

3 Problems and limitations in conducting systematic reviews

MATTHIAS EGGER, KAY DICKERSIN,
GEORGE DAVEY SMITH

Summary points

- There are numerous ways in which bias can be introduced in reviews and meta-analyses of controlled clinical trials.
- If the methodological quality of trials is inadequate then the findings of reviews of this material may also be compromised.
- Publication bias can distort findings because trials with statistically significant results are more likely to get published, and more likely to be published without delay, than trials without significant results.
- Among published trials, those with significant results are more likely to get published in English, more likely to be cited, and more likely to be published more than once which means that they will also be more likely to be identified and included in reviews.
- The choice of the outcome that is reported can be influenced by the results. The outcome with the most favourable findings will generally be reported, which may introduce bias.
- Criteria for inclusion of studies into a review may be influenced by knowledge of the results of the set of potential studies.
- The definition of eligibility criteria for trials to be included, a comprehensive search for such trials, and an assessment of their methodological quality are central to systematic reviews. Systematic reviews are thus more likely to avoid bias than traditional, narrative reviews.

Systematic reviews and meta-analyses have received a mixed reception since the outset. Those on the receiving end have rejected what they see as exercises in “mega-silliness”¹ and the authors of a highly distinguished series of systematic reviews of care during pregnancy and childhood² have been dubbed as terrorists (“an obstetrical Baader-Meinhof gang”³). Some

statisticians think that meta-analysis “represents the unacceptable face of statisticism”⁴ and to clinicians objecting to the findings of meta-analyses “a tool has become a weapon”.⁵ Others “still prefer the conventional narrative review article”.⁶ At the other end of the spectrum, the application of a technique which basically consists of calculating a weighted average has been described as “Newtonian”⁷ and it has been suggested that with the advent of meta-analysis there is no place left for the narrative review article.⁸ As may be imagined, the truth is likely to lie somewhere between these extreme views.

This mixed reception is not surprising considering that several examples exist of meta-analyses of small trials whose findings were later contradicted by a single large randomised trial (Figure 3.1).^{9,10} Also, systematic reviews addressing the same issue have reached opposite conclusions.¹¹ For example, one group reviewing trials comparing low-molecular-weight (LMW) heparins and standard heparin in the prevention of thrombosis following surgery concluded that “LMW heparins seem to have a higher benefit to risk ratio than unfractionated heparin in preventing perioperative thrombosis”,¹² while another group of reviewers considered that “there is at present no convincing evidence that in general surgery patients LMW heparins, compared with standard heparin, generate a clinically important improvement in the benefit to risk ratio”.²¹ The differences between these

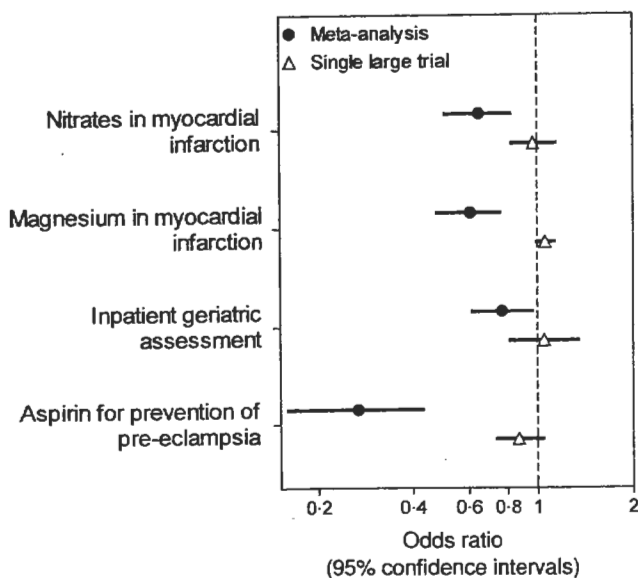


Figure 3.1 Results from discordant pairs of meta-analyses of small trials and single large trials: effect of nitrates^{13,14} and magnesium^{15,16} on mortality in acute myocardial infarction, effect of inpatient geriatric assessment on mortality in the elderly,^{17,18} and effect of aspirin on the risk of pre-eclampsia.^{19,20}

Table 3.1 Characteristics of two systematic reviews of clinical trials which addressed the same issue but reached opposite conclusions: low-molecular-weight (LMW) heparin compared with standard heparin in the prevention of perioperative thrombosis.

Characteristic	Leizorovicz <i>et al.</i> ¹²	Nurmohamed <i>et al.</i> ²¹
Literature search		
Years covered	1984-91	1984-91
Search for unpublished data	Yes	No
Language restrictions	None	English, German, French
Accepted methods to detect deep-vein thrombosis	Fibrinogen uptake test, thermographic DeVeTherm test, Doppler ultrasonography, phlebography	Fibrinogen uptake test in general surgery patients, phlebography for orthopaedic surgery patients
Assessment of trial quality	No	Eight-point scale
Analysis		
Number of studies included	39	23*
Number of patients included	12 375	8172
Statistical model	Fixed effects	Fixed effects
Stratified by trial quality	No	Yes
Conclusions	LMW heparins have higher benefit to risk ratio in preventing perioperative thrombosis	There is no convincing evidence that in general surgery patients LMW heparins have a higher benefit to risk ratio

* Thirteen studies were considered to be of high methodological quality.

reviews are summarised in Table 3.1. The literature search was free of language restrictions and attempts were made to identify unpublished studies in one review¹² but not the other.²¹ One group of reviewers²¹ based their conclusion on a subgroup of trials, which they considered to possess the highest methodological strength, while the other group¹² did not consider the quality of trials.

Contrary to one of the central objectives of systematic reviews, to reduce uncertainty, such contradictory reports may contribute to the confusion, a situation that has arisen in other fields, for example when assessing calcium antagonists or cholesterol-lowering interventions in hypertension and coronary heart disease, or mammography for breast cancer screening.²²⁻²⁴ In this chapter we will review the problems and limitations of systematic reviews and meta-analyses.

Garbage in – garbage out?

The quality of component trials is of crucial importance: if the “raw material” is flawed, then the findings of reviews of this material may also be compromised. Clearly, the trials included in systematic reviews and meta-analyses should ideally be of high methodological quality and free of bias

such that the differences in outcomes observed between groups of patients can confidently be attributed to the intervention under investigation. The biases that threaten the validity of clinical trials are reviewed in detail in Chapter 5. These relate to systematic differences in the patients' characteristics at baseline (*selection bias*), unequal provision of care apart from the treatment under evaluation (*performance bias*), biased assessment of outcomes (*detection bias*), and bias due to exclusion of patients after they have been allocated to treatment groups (*attrition bias*).²⁵ Several studies²⁶⁻²⁸ have recently attempted to quantify the impact these biases have on the results of controlled clinical trials (see Chapter 5). For example, Schulz *et al.*²⁶ assessed the methodological quality of 250 trials from 33 meta-analyses from the Cochrane Pregnancy and Childbirth Database and examined the association between dimensions of trial quality and estimated treatment effects. Compared to trials in which authors reported adequately concealed treatment allocation, failure to prevent foreknowledge of treatment allocation or unclear concealment were associated, on average, with an exaggeration of treatment effects by 30 to 40%. Trials that were not double-blind also yielded larger effects.

The meta-analyses^{12,21} of trials comparing LMW heparin with standard heparin for the prevention of postoperative deep-vein thrombosis mentioned earlier are a case in point. Jüni *et al.*²⁹ recently showed that in these trials, blinding of outcome assessments is a crucial quality feature: trials that were not double-blind showed a spurious benefit of LMW heparin that disappeared when restricting the analysis to trials with blinded outcome assessment. This is not entirely surprising considering that the interpretation of fibrinogen leg scanning, which is used to detect thrombosis, can be subjective.³⁰ One of the two reviews summarised in Table 3.1 produced discordant results precisely because the authors chose to ignore the quality of component trials. It is somewhat ironic that the same reviewers were considerably more thorough in their attempt to identify all relevant trials, independent of publication status or language of publication. Although the quality of component trials happened to be more important in this particular situation, the dissemination of findings from clinical trials is known to be biased, and a comprehensive literature search is an essential ingredient of high-quality reviews.

The dissemination of research findings

The dissemination of research findings is not a dichotomous event but a continuum ranging from the sharing of draft papers among colleagues, presentations at meetings, published abstracts to papers in journals that are indexed in the major bibliographic databases.³¹ It has long been recognised that only a proportion of research projects ultimately reach publication in

an indexed journal and thus become easily identifiable for systematic reviews.³² Scherer *et al.*³³ showed that only about half of abstracts presented at conferences are later published in full (see Box 3.1). Dickersin and Meinert examined the fate of doctoral theses from the Department of Epidemiology at Johns Hopkins University School of Hygiene and Public Health and found that one-third of graduates had not published a single article from their thesis (see Box 3.2). Similar results were found for trainees in public health in the UK.⁴⁴ Four separate studies followed up research

Box 3.1 Full publication of results initially published as abstracts

The dissemination of research findings follows a continuum from oral report to published abstract to full publication in an indexed and accessible journal. Since the main purpose of publication is to report a study in sufficient detail to allow critical appraisal and decision-making on clinical or other questions, neither oral nor "abstract" presentations are considered sufficient to qualify as "complete" dissemination. In a 1994 systematic review, Roberta Scherer and colleagues³³ summarised the results from 11 studies³³⁻⁴³ describing subsequent full publication of research initially presented in abstract or short report form. To be included, studies had to have followed published abstracts for at least two years to assess full publication. Studies followed a total of 2391 abstracts published in various fields of medicine, including vision research, anaesthesiology, perinatology, and paediatrics. The authors obtained a weighted average rate of full publication of abstracts by weighting by the square root of the total number of abstracts in each report. The average rate of full publication was 51% (95% confidence interval 45% to 57%), and individual study rates ranged from 32% to 66%. Average publication rates were similar for the two studies^{33,34} that were confined to randomised controlled trials (50%). The data from eight reports that included data on cumulative rates of publication are summarised in Figure 3.2. The findings from this systematic review are reason for concern. On average, only half of health-related abstracts are published in full, which means that those performing systematic reviews should not omit abstracts from their consideration. While publication in abstract form is better than no publication at all, the format does not allow presentation of methodology or other details that allow the reader to critically assess the findings. And, since investigators change institutions and are often poor responders to mailed or telephoned request, one cannot rely on contact with authors to fill in the details if only abstracts are available. Abstracts are also difficult to locate, as most appear in conference proceedings and these are not typically indexed in bibliographic databases. This situation leads reviewers to focus their analyses mainly on data

from full text reports, which may lead to biased and incorrect conclusions (see main text). Scherer is currently undertaking an update of her systematic review for the Cochrane Collaboration.

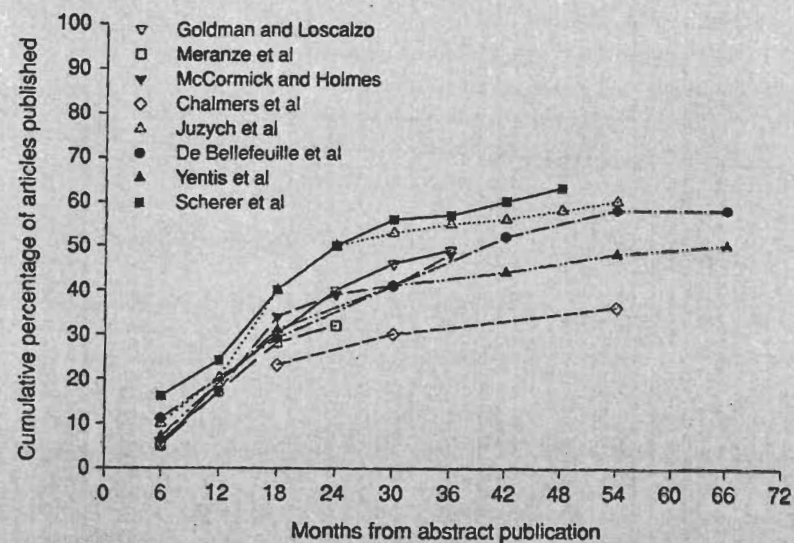


Figure 3.2 Percentages of total abstracts published over time, calculated for eight studies^{33,35-40,42} that followed up research presented at meetings and conferences.

proposals approved by ethics committees or institutional review boards in Oxford,⁴⁵ Sydney,⁴⁶ and at the Johns Hopkins School of Medicine⁴⁷ and School of Hygiene and Public Health in Baltimore.⁴⁷ For each cohort of research proposals the principal investigators were contacted several years later in order to determine the publication status of each completed study. The rates of full publication as journal articles ranged from 49 to 67% (Table 3.2). Similarly, 20% of trials funded by the National Institutes of Health (NIH) and 45% of trials on HIV infection funded by the National Institute of Allergy and Infectious Diseases (NIAID) were still unpublished several years after completion.⁴⁸⁻⁵⁰ The fact that a substantial proportion of studies remains unpublished even a decade after the study had been completed and analysed must be of concern as potentially important information remains hidden from reviewers. Making things worse, the dissemination of research findings is not a random process; rather it is strongly influenced by the nature and direction of results. Statistically

Box 3.2 Publication of doctoral dissertations

A study on publication bias, conducted at the Johns Hopkins University Schools of Medicine and Hygiene and Public Health, found that projects associated with thesis work (doctoral and master's level) were less likely than other work to be published.⁴⁷ As a follow-up to these findings, Kay Dickersin and Curtis Meinert surveyed 90 graduates of the Department of Epidemiology at Johns Hopkins University School of Hygiene and Public Health who received a doctorate in the years 1967 to 1987 and asked the graduates whether they had published their dissertation work in a full-length report. Work was considered published if it appeared in a journal or book. Eighty-one graduates responded. Overall, 67.9% of respondents published at least one full text report; 8.6% published an abstract only; and 4.9% made a presentation only. Ninety-six per cent of those who published a full-length article did so in a MEDLINE-indexed journal. Publication did not appear to be related to year the degree was granted, current employment status, type of degree (DrPH, PhD, or ScD), or sex of the student. Although numbers were small and differences statistically non-significant, several observations from the study bear testing in future studies of dissertation publication experience. Men and women students published in similar proportions (72.5% versus 60.0%, respectively). Students with women advisors, however, had better publication records compared with those with men advisors (87.5% versus 63.1%). Women students with men for advisors had the lowest publication rate of all combinations (47.6%). The publication rate observed for 1967 to 1987 epidemiology doctoral graduates (67.9%) was similar to that for Johns Hopkins School of Public Health faculty in 1980 (66%), and for British public health trainees (70%).⁴⁴ The study represents a special and small population. One might assume, however, that publication rates for graduates of Johns Hopkins are at least as high as those for graduates of other programmes. These findings imply that those performing systematic reviews must include searches of dissertations to assure comprehensive identification of all relevant studies. In addition, it is likely that "positive", statistically significant findings are selectively published from theses. In education research, it has been shown that findings published in journals were more supportive of the hypotheses favoured by the investigators than findings published in theses or dissertations.⁵¹

significant, "positive" results that indicate that a treatment works are more likely to be published, more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, and more likely to be cited by others. When discussing these reporting biases, which are summarised in Table 3.3, we will denote trials with statistically significant ($P < 0.05$) and non-significant results as trials with "positive" and "negative"

Table 3.2 Publication status of four cohorts of research projects approved by ethics committees or institutional review boards which had been completed and analysed at the time of follow up (adapted from Dickersin.⁵⁰)

Reference	Johns Hopkins University, Baltimore		Central Research Ethics Committee Oxford		Royal Prince Alfred Hospital, Sydney	
	Medicine	Public Health				
Number approved	Dickersin <i>et al.</i> ⁴⁷ 342 (100%)	172 (100%)	Easterbrook <i>et al.</i> ⁴⁵ 285 (100%)		Stern and Simes ⁴⁶ 321 (100%)	
Period of approval	1980	1980	1984-87		1979-88	
Year of follow up	1988	1988	1990		1992	
Published	230 (67%)	104 (61%)	138 (49%)		189 (59%)	
Full publication	36 (11%)	7 (4%)	69 (24%)		n.a.	
Abstract only	11 (3%)	2 (1%)	0 (0%)		0 (0%)	
Other/unclear	65 (19%)	59 (34%)	78 (27%)		132 (41%)	
Unpublished						

n.a. = not assessed.

Table 3.3 Reporting biases: definitions

Type of reporting bias	Definition
Publication bias	The <i>publication</i> or <i>non-publication</i> of research findings, depending on the nature and direction of the results
Time lag bias	The <i>rapid</i> or <i>delayed</i> publication of research findings, depending on the nature and direction of the results
Multiple (duplicate) publication bias	The <i>multiple</i> or <i>singular</i> publication of research findings, depending on the nature and direction of the results
Citation bias	The <i>citation</i> or <i>non-citation</i> of research findings, depending on the nature and direction of the results
Language bias	The publication of research findings in a <i>particular language</i> , depending on the nature and direction of the results
Outcome reporting bias	The <i>selective reporting</i> of some outcomes but not others, depending on the nature and direction of the results

results. However, the contribution made to the totality of the evidence by trials with non-significant results is as important as that from trials with statistically significant results.

Publication bias

In a 1979 article on "The 'file drawer problem' and tolerance for null results" Rosenthal described a gloomy scenario where "the journals are filled with the 5 per cent of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant (e.g., $P > 0.05$) results."⁵² The file drawer problem has long been recognised in the social sciences: a review of psychology journals found that of 294 studies published in the 1950s, 97.3% rejected the null hypothesis at the 5% level ($P < 0.05$).⁵³ The study was recently updated and complemented with three other journals (*New England Journal of Medicine*, *American Journal of Epidemiology*, *American Journal of Public Health*).⁵⁴ Little had changed in the psychology journals (95.6% reported significant results) and a high proportion of statistically significant results (85.4%) was also found in the general medical and public health journals. Similar results have been reported for emergency medicine⁵⁵ and, more recently, in the area of alternative and complementary medicine.^{56,57} It is thus possible that studies which suggest a beneficial treatment effect are published, while an equal mass of data pointing the other way remains unpublished. In this situation, a systematic review of the published trials could identify a spurious beneficial treatment effect, or miss an important adverse effect of a treatment. In

the field of cancer chemotherapy such *publication bias* has been demonstrated by comparing the results from studies identified in a literature search with those contained in an international trials registry^{58,59} (see Box 3.3). In cardiovascular medicine, investigators who, in 1980, found an increased death rate among patients with acute myocardial infarction treated with a class I anti-arrhythmic dismissed it as a chance finding and did not publish

Box 3.3 A demonstration of publication bias

Studies with statistically significant results are more likely to get published than those with non-significant results. Meta-analyses that are exclusively based on published literature may therefore produce biased results. Conversely, the inclusion of a study in a trials register can be assumed not to be influenced by its results: registration generally takes place before completion of the study and the criteria that qualify for registration are exclusively based on design features. The studies enlisted in a register are therefore likely to constitute a more representative sample of all the studies that have been performed in a given area than a sample of published studies. John Simes examined this issue for trials of different cancer chemotherapies by comparing the results from meta-analysis of trials identified in a literature search and of trials registered with the International Cancer Research Data Bank.⁵⁸ As shown in Figure 3.3, an analysis restricted to 16 published clinical trials indicates that survival of patients with advanced ovarian cancer is improved with combination chemotherapy as compared to alkylating agent monotherapy (survival ratio 1.16, 95% confidence interval 1.06 to 1.27, $P = 0.004$). However, an analysis of all registered trials (eight published and five unpublished trials) showed only a modest benefit of combination chemotherapy which was not statistically significant (survival ratio 1.06, 95% confidence interval 0.97 to 1.15, $P = 0.17$) (adapted from Simes⁵⁸).

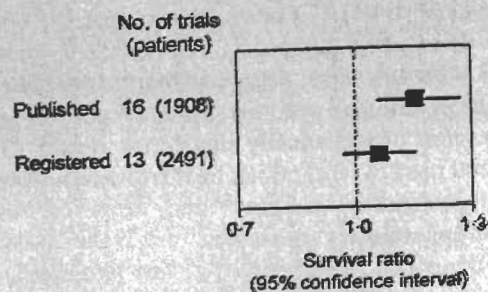


Figure 3.3

their trial at the time.⁶⁰ As discussed in Chapter 24 their findings would have contributed to a more timely detection of the increased mortality that has since become known to be associated with the use of class I anti-arrhythmic agents.^{19,61}

The proportion of all hypotheses tested for which the null hypothesis is truly false is of course unknown and surveys of published results can therefore only provide indirect evidence of publication bias. Convincing, direct evidence is available from the four cohort studies of proposals submitted to ethics committees mentioned earlier,⁴⁵⁻⁴⁷ from cohorts of trials funded by the National Institutes of Health,⁴⁸ trials submitted to licensing authorities,⁶² trials conducted by multicentre trial groups in the domain of human immunodeficiency virus (HIV) infection⁴⁹ and from analyses of trial registries.⁵⁸ In all these studies publication was more likely if effects were large and statistically significant. A meta-analysis of the four ethics committee cohorts is shown in Figure 3.4. In each study possible predictors of publication were examined in multivariate analyses.⁵⁰ The odds of publication were 2.4 times greater if results were statistically significant. Other factors such as the design of the study, its methodological quality, study size and number of study centres, were not consistently associated with the probability of publication.⁵⁰ There was some evidence that the source of funding was associated with publication. We will revisit this finding later in the chapter.

Time lag bias

Studies continued to appear in print many years after approval by the ethics committee. Among proposals submitted to the Royal Prince Alfred Hospital Ethics Committee in Sydney, an estimated 85% of studies with significant results as compared to 65% of studies with null results had been published after 10 years.⁴⁶ The median time to publication was 4.8 years for studies with significant results and 8.0 years for studies with null results. Similarly, trials conducted by multicentre trial groups in the field of HIV infection in the United States appeared on average 4.2 years after the start of patient enrolment if results were statistically significant but took 6.4 years to be published if the results were negative.⁴⁹ As shown in Figure 3.5, trials with positive and negative results differed little in the time they took to complete follow-up. Rather, the time lag was attributable to differences in the time from completion to publication. These findings indicate that *time lag bias*,⁴⁹ may be introduced in systematic reviews even in situations when most or all trials will eventually be published. Trials with positive results will dominate the literature and introduce bias for several years until the negative, but equally important, results finally appear.

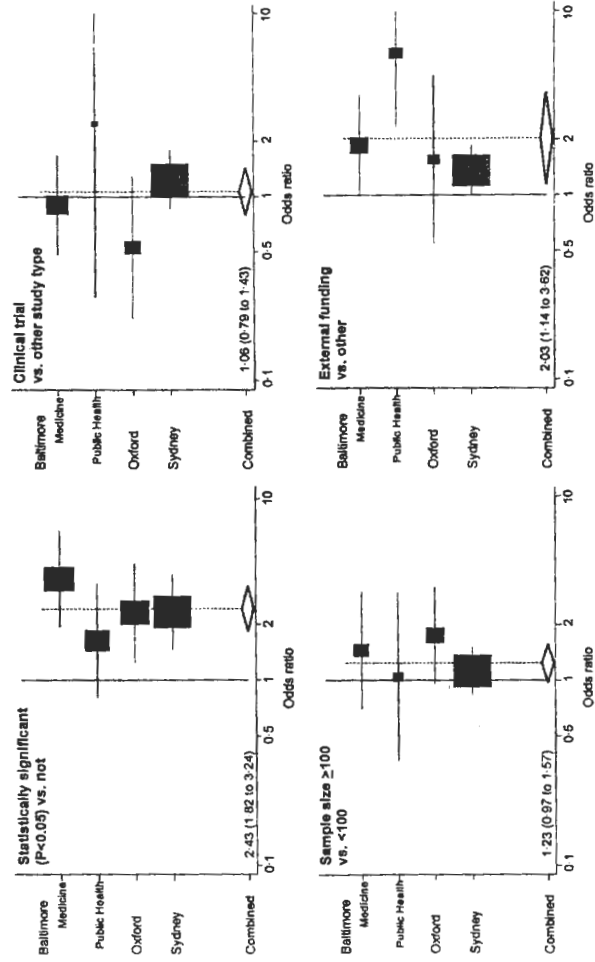


Figure 3.4 Predictors of publication: Meta-analysis of results from four cohort studies of research projects approved by ethics committees or institutional review boards. The analyses of the Johns Hopkins University, Baltimore studies^{47,50} were adjusted for all variables shown, plus number of study groups and number of study centres. The Central Oxford Research Ethics Committee study⁵ was adjusted for all variables shown, plus number of study groups, pilot study and rating on study importance. The Royal Prince Alfred Hospital, Sydney study⁶ was adjusted for all variables shown except sample size, plus study importance for qualitative studies. Meta-analysis was by fixed effects model except for external funding in which case a random effects model was used. Adapted from Dickersin.⁵⁰

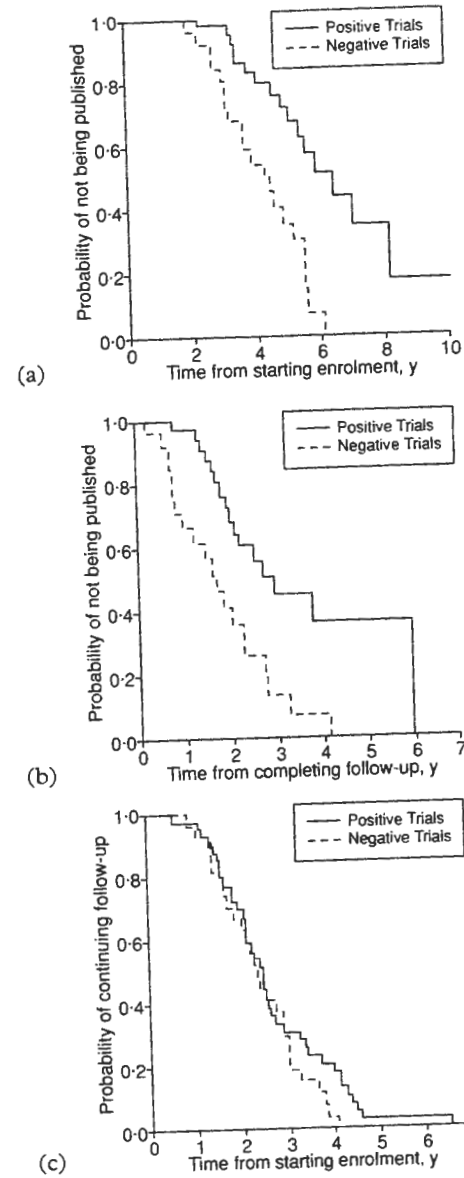


Figure 3.5 Time to publication of 66 clinical trials conducted by multicentre trial groups in HIV infection in the United States: (a) time to publication from start of enrolment, (b) time to publication from completion of follow up, and (c) time to completion from start of enrolment. Reproduced with permission from Ioannidis.⁴⁹

Who is responsible for publication bias: authors, reviewers or editors?

Trials with negative results could remain unpublished because authors fail to write them up and submit to journals, because such trials are reviewed less favourably, or because editors simply don't want to publish negative results. The peer review process is notoriously unreliable and susceptible to subjectivity, bias and conflict of interest.^{63,64} Experimental studies in which test manuscripts were submitted to reviewers or journals showed that reviewers are more likely to referee favourably if results were in accordance with their own views.⁶⁵⁻⁶⁷ For example, when a selected group of authors was asked to review a fictitious paper on transcutaneous electrical nerve stimulation (TENS) they were influenced by their own findings and preconceptions.⁶⁷ A similar study using a fabricated trial of a herbal preparation for the treatment of claudication found that a larger, unselected group of reviewers was not influenced by the direction of the results.⁶⁸ It thus appears that although reviewers may hold strong beliefs, which will influence their assessments, there is no general bias for or against positive findings.

When authors were directly asked why they had not published their findings, the most frequent answer was that they were not interesting enough to merit publication.⁴⁵⁻⁴⁷ Rejection of a manuscript by a journal was rarely mentioned as a reason for not publishing. Selective submission of papers by authors rather than selective recommendation by reviewers and selective acceptance by editors thus appears to be the dominant contributor to publication bias. However, that the latter does occur is illustrated by the "instructions to authors" section of one major diabetes journal, which stated that "mere confirmation of known facts will be accepted only in exceptional cases; the same applies to reports of experiments and observations having no positive outcome".⁶⁹ Such statements have disappeared from guidelines but authors may rightly be reluctant to submit studies with negative results in anticipation of rejection.

The influence of external funding and commercial interests

External funding was associated with publication independently of the statistical significance of the results. However, results were heterogeneous (Figure 3.3) and the effect appears to depend on the source of funding. Funding by government agencies was significantly associated with publication in three cohorts of proposals submitted to ethics committees⁴⁵⁻⁴⁷ whereas pharmaceutical industry sponsored studies were less likely to be published in two studies.^{45,47} Indeed, a large proportion of clinical trials submitted by drug companies to licensing authorities remain unpublished.^{62,70} This is in agreement with a review of publications of clinical trials which separated them into those which were sponsored by the

pharmaceutical industry and those supported by other means.⁷¹ The results of 89% of published industry-supported trials favoured the new therapy, as compared to 61% of the other trials. Similar results have been reported for non-steroidal anti-inflammatory drug trials⁷² and drug studies published in symposium proceedings.⁷³ The implication is that the pharmaceutical industry tends to discourage the publication of negative studies which it has funded. For example, a manuscript reporting on a trial comparing the bioequivalence of generic and brand levothyroxine products, which had failed to produce the results desired by the sponsor of the study, Boots Pharmaceuticals, was withdrawn because Boots took legal action against the university and the investigators. The actions of Boots, recounted in detail by one of the editors of *JAMA*, Drummond Rennie,⁷⁴ meant that publication of the paper⁷⁵ was delayed by about seven years. In a national survey of life-science faculty members in the United States, 20% of faculty members reported that they had experienced delays of more than six months in publication of their work and reasons for not publishing included "to delay the dissemination of undesired results".⁷⁶ Delays in publication were associated with involvement in commercialisation and academic-industry research relationship, as well as with male sex and higher academic rank of the investigator.⁷⁶

Should unpublished data be included in systematic reviews?

Publication bias clearly is a major threat to the validity of any type of review, but particularly of unsystematic, narrative reviews. Obtaining and including data from unpublished trials appears to be the obvious way of avoiding this problem. However, the inclusion of data from unpublished studies can itself introduce bias. The trials that can be located may be an unrepresentative sample of all unpublished studies. Unpublished trials may be of lower methodological quality than published trials: a recent study of 60 meta-analyses that included published and unpublished trials found that unpublished trials were less likely to adequately conceal treatment allocation and blind outcome assessments.⁷⁷ A further problem relates to the willingness of investigators of located unpublished studies to provide data. This may depend upon the findings of the study, more favourable results being provided more readily. This could again bias the findings of a systematic review. Interestingly, when Hetherington *et al.*,⁷⁸ in a massive effort to obtain information about unpublished trials in perinatal medicine, approached 42 000 obstetricians and paediatricians in 18 countries they identified only 18 unpublished trials that had been completed for more than two years.

A questionnaire assessing the attitudes toward inclusion of unpublished data was sent to the authors of 150 meta-analyses and to the editors of the journals which published them.⁷⁹ Support for the use of unpublished material was evident among a clear majority (78%) of meta-analysts.

Journal editors were less convinced – only 47% felt that unpublished data should be included.⁷⁹ The condemnation of the inclusion of unpublished trial data by some editors relates to the issue that the data have not been peer reviewed. It should be kept in mind, however, that the refereeing process has not always been a successful way of ensuring that published results are valid.⁶³ On the other hand, meta-analyses of unpublished data from interested sources is clearly of concern. Such unchallengeable data have been produced in circumstances in which an obvious financial interest exists, as discussed in Box 3.4.

Box 3.4 The controversy over selective serotonin-reuptake inhibitors and depression

Selective serotonin-reuptake inhibitors (SSRIs) are widely used for the treatment of depression, although their clinical advantages over the much less expensive tricyclic antidepressants have not been well established. In their meta-analysis Song *et al.*,⁸⁰ used the drop-out rate among randomised controlled trial participants on SSRIs and those on conventional antidepressants as an indicator of therapeutic success: patients who stop taking their medication because of inefficacy or side-effects are the ones who are not benefiting, and thus the class of drug with the lower drop-out rate can be considered the one with the more favourable effects. There was little difference between SSRIs and the other, usually tricyclic, antidepressants. In response to this analysis, Lilly Industries, the manufacturers of the SSRI fluoxetine, presented a meta-analysis of 14 investigational new drug studies which they stated included every study completed by December 1990.⁸¹ This included what were called (in the usual industry terminology) “unpublished data on file”. As shown in Table 3.4, the pooled drop out rates calculated by Lilly Industries differed markedly from the literature-based analysis. Lilly Industries claimed that their analysis was not “subject to biases introduced by selective publication and literature searches” but this is difficult to assess if the trials included represent unpublished “data on file”.

Table 3.4

	Trials (n)	Fluoxetine		Tricyclic antidepressant		P
		Patients (n)	Drop-out rate (%)	Patients (n)	Drop-out rate (%)	
Song <i>et al.</i> ⁸⁰	18*	913	34.5	916	36.7	0.40
Lilly Industries ⁸¹	14	781	36.5	788	47.5	<0.0001

* References 6,12-15,18,29,31,33-35,44,47,63,65-67,69 in Song *et al.*⁸⁰

Other reporting biases

While publication bias has long been recognised⁵³ and much discussed, other factors can contribute to biased inclusion of studies in meta-analyses. Indeed, among published studies, the probability of identifying relevant trials for meta-analysis is also influenced by their results. These biases have received much less consideration than publication bias, but their consequences could be of equal importance.

Duplicate (multiple) publication bias

In 1987, Gøtzsche⁸² found that among 244 reports of trials comparing non-steroidal anti-inflammatory drugs in rheumatoid arthritis 44 (18%) were redundant, multiple publications, which overlapped substantially with an already published article.⁸³ Twenty trials were published twice, 10 trials three times and one trial times.⁸⁴ The production of multiple publications from single studies can lead to bias in a number of ways.⁸⁵ Most importantly, studies with significant results are more likely to lead to multiple publications and presentations,⁴⁵ which makes it more likely that they will be located and included in a meta-analysis. The inclusion of duplicated data may therefore lead to overestimation of treatment effects, as recently demonstrated for trials of the efficacy of ondansetron to prevent post-operative nausea and vomiting⁸⁶ (Figure 3.6). It is not always obvious that

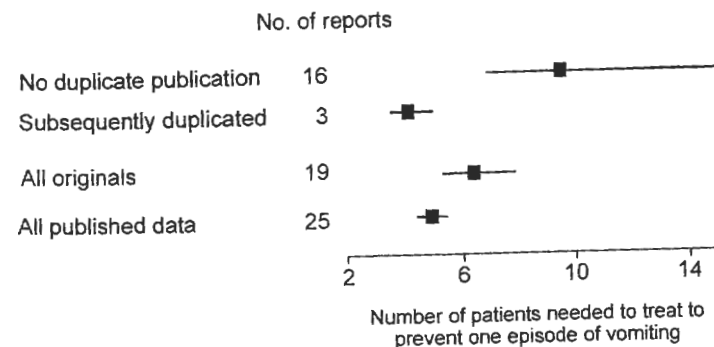


Figure 3.6 The inclusion of duplicated data may lead to overestimation of treatment effects: Tramèr *et al.*⁸⁶ found that of 19 trials that compared prophylactic intravenous ondansetron, data from three large multicentre trials had been duplicated in six further reports. In the 16 reports which were not duplicated, the number needed to treat (NNT) to prevent one episode of vomiting was 9.5 compared to 3.9 in the three reports that were subject to duplication. When all 19 original reports were combined, the NNT was 6.4; when original and duplicate reports were combined, a biased NNT of 4.9 was obtained.

multiple publications come from a single study, and one set of study participants may thus be included in an analysis twice. Huston and Moher⁸⁵ and, more recently, Johansen and Gøtzsche⁸⁷ vividly described the difficulties and frustration caused by redundancy and the “disaggregation” of medical research when results from a multicentre trial are presented in several publications. Indeed, it may be extremely difficult – if not impossible – for reviewers to determine whether two papers represent duplicate publications of one trial or two separate trials, since examples exist where two articles reporting the same trial do not share a single common author.^{84,86}

Citation bias

The perusal of the reference lists of articles is widely used to identify additional articles that may be relevant. The problem with this approach is that the act of citing previous work is far from objective and retrieving literature by scanning reference lists may thus produce a biased sample of studies. There are many possible motivations for citing an article, ranging from decoration to showing up-to-dateness and knowledge. Brooks⁸⁸ interviewed academic authors from various faculties at the University of Iowa and asked for the reasons for citing each reference in one of the authors' recent articles. Persuasiveness, the desire to convince peers and substantiate their own point of view emerged as the most important reason for citing articles. Brooks concluded that authors advocate their own opinions and use the literature to justify their point of view: “Authors can be pictured as intellectual partisans of their own opinions, scouring the literature for justification”.⁸⁸ In Gøtzsche's analysis⁸² of trials of non-steroidal anti-inflammatory drugs in rheumatoid arthritis, trials demonstrating a superior effect of the new drug were more likely to be cited than trials with negative results. Similarly, trials of cholesterol lowering to prevent coronary heart disease were cited almost six times more often if they were supportive of cholesterol lowering (see also Box 3.5).⁸⁹ Overcitation of unsupportive studies can also occur. Hutchinson *et al.*⁹⁰ examined reviews of the effectiveness of pneumococcal vaccines and found that unsupportive trials were more likely to be cited than trials showing that vaccines worked.

Language bias

Reviews are often exclusively based on trials published in English. For example, among 36 meta-analyses reported in leading English-language general medicine journals from 1991 to 1993, 26 (72%) had restricted their search to studies reported in English.⁹¹ Investigators working in a non-English speaking country will, however, publish some of their work in local journals.⁹² It is conceivable that authors are more likely to report in an

Box 3.5 Cholesterol lowering after myocardial infarction: citation bias and biased inclusion criteria

A meta-analysis of trials of cholesterol-lowering after myocardial infarction⁹³ defined its inclusion criteria as those single-factor randomised trials with at least 100 participants per group, with at least three years of follow up and without the use of hormone treatment to lower cholesterol. Seven trials – one with two treatment arms – were included in this analysis. The pooled odds ratio for all-cause mortality for these trials was reported as 0.91 (95% confidence interval 0.82 to 1.02), indicating a favourable trend. One trial that met all the entry criteria was, however, not included.⁹⁴ In this study, the odds ratio for overall mortality was an unfavourable 1.60 (0.95 to 2.70). For the seven trials included in the analysis the mean annual citation count per study for the period up to five years after publication was 20; for the study which was not included it was less than one.⁹⁵ It is likely that the latter was missed precisely because it was infrequently quoted. The inclusion criteria relating to study size and length of follow up were somewhat arbitrary: the results of at least 11 other randomised secondary prevention trials of cholesterol lowering were available at the time this analysis was published (references 2a, 3a, 4a, 5a, 9a, 10a, 12a, 13a, 15a, 22a, 35a in Davey Smith *et al.*⁹³). The pooled odds ratio for all-cause mortality for these trials is 1.14 (1.03 to 1.26). As shown below, the selective identification of the much-quoted supportive studies, citation bias, and biased inclusion criteria may have distorted the results of this meta-analysis.

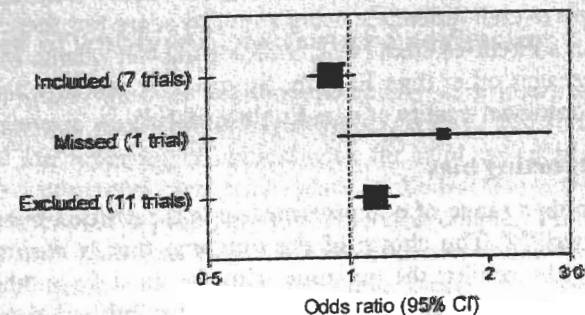


Figure 3.7 (adapted from Rossouw *et al.*⁹³).

international, English-language journal if results are positive whereas negative findings are published in a local journal. This has recently been demonstrated for the German language literature.⁹⁶ When comparing pairs of articles published by the same first author, 63% of trials published in English had produced significant ($P < 0.05$) results as compared to 35% of

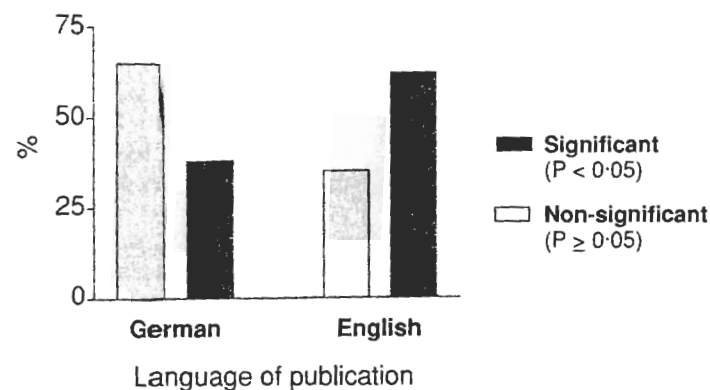


Figure 3.8 Language bias. The proportion of controlled trials with statistically significant results was higher among reports published in English. Analysis based on 40 pairs of trials published by the same author, with one trial published in English and the other in German. Adapted from Egger *et al.*⁹⁶

trials published in German (Figure 3.8). Bias could thus be introduced in meta-analyses exclusively based on English-language reports.^{91,97} On the other hand, as with unpublished trials, the lower quality of trials published in languages other than English may in fact introduce bias. Moher *et al.* compared the quality of 133 randomised controlled trials published in English with 96 trials published in French, German, Italian or Spanish.⁹⁷ They found no overall difference using a quality score but there were some differences on an item-by-item basis, indicating lower quality of trials published in languages other than English. Sterne *et al.* recently also reported lower methodological quality of non-English trials.⁷⁷

Outcome reporting bias

In many trials a range of outcome measures is recorded but not all are always reported.^{98,99} The choice of the outcome that is reported can be influenced by the results: the outcome with the most favourable findings will generally be reported. An example of how published results can be misleading comes from two separate analyses^{100,101} of a double-blind placebo-controlled trial assessing the efficacy of amoxicillin in children with non-suppurative otitis media. Opposite conclusions were reached, mainly because different weight was given to the various outcome measures that were assessed in the study. This disagreement was conducted in the public arena, since it was accompanied by accusations of impropriety against the team producing the findings favourable to amoxicillin. The leader of this team had received large monetary sums, both in research grants and as personal honcraia, from the manufacturers of amoxicillin.¹⁰² It is a good

example of how reliance upon the data chosen to be presented by the investigators can lead to distortion.¹⁰³ Reporting bias may be particularly important for adverse effects. Hemminki⁷⁰ examined reports of clinical trials submitted by drug companies to licensing authorities in Finland and Sweden and found that unpublished trials gave information on adverse effects more often than published trials.

Biased inclusion criteria

Once studies have been located and data obtained, there is still potential for bias in setting the inclusion criteria for a meta-analysis. If, as is usual, the inclusion criteria are developed by an investigator familiar with the area under study, they can be influenced by knowledge of the results of the set of potential studies. Manipulating the inclusion criteria could lead to selective inclusion of positive studies and exclusion of negative studies. For example, some meta-analyses of trials of cholesterol-lowering therapy^{104,105} have excluded certain studies on the grounds that the treatments used appear to have had an adverse effect that was independent of cholesterol lowering itself. These meta-analyses have, however, included trials of treatments that are likely to favourably influence risk of coronary heart disease, independent of cholesterol lowering. Clearly such an asymmetrical approach introduces the possibility of selection bias, with the criteria for inclusion into the meta-analysis being derived from the results of the studies (see Box 3.5).

The future of unbiased, systematic reviewing

Reporting biases and the inadequate quality of primary research are potentially serious problems for systematic reviews. We need both fair conduct and fair reporting of clinical trials for valid systematic reviews.¹⁰⁶ Important developments have taken place in the last few years which will eventually overcome the problems outlined in this chapter. Firstly, a variety of graphical and statistical methods have been developed for evaluating whether publication and related reporting biases are operating. For example, Hutton and Williamson¹⁰⁷ recently proposed sensitivity analyses to examine the potential impact of biased selection of outcomes. Their approach is similar to the selection models advocated by Copas¹⁰⁸ to examine publication bias which, along with other methods, is discussed in Chapter 11. Secondly, and more importantly, reporting biases are more likely to be prevented nowadays. Only a few years ago searching electronic databases such as MEDLINE or EMBASE was unreliable and reviewers were likely to miss substantial proportions of relevant trials. Indeed, in 1995 only 19 000 reports were readily identifiable as randomised controlled trials in the widely used MEDLINE database, although many more trials were

included in that database. As discussed in Chapter 4 this situation is now much improved. The regularly updated Cochrane Controlled Trials Register contains over a quarter of a million of reports of controlled trials and is clearly the best single source of published trials for inclusion in systematic reviews and meta-analyses. The identification of ongoing and unpublished studies has also become more practicable. National research registers, a "Register of Registers", and a "metaRegister" have been set up (see Chapters 4 and 24). One of these registers, Current Controlled Trials (www.controlled-trials.com) will publish trial protocols and full reports of controlled trials. There have also been important initiatives from within the pharmaceutical industry to improve access to information on trials, and to prevent duplicate or selective publication.^{109,110} Finally, an "amnesty" for unpublished trials was launched by journal editors.¹¹¹

These initiatives mean that the identification of published and unpublished trials for systematic reviews has become an easier task, but what about the quality of the trials that are identified? There is growing consensus that the methodological quality should routinely be assessed but this is hampered by the quality of reporting, which is often inadequate. With the adoption of the CONSORT (Consolidated Standards of Reporting Trials) guidelines (see <http://www.consort-statement.org>) by an increasing number of journals this situation is now also improving.^{112,113} Considerable progress has thus been made in a short time. This is not to say, however, that in the future unbiased, systematic reviewing will always produce conclusive answers. Many systematic reviews will continue to be based on a small number of trials of doubtful quality. These will have to be inconclusive, even if meta-analysis indicates a statistically significant effect of the intervention. Clearly demonstrating the inadequacy of existing evidence is an important objective of systematic reviews, which should serve as a stimulus for conducting the appropriate and necessary trials.

Acknowledgements

We are grateful to Fujian Song and Sally Hopewell for their help with the literature search, and to Iain Chalmers and Nicola Low for helpful comments on an earlier draft of this chapter.

- 1 Eysenck HJ. An exercise in mega-silliness. *Am Psychol* 1978;33:517.
- 2 Chalmers I, Enkin M, Keirse M. *Effective care during pregnancy and childbirth*. Oxford: Oxford University Press, 1989.
- 3 Mann C. Meta-analysis in the breech. *Science* 1990;249:476-80.
- 4 Oakes M. *Statistical inference: a commentary for the social and behavioural sciences*. Chichester: John Wiley, 1986.
- 5 Boden WE. Meta-analysis in clinical trials reporting: has a tool become a weapon? *Am J Cardiol* 1992;69:681-6.

- 6 Bailar JC. The promise and problems of meta-analysis. *N Engl J Med* 1997;337:559-61.
- 7 Mann C. Richard Peto: statistician with a mission. *Science* 1990;249:479.
- 8 Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990;263:1392-5.
- 9 Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
- 10 LeLorier J, Grégoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-42.
- 11 Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Ass J* 1997;156:1411-6.
- 12 Leizorovicz A, Haugh MC, Chapuis F-R, et al. Low molecular weight heparin in prevention of perioperative thrombosis. *BMJ* 1992;305:913-20.
- 13 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;i:1088-92.
- 14 Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). 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.
- 15 Teo KK, Yusuf S. Role of magnesium in reducing mortality in acute myocardial infarction. A review of the evidence. *Drugs* 1993;46:347-59.
- 16 ISIS-4 Collaborative Group. ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995;345:669-87.
- 17 Stuck AE, Siu AL, Wieland GD, et al. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 1993;342:1032-6.
- 18 Reuben DB, Borok GM, Wolde-Tsadik G, et al. Randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. *N Engl J Med* 1995;332:1345-50.
- 19 The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;321:406-12.
- 20 Imperiale TF, Stollenwerk Petrullis A. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. *JAMA* 1991;266:261-5.
- 21 Nurmohamed MT, Rosendaal FR, Bueller HR, et al. Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152-6.
- 22 Yusuf S. Calcium antagonists in coronary heart disease and hypertension. Time for reevaluation? *Circulation* 1995;92:1079-82.
- 23 Dunnigan MG. The problem with cholesterol. No light at the end of this tunnel? *BMJ* 1993;306:1355-6.
- 24 de Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;355:80-1.
- 25 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
- 26 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
- 27 Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
- 28 Jüni P, Tallon D, Egger M. "Garbage in - garbage out?" Assessment of the quality of controlled trials in meta-analyses published in leading journals. *3rd Symposium on Systematic Reviews: beyond the basics*. Oxford, 3-5 July 2000.
- 29 Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trial for meta-analysis. *JAMA* 1999;282:1054-60.
- 30 Lensing AW, Hirsh J. ¹²⁵I-fibrinogen leg scanning: reassessment of its role for the diagnosis of venous thrombosis in post-operative patients. *Thromb Haemost* 1993;69:2-7.
- 31 Smith R. What is publication? A continuum. *BMJ* 1999;318:142.
- 32 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-9.

- 33 Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;272:158-62.
- 34 Dudley HA. Surgical research: master or servant. *Am J Surg* 1978;135:458-60.
- 35 Goldman L, Loscalzo A. Fate of cardiology research originally published in abstract form. *N Engl J Med* 1980;303:255-9.
- 36 Meranze J, Ellison N, Greenhow DE. Publications resulting from anesthesia meeting abstracts. *Anesth Analg* 1982;61:445-8.
- 37 McCormick MC, Holmes JH. Publication of research presented at the pediatric meetings. Change in selection. *Am J Dis Child* 1985;139:122-6.
- 38 Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. *JAMA* 1990;263:1401-5.
- 39 Juzych MS, Shin DH, Coffey JB, Parrow KA, Tsai CS, Briggs KS. Pattern of publication of ophthalmic abstracts in peer-reviewed journals. *Ophthalmology* 1991;98:553-6.
- 40 De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Ann Oncol* 1992;3:187-91.
- 41 Juzych MS, Shin DH, Coffey J, Juzych L, Shin D. Whatever happened to abstracts from different sections of the association for research in vision and ophthalmology? *Invest Ophthalmol Vis Sci* 1993;34:1879-82.
- 42 Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. *Can J Anaesth* 1993;40:632-4.
- 43 Dirk L. Incidence of acceptance, rejection, re-submission, and nonpublication of full reports of research presented at academic meetings. *Second International Conference on Peer Review*. Chicago, IL, 10 September 1993.
- 44 Donaldson LJ, Cresswell PA. Dissemination of the work of public health medicine trainees in peer-reviewed publications: an unfulfilled potential. *Public Health* 1996;110:61-3.
- 45 Easterbrook PJ, Berlin J, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- 46 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640-5.
- 47 Dickersin K, Min YL, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374-8.
- 48 Dickersin K, Min YL. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993;2. Doc No 50:4967.
- 49 Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281-6.
- 50 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9:15-21.
- 51 Smith ML. Publication bias and meta-analysis. *Evaluation Education* 1980;4:22-4.
- 52 Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychological Bulletin* 1979;86:638-41.
- 53 Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *J Am Stat Assoc* 1959;54:30-4.
- 54 Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;49:108-12.
- 55 Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med* 1994; 1:267-71.
- 56 Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19:159-66.
- 57 Pittler MH, Abbot NC, Harkness EF, Ernst E. Location bias in controlled clinical trials of complementary/alternative therapies. *J Clin Epidemiol* 2000;53:485-9.
- 58 Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11-29.
- 59 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-41.
- 60 Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int J Cardiol* 1993;40:161-6.
- 61 Teo KK, Yusuf S, Furberg CD. Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction. An overview of results from randomized controlled trials. *JAMA* 1993;270:1589-95.
- 62 Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol* 1998;46:147-50.
- 63 Godlee F, Dickersin K. Bias, subjectivity, chance, and conflict of interest in editorial decisions. In Godlee F, Jefferson T, eds. *Peer review in health sciences*, pp. 57-78. London: BMJ Books, 1999.
- 64 Peters DP, Ceci SJ. Peer review practices of psychology journals: the fate of published articles, submitted again. *Behav Brain Sci* 1982;5:187-255.
- 65 Mahoney MJ. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* 1977;1:161-75.
- 66 Epstein WM. Confirmation bias among social work journals. *Sci Tech Hum Values* 1990;15:9-37.
- 67 Ernst E, Resch KL. Reviewer bias: a blinded experimental study. *J Lab Clin Med* 1994;124:178-82.
- 68 Abbot NC, Ernst E. Publication bias: direction of outcome less important than scientific quality. *Perfusion* 1998;11:182-4.
- 69 Anon. Manuscript guideline. *Diabetologia* 1984;25:4A.
- 70 Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;280:833-6.
- 71 Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;1:155-8.
- 72 Rochon PA, Gurwitz JH, Simms RW, et al. A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Intern Med* 1994;154:157-63.
- 73 Cho M, Bero L. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996;124:485-9.
- 74 Rennie D. Thyroid storms. *JAMA* 1997;277:1238-43.
- 75 Dong BJ, Hauck WW, Gambertoglio JG, et al. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JAMA* 1997;277:1205-13.
- 76 Blumenthal D, Campbell EG, Anderson MS, et al. Withholding research results in academic life science. Evidence from a national survey of faculty. *JAMA* 1997;277:1224-8.
- 77 Sterne JAC, Bartlett C, Juni P, Egger M. Do we need comprehensive literature searches? A study of publication and language bias in meta-analyses of controlled trials. *3rd Symposium on Systematic Reviews: beyond the basics*. Oxford, 3-5 July 2000.
- 78 Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;84:374-80.
- 79 Cook DJ, Guyatt GH, Ryan G, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993;269:2749-53.
- 80 Song F, Freemantle N, Sheldon TA, et al. Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *BMJ* 1993;306:683-7.
- 81 Nakielnny J. Effective and acceptable treatment for depression. *BMJ* 1993;306:1125.
- 82 Gotzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
- 83 Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. *JAMA* 1997;277:927-34.
- 84 Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;36:429-32.
- 85 Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;347:1024-6.
- 86 Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635-40.
- 87 Johansen HK, Gotzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *JAMA* 1999;282:1752-9.

- Brooks TA. Private acts and public objects: an investigation of citer motivations. *J Am Soc Information Sci* 1985;36:223-9.
- Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;305:15-19.
- Hutchison BG, Oxman AD, Lloyd S. Comprehensiveness and bias in reporting clinical trials. *Can Family Physician* 1995;41:1356-60.
- Grégoire G, Derderian F, LeLorier 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.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286-91.
- Rossouw JE, Lewis B, Rifkind BM. The value of lowering cholesterol after myocardial infarction. *N Engl J Med* 1990;323:1112-19.
- Woodhill JM, Palmer AJ, Leelarthaepin B, McGilchrist C, Blacket RB. Low fat, low cholesterol diet in secondary prevention of coronary heart disease. *Adv Exp Med Biol* 1978;109:317-30.
- Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ* 1993;306:1367-73.
- Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326-9.
- Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.
- Pocock S, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;317:426-32.
- Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996;88:206-7.
- Mandel EH, Rockette HE, Bluestone CD, et al. Efficacy of amoxicillin with and without decongestant-antihistamine for otitis media with effusion in children. *N Engl J Med* 1987;316:432-7.
- Cantekin EI, McGuire TW, Griffith TL. Antimicrobial therapy for otitis media with effusion ("secretory" otitis media). *JAMA* 1991;266:3309-17.
- Rennie D. The Cantekin affair. *JAMA* 1991;266:3333-7.
- Anon. Subjectivity in data analysis. *Lancet* 1991;337:401-2.
- Peto R, Yusuf S, Collins R. Cholesterol-lowering trial results in their epidemiologic context. *Circulation* 1985;72 (suppl 3):451.
- MacMahon S. Lowering cholesterol: effects on trauma death, cancer death and total mortality. *Aust NZ J Med* 1992;22:580-2.
- Rennie D. Fair conduct and fair reporting of clinical trials. *JAMA* 1999;282:1766-8.
- Hutton JL, Williamson RR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-70.
- Copas J. What works?: selectivity models and meta-analysis. *J Roy Stat Soc (Series A)* 1999;162:95-109.
- Sykes R. Being a modern pharmaceutical company: involves making information available on clinical trial programmes. *BMJ* 1998;317:1172.
- Wager E, Tumas JA, Field EA, Glazer NB, Schulz G, Grossman L. Improving the conduct and reporting of clinical trials. *JAMA* 2000;283:2788-9.
- Smith R, Roberts I. An amnesty for unpublished trials. *BMJ* 1997;315:612.
- Moher D, Jones A, LePage L, for the CONSORT Group. Does the CONSORT statement improve the quality of reports of randomized trials: a comparative before and after evaluation? *JAMA* (in press).
- Egger M, Jüni P, Bartlett C, for the CONSORT Group. The value of flow diagrams in reports of randomized controlled trials: bibliographic study. *JAMA* (in press).