria have been set, although they may not be so specific as to dictate which studies are included or excluded from the meta-analysis. This leaves the meta-analyst selector free to choose studies, a choice which is susceptible to bias. Several suggested methods can limit selector bias. The most common is to blind methods and results of studies to make it hard for the meta-analyst selector to determine the study results. In this method, there are often two selectors who work independently. Any disagreement in study selection is arbitrated by a joint meeting or by a third selector. This process certainly decreases the chance of selector bias, but it does not eliminate it. Blinding of studies is often difficult, and some important studies will be so prominent that selectors will be familiar with them, even though study results are masked. One important way to limit selection bias is to create extremely specific and clear study inclusion criteria so that the selector has little leeway to inject bias into the selection decision.

Selection bias of studies is probably the central reason for discrepant results in meta-analyses. For example, in one meta-analysis evaluating the relationship of corticosteroids to gastrointestinal (GI) bleeding, Conn and Blitzer [25] suggested no significant relationship between steroids and GI bleeding. However, a later meta-analysis on the same issue [26] found a significant, albeit small, increased risk of GI bleeding among steroid users. The main difference between these two meta-analyses was that the second study contained a larger and overlapping group of primary studies and excluded several of the primary studies that were included in the first. Other repeat meta-analyses have had markedly different study samples [27], but this has not always led to discrepant meta-analytic results.

WITHIN STUDY BIASES

Once studies are selected for a meta-analysis, data should be accurately extracted from the study. There are several opportunities for bias here, both on the part of the meta-analyst and in the study report. From the perspective of the meta-analyst, the most likely bias is **extractor bias**, which occurs when the data is not extracted accurately from the study. Although it may be random, extractor bias can create systematically biased results. There may be considerable inter- and intraobserver variability in extracting data from studies. For example, Smith and Glass [28], in a meta-analysis focusing on the therapeutic efficacy of psychoactive drugs vs psychotherapy, tested the reliability of two judges and compared them to a third judge in obtaining data from five studies. Judges extracted information on variables such as patient age and duration of treatment. Of these data, 75% of extractions were identical for both judges, 80% were within one or two scale points on a five point scale, and 17% were placed in a wrong category or were off by more than one scale point. More importantly, for the outcome of the study, an effect size extracted in duplicate, the average score was 0.60 with an average

difference between the two judges of 0.07, a difference slightly >10% of the effect size. Therefore, interobserver variability in noting outcome data from these trials was modest, but there were major observer differences in extracting other data from the studies. Rosenthal [29] has reported a wide variation in the consistency of data extraction from studies.

To maximize interobserver reliability and minimize extractor bias, an extraction sheet should lay out specific rules for data extraction with clarity. Furthermore, extraction of results from trials may be troubled when a series of calculations must be made by the extractor. Another way of minimizing the variability of data extraction is to have only one data extractor, although intraobserver consistency may still be low if rules for data extraction are not well defined in advance. Furthermore, a single extractor may infuse their own idiosyncrasies. Therefore, if resources permit, it is best to have two extractors perform the work in duplicate, following a specific extraction sheet.

Meta-analyst bias may affect the scoring of studies for quality. If study results are weighted for quality in the analysis, **a bias in scoring study quality** may have a real impact on meta-analysis results. Once again, rigid *a priori* rules on how to measure the quality of trials may help lessen observer variability and mitigate bias.

Even if the meta-analyst carefully avoids all biases by standardizing extraction instruments, the primary study paper itself may not accurately report the study's result. One type of bias that can be introduced is **reporting bias**. In this bias, the study has several outcomes which were measured, but the only results reported are those which reach statistical significance. In a recent meta-analysis evaluating rheumatoid arthritis second line drugs [30], we found trials that had included up to ten efficacy outcomes, yet reported numerically only those that reached significance. Trial reports either did not mention other end points or reported them in the text as having no significant change. Including these trials in the meta-analysis would have introduced bias in favor of the drug studied. Therefore, when reporting bias could be identified, we excluded the trial from consideration. An alternative might have been to assume that the nonreported results were statistically null and score them as having effect sizes of 0. Unfortunately, some trial reports may not reveal that multiple outcomes were assessed but may report on only those that had positive results. The prevalence of reporting bias is unknown, but we suspect it is a widespread problem which could serve to substantially bias meta-analysis results.

Also, trial reports can contain inconsistency of results. For instance, such data as the number of dropouts or the number of patients experiencing toxicity may be reported differently in different parts of the study report.

Another type of misinformation is **recording error bias** which is when the actual study results and the recorded results in the published paper differ. Rosenthal [29] meta-analyzed 27 studies in which investigators had checked for errors in the transfer of collected data to the actual trial report. He found an error rate of approx. 1% for all data. Interestingly, about two thirds of the errors were biased in favor of the observer's hypothesis, suggesting that the data errors were not random. Furthermore, the errors often led to study results which were barely significant ($p < 0.05$), while the real data, properly analyzed, yielded p values that did not quite reach significance. While recording errors of data in trials do not, therefore, appear to be a problem of great magnitude, they add another element into the imperfect validity of a meta-analyzed result.

We have not covered intrastudy bias, such as a clinical trial with flawed (biased) randomization or a case–control study with a selection bias of controls. Including biased studies in a meta-analysis can certainly threaten the validity of a meta-analyzed result. Furthermore, we have not discussed analysis procedures in meta-analyses or methods of combining data from different studies which are well covered elsewhere (e.g. [29, 31]). If bias consists of misrepresentation of analytic results, performing incorrect analyses [32] or misinterpreting results, then these latter stages of a meta-analysis are also susceptible to important biases.

In summary, bias can occur at multiple steps in the process of meta-analysis. To perform a valid meta-analysis, the investigator must carefully avoid many of the problems elucidated here. A thorough search for all studies using multiple sources of retrieval is necessary. An honest set of specific inclusion criteria which fit the question of interest and not the results of studies is needed. Finally, there should be a rigorous attempt using a detailed and comprehensive instrument to extract data from trials.

Acknowledgments—The author appreciates the helpful comments of Robert Meenan and Elaine Landes and the

technical assistance of Gaynell Walker and Pearl Siew. This work was supported by Arthritis Center Grant AR 20613 from the NIH.

REFERENCES

1. Jenicek M. Meta-analysis in medicine: where we are and where we want to go. **J Clin Epidemiol** 1989; 42: 35–44.
2. Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH. Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. **N Engl J Med** 1982; 307: 1180–1182.
3. Early Breast Cancer Trialists' Collaborative Group: Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomized trials among 28,896 women. **N Engl J Med** 1988; 319: 1681–1692.
4. Himel NH, Liberati A, Gelber RD, Chalmers TC. Adjuvant chemotherapy for breast cancer: a pooled estimate based on published randomized control trials. **JAMA** 1986; 256: 1148–1159.
5. Felson DT, Anderson J. Evidence for the superiority of immunosuppressive drugs and prednisone over prednisone alone in lupus nephritis: results of a pooled analysis. **N Engl J Med** 1984; 331: 1528–1533.
6. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. **N Engl J Med** 1987; 316: 450–455.
7. Last JM. **A Dictionary of Epidemiology.** 2nd edn. New York: Oxford University Press; 1988: 13–14.
8. Glass GV, McGaw B, Smith ML. **Meta-analysis in Social Research.** Beverly Hills: Sage; 1987.
9. Simes RJ. Publication bias: The case for an international registry of clinical trials. **J Clin Oncol** 1986; 4: 1529–1541.
10. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. **J Natl Cancer Inst** 1989; 81: 107–114.
11. Easterbrook PJ, Berlin JA, Gopalan T *et al.* Publication bias in clinical research. **Lancet** 1991; 337 (8746): 867–872.
12. Bradley JD, Brandt KD, Barry PK *et al.* Comparison of an antiinflammatory dose of ibuprofen, and analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. **N Engl J Med** 1991; 325: 87–91.
13. Epstein WV, Henke CJ, Yelin EH, Katz PP. Effects of parentally administered gold therapy on the course of adult rheumatoid arthritis. **Ann Intern Med** 1991; 114: 437–443.
14. Rosenthal R. The "file-drawer problem" and tolerance for null results. **Psychol Bull** 1979; 86: 638–641.
15. Dickerson K. The existence of publication bias and risk factors for its occurrence. **JAMA** 1990; 263: 1385–1389.
16. Light RJ, Pillemer DB. Quantitative procedures. In **Summing up: The Science of Reviewing Research.** Cambridge, MA: Harvard University Press; 1984.
17. Collins R, Gray R, Godwin J *et al.* Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. **Stat Med** 1987; 6: 245–250.
18. Chalmers TC, Levin H, Sacks H *et al.* Meta-analysis of clinical trails as a scientific discipline. I: Control of bias and comparison with large co-operative trials. **Stat Med** 1987; 6: 315–325.
19. Hewitt P, Chalmers TC. Perusing the literature: Methods of accessing MEDLINE and related databases. **Controlled Clin Trials** 1985; 6: 168–177.
20. Dickersin K, Hewitt P, Mutch L, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. **Controlled Clin Trials** 1985; 6: 306–317.
21. Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature: A comparison of MEDLARS and manual methods. **Controlled Clin Trials** 1985; 6: 271–279.
22. Bernstein F. The retrieval of randomized clinical trials in liver diseases from the medical literature: manual versus MEDLARS searches. **Controlled Clin Trials** 1988; 9: 23–31.
23. Gofton JP, O'Brien WM, Hurley JN, Scheffler BJ. Radiographic evaluation of erosion in rheumatoid arthritis: double blinded study of auranofin vs placebo. **J Rheumatol** 1984; 11: 768–771.
24. Katz WA, Alexander S, Bland J *et al.* The efficacy and safety of auranofin compared to placebo in rheumatoid arthritis. **J Rheum** 1982; 9: 173–178 (Suppl 8).
25. Conn HO, Blizter BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. **N Engl J Med** 1983; 294: 473–479.
26. Messer J, Reitman D, Sacks HS *et al.* Association of adrenocorticosteroid therapy and peptic-ulcer disease. **N Engl J Med** 1983; 309: 21–24.
27. 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.
28. Smith ML, Glass GV. Metaanalysis of psychotherapy outcome studies. **Am Psychol** 1977; 32: 752–760.
29. Rosenthal R. **Meta-Analytic Procedures for Social Research,** 2nd edn. Beverly Hills: Sage; 1986: 63–123.
30. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis: results of two metaanalyses. **Arthritis Rheum** 1990; 33: 1449–1461.
31. Hedges LV, Olkin I. **Statistical Methods for Meta-Analysis.** Orlando, FL: Academic Press; 1985.
32. Goldman L, Feinstein AR. Anticoagulants and mycardial infarction: The problems of pooling, drowning, and floating. **Ann Intern Med** 1979; 90: 92–94.