ELSEVIER

Journal of Clinical Epidemiology 53 (2000) 207–216

Journal of Clinical Epidemiology

Publication bias in meta-analysis: its causes and consequences

Alison Thornton, Peter Lee*

P.N. Lee Statistics and Computing Ltd, 17 Cedar Road, Sutton, Surrey, SM2 5DA, England

Received 16 March 1998; received in revised form 21 June 1999; accepted 26 July 1999

Abstract

Publication bias is a widespread problem that may seriously distort attempts to estimate the effect under investigation. The literature is reviewed to determine features of the design and execution of both single studies and meta-analyses leading to publication bias, and the role the author, journal editor, and reviewer play in selecting studies for publication. Methods of detecting, correcting for, and preventing publication bias are reviewed. The design of the meta-analysis itself, and the studies included in it, are shown to be important among a number of sources of publication bias. Various factors influence an author's decision to submit results for publication. Journal editors and reviewers are crucial in deciding which studies to publish. Various methods proposed for detecting and correcting for publication bias, though useful, all have limitations. However, prevention of publication bias by registering every trial undertaken or publishing all studies is an ideal that is hard to achieve.

Keywords: Publication bias

1. Introduction

The "dictionary of epidemiology" [1] defines publication bias as "an editorial predilection for publishing particular findings, e.g., positive results, which leads to the failure of authors to submit negative findings for publication." Rosenthal, in his "file drawer problem," described an extreme view in which journals are filled with the 5% of studies showing a false-positive result, the other 95%, showing nonsignificant results (at $P < 0.05$), being left to fill file drawers [2]. Awareness of publication bias began in 1956 when the editor of the *Journal of Abnormal Social Psychology* indicated that negative studies were less likely to be published in his journal [3]. In 1959, it was found that very few negative results were reported in four psychological journals, a finding regarded as strongly suggesting publication bias [4]. However, no attempt was made to quantify the problem until 1964 [5]. The existence of publication bias is now widely accepted. Attempts to summarize evidence relating to a specific hypothesis, whether by narrative review or meta-analysis, can be seriously distorted by publication bias. For example, one recent analysis estimated that 45% of an observed association could be due to publication bias [6].

This article aims to explore publication bias and issues related to it, and the effect it may have on attempts to review evidence relating to various hypotheses. Features of the design and execution of both single studies and meta-analyses that may lead to publication bias are examined, along with factors that may influence the author's decision to submit his results for publication. The role of journal editors and reviewers in deciding which studies to publish is also considered. Methods aimed at confirming the existence of, correcting for, and preventing publication bias are reviewed. It is shown that one can estimate the extent to which such a bias may have occurred, and even correct for it, helping authors of future reviews not only to be fully aware of the problem, but also to take steps to minimize it.

2. Publication bias arising from the design or execution of single studies

Several facets of the design or execution of a study, including sample size and the method of reporting the data, may lead to publication bias. The investigator's own beliefs and expectations may also influence the outcome. A small sample size leads to lack of power [7], and significance may then only be obtained if chance exaggerates any true differences between the groups under study [8]. Though the obvious likely effect of inadequate sample size is failure to demonstrate statistical significance for a clinically important effect [8], it will also lead to publication bias if results from

* Corresponding author. Tel.: +44 181 642 8265; fax: +44 181 642 2135.

0895-4356/00/$ – see front matter
PII: S0895-4356(99)00161-4

small studies are very unlikely to be published unless significant. If there is no preset trial size results may be presented at an arbitrary time, with the magnitude of the treatment difference possibly affecting the decision to report [9]. Studies that have been stopped early tend to decline in significance with further follow-up, even without the recruitment of additional subjects [10], so that delaying publication may lead to a smaller estimate of treatment effect. Studies selected for publication not only contain exaggerated estimates of the main effects under study but also usually underestimate variance, these biases operating more strongly the more inadequate the sample size [8].

Unlike small studies, large studies have the potential to answer hundreds of questions. An author may be strongly tempted to dredge through the data from an essentially negative study to find positive results and publish only those [7,11,12]. The probability of finding a significant result increases markedly with the number of end points or subgroups analyzed [9,13]. Multitreatment trials and repeating measurements over time also increase the chances of finding a false-positive effect of treatment [9]. However, most studies are constructed with a stated primary focus, other factors being collected either to permit adjustment for potential confounding or for hypothesis-generation purposes. Thus, the results pertaining to the primary hypotheses are normally given greater credibility than unexpected or casual observations from the remaining data [14].

Errors in the recording of data may also lead to bias. Although differences between the actual study results and the results recorded in subsequent publications are rare, most favor the observer's hypothesis, thus suggesting the mistakes may not be random [11]. The usual effect is to confer significance on data that in reality are just below the level of statistical significance. Some researchers may go to more extreme lengths and deliberately tamper with their data. Such fraud, although extremely rare, may not be detected even after peer review and replication [15].

Finally, there is evidence that, for some experimenters, their preconceptions of what their results should look like may influence the data they obtain [16,17]. By analogy, it also seems possible that the decision to publish may be affected by preconceptions on the part of the reviewer.

3. Publication bias arising from the researcher deciding whether or not to submit results

An early study found that dissertations and theses were three or four times more likely to be published if they were positive than if they were negative [5]. Such findings may be more because researchers decide not to submit their findings than because journal editors reject their papers [7,11,18–21], statistically significantly positive studies being up to 10 times more likely to be submitted for publication [13,22]. The main reasons given for nonsubmission of studies are the negative results themselves and lack of interest by the researcher [7,19,20]. There is also an assumption that editors and reviewers are biased against negative studies, considering them to be of lesser interest [7].

Researchers may also not submit their work because they are aware of serious limitations in it [18,20], and indeed the wisdom of publishing such studies is questionable. Such flaws may occur more often in negative studies than in positive ones [7]. However, one investigator found that the subsequent full publication of abstracts showed no association with either study quality or results, though researchers who have had two or more studies published were likely to submit further papers [19]. Study size may also be important, as researchers are often keen to publish large studies regardless of outcome, due to the effort involved in conducting them [23,24].

4. Publication bias arising from the tendency of journals to reject negative studies

Some editors and reviewers strongly dislike negative studies [7,8,20,25]. The *British Medical Journal* states that "negative results have never made rivetting reading." Their ideal article is one that affects clinical practice, improves prognosis, or simplifies management [19]. While some negative reports may legitimately be rejected due to poor quality [3,19], even negative studies that appear to be better conducted than positive ones may be much less likely to be accepted for publication [25]. Negative studies sometimes deal with implausible hypotheses, leading to rejection however well-conducted the study [7]. Reviewers may block or delay the publication of work comparable with their own, they may be biased against work in other fields, or the article may only be appreciated by a worker in the same field [26]. In addition, negative replications of previously positive findings are particularly unpopular among journal editors [5,27], but where they are published it may be appropriate to scrutinize the methods of such studies more carefully [3].

Negative studies may be subjected to more critical scrutiny than papers reporting positive results [5,27]. In one study [28], reviewers were asked to examine manuscripts falling into one of five groups: three with no discussion but positive, negative, and no results, respectively, and two with mixed results and a positive or negative discussion. Ratings of topic relevance did not differ across the groups, but despite having identical experimental procedures manuscripts with positive results were rated as methodologically better than those reporting negative results. Data presentation was also rated more highly in positive manuscripts, while the discussion section did not appear to influence the reviewers' evaluation of manuscripts with mixed results. A similar trend was seen for evaluation of scientific contribution. Most significantly, positive manuscripts were usually recommended for publication with only minor revisions, while the reviewers recommended rejection or major revisions of negative reports. Manuscripts with mixed results were consistently rejected. Furthermore, a contradiction in the methods section was noted by only 25% of reviewers of positive

papers, compared to over 70% of those reading negative reports. This study clearly suggests that peer review is not an unbiased procedure.

5. Sponsorship

A study's source of funding may also unduly influence the probability of subsequent publication of the results. For instance, studies showing no association between exposure and disease may be published by groups with a presumed special interest in demonstrating a lack of causation,such as the companies that introduced the risk factor [13,29]. Similarly, reports submitted to governments by Scandinavian pharmaceutical companies showed a lower proportion of published than unpublished studies providing evidence of adverse drug effects [13]. Furthermore, pharmaceutical companies may discourage the publication of studies that show null effects of their drugs [11,29]. Up to 89% of studies sponsored by the pharmaceutical industry favor a new therapy, compared to only 61% of trials funded from other sources [13]. Additionally, 72% of articles in tobacco industry-sponsored symposia agree with the viewpoint of tobacco companies, compared to 41% of articles in nonindustry-sponsored symposia and only 20% for journal articles [30]. In this respect, it should be borne in mind that there is currently considerable pressure from antismoking organizations for journals not to accept papers supporting any aspect of the tobacco industry's position on smoking and health, regardless of the scientific merits of the paper, presumably so that the message to smokers to give up will come over as clearly as possible.

Biasing factors in relation to sponsorship would therefore seem to operate in both directions.

6. Bias arising from the design and execution of reviews and meta-analyses

There are likely to be unpublished studies relevant to any given hypothesis. As published studies may systematically differ from unpublished ones [31,32], reviews or meta-analyses based only on published data may reach misleading conclusions [33]. It is widely thought, therefore, that as many studies as possible should be included, both published and unpublished [22,31,34–36].

However, there are some problems with this simple view. Firstly, it should be noted that it is often impossible to obtain details of every relevant study [12,33] and expending great effort in acquiring unpublished data may be of limited use if a complete sample still cannot be obtained [37]. Furthermore, while published studies have the advantage of full data analysis and have been subjected to peer review (imperfect as it may be, as discussed above), unpublished data have not [7,24,32,34,36]. Results from unpublished studies may be less reliable, as such data may be more subject to fraud or distortion [37], and, even if honest, the studies may not have been conducted with the same rigor [36,37]. Studies with obvious flaws should not be included in a meta-analysis [22]. To do so might skew the results at least as much as does publication bias [7]. As discussed earlier though, publication does not guarantee study quality, aspects of which should be taken into account when conducting meta-analyses [34].

Some authors choose to exclude particular studies, arguing that the results of a meta-analysis are meaningful only if the same protocol has been followed in all of the studies to be combined, and if each study has allocated treatment randomly to its subjects. According to this view, meta-analysis should only be carried out on randomized clinical trials [38,39], as meta-analysis of observational studies may provide a biased estimate of the association under review, even in the absence of publication bias [38,40]. Additionally, although the quality of a randomized clinical trial can be assessed against a quality scoring system [32], it is less easy to do this for observational studies, where there is much more variation in study design [41]. Studies that are not randomized or controlled not only have more potential for bias [41,42], but tend to show greater treatment effects and greater heterogeneity [42]. However, many epidemiological studies are observational in nature, and excluding them may seriously bias the results of a meta-analysis, or prevent it being performed at all.

Studies may also be excluded from meta-analyses for less appropriate reasons. For example, nearly 80% of reviews have language restrictions [43], and only include papers published in English [43–45]. Reasons for this may include difficulties in identifying relevant papers published in other languages, or the presumed greater importance and quality of English-language publications. Reanalysis of some meta-analyses originally based only on English-language papers have produced different results when relevant papers published in other languages were included. While three meta-analyses showed a difference in results but no change to the actual level of significance, a fourth, the original results of which failed to reach statistical significance, would have arrived at a different conclusion had a German-language paper been included [45]. However, certain countries only publish positive results [46] so that including papers in all languages may actually introduce more bias into a meta-analysis.

Authors must be careful to avoid the multiple inclusion of studies from which more than one publication has arisen. Nearly 20% of articles identified for review purposes may be repeat publications of studies already reported, some trials giving rise to as many as five papers [47,48]. Perhaps unsurprisingly, positive studies are particularly prone to multiple reporting [44,48]. Difficulty in identifying duplicate publications arises due to variations in the title, the name of the first author, and the number of authors. A failure to indicate the institution to which the authors were affiliated and to refer to previously published articles may also lead to confusion [11,47,48]. Undetected multiple publication in a meta-analysis may lead to overweighting by the

duplicated study or studies [34,47,48], and can overestimate a treatment's efficacy by nearly 25% [48].

Two models commonly used for the combining of data are the fixed-effects and random-effects models. The choice of model may influence the conclusions of a meta-analysis. Each method calculates a weighted average of the estimates from the original studies [14,41,49], with the assumptions underlying the method determining the weights to be used [49]. The fixed-effects model assumes homogeneity of effects across the studies being combined [14,32,49–51], and models the results from individual studies by

$$Y_j = \Delta + \epsilon_j$$

where $\epsilon_j \sim N(0, \sigma_j^2)$, so that Δ is interpreted as the overall effect [14].

In reality, it is unlikely that every study would have the same outcome, considering the different populations and treatments used [32,41,49,52], particularly in observational studies [41]. Jones [41] suggests relaxing the assumption of a single fixed-effect to a set of fixed-effects reflecting the relative risks in several subgroups of interest. However, although these factors will be identifiable in principle, in practice it may not always be possible to estimate them from the available data. Further problems arise from the fact that combinations of effects may be of importance, and in these instances use of regression models may be indicated.

While the fixed-effects model may be useful where between-study variations that may affect relative risk estimates are knowable, the random-effects model not only allows for the occurrence of variation of true effects between studies but regards them as unknown, to be estimated by assuming that the effects observed in the sample of studies analyzed are drawn from a population of studies [41]. This model has an extra term compared with the fixed-effects model, as follows:

$$Y_j = \Delta + \beta_j + \epsilon_j,$$

where $\beta_j \sim N(0, \tau^2)$ is introduced to allow for heterogeneity between studies, and $\epsilon_j \sim N(0, \sigma_j^2)$ represents within-study variability of study j as before [14]. In the special case in which $\tau^2 = 0$, indicating homogeneity between studies, the random-effects model reduces to the fixed-effects model [6,14,50]. In principle, the random-effects model could be extended to include covariate information, the inclusion of which may substantially reduce heterogeneity of effects. In practice, however, covariate information is often missing from at least some studies [42].

A potential drawback of the random-effects model is the greater weight it gives to smaller trials [32,50,52,53], leading random-effects summaries to be more strongly biased than fixed-effects summaries by any tendency not to publish small nonsignificant studies [52]. Additionally, the random-effects model involves the assumption of a specific statistical distribution of effects that in practice may be difficult to justify [41,53]. Furthermore, although this model addresses the question of heterogeneity mathematically, it is of no use in identifying the source of the variation among the studies [49]. Indeed, it has been suggested that where use of the random-effects model makes a difference to the results, the analysis is incomplete and the investigator should search carefully for the source of the discrepancy between two models [41,52]. Despite these drawbacks, use of the random-effects model may be more appropriate in meta-analysis, as it tends towards the fixed-effects model if homogeneity is present, and allows for between-study heterogeneity if it is not [32,50,51]. However, some authors believe neither model gives a completely informative summary of the data when heterogeneity is present [52,53].

Although unpublished studies undoubtedly exist, the magnitude of any bias caused by failure to include them has never been well quantified. The degree of bias may depend on many other factors. Those identified include sampling biases that may influence the location of suitable studies, inclusion criteria and selection biases influencing which of the located studies are included in a meta-analysis, and biases in obtaining accurate data from the selected studies, particularly in scoring the quality of the studies under examination. Further biases may arise from the failure of the original author to present the study results accurately, either due to reporting or recording error [11,18,31,34,36]. It is also probable that other factors, some of which may be unknown, will influence any estimation of effect size, either in individual studies included in a meta-analysis, or in the meta-analysis itself. The interaction with publication bias of several of these factors, such as study size, reporting of data, the combining of randomized controlled trials and observational studies, and the model of meta-analysis chosen, has already been discussed, but how all these variables work together to influence the outcome of a meta-analysis is little understood. It should be remembered that publication bias is just one variable that may influence the outcome of a meta-analysis.

7. Methods of detecting and correcting for publication bias

As publication bias may seriously distort the findings of a meta-analysis, various methods have been devised for detecting its presence. Each of the methods is described below, in some cases with examples of its use, its chief advantages and limitations being listed in Table 1.

7.1 Proportion of significant studies

A simple way of attempting to detect publication bias is to look at the proportion of published studies that are significant. A survey of studies reported in four psychological journals found that over 94% of 294 papers confirmed the experimental hypotheses being tested [4]. High proportions were also found in a later study, based on papers published

Table 1
Methods of detecting and correcting for publication bias

Method	Advantages	Limitations
Proportion of significant studies	Simplicity	Does not actually demonstrate publication bias as no expected percentage of positive studies exists
Funnel graphs	Only requires published data	Symmetry defined informally, therefore open to interpretation, e.g., Vandenbroucke's [56] funnel graph of literature relating passive smoking to lung cancer in men [52]
Egger's method	Formal test for asymmetry in funnel graph	Statistical properties of method not described [73] Test may itself be biased [73] Even if asymmetry proved cause remains unknon [74]
Rank correlation test	Statistical analogue of funnel graph	Power of test highly variable; depends on characteristics of meta-analysis that may be unknown Publication bias cannot be ruled out in small meta-analyses if test not significant
Begg's method	Easy to carry out	Requires knowledge of the relative quantity of published and unpublished data Method assumes subjects in published studies are similar to those in unpublished studies in terms of relevant prognostic factors Assumes the sample size is constant for all unpublished studies
Truncated sampling	May have role in correcting for publication bias; however, limitations reduce this role	Strongly dependent on nature of distribution of P-values in range 0.00–0.05 that may be difficult to assess accurately Assumption that all significant studies are published may lead to bias if in reality only most significant ones are published
Weighted distribution theory	Estimated weight function is direct reflection of selection probabilities which give rise to publication bias: if constant no bias occurs, if proportional to effect size substantial bias occurs	Complex analysis with methodological problems Entire line of research in early stage of development Lack of commercially available software
Maximum likelihood	Tests explicitly whether certain studies have been censored	Assumes censorship is based on effect size rather than statistical significance Failure to adjust for differences in sample size across studies considered Good results cannot be guaranteed unless large sample is used Inadequate reporting may lead to results being excluded from meta-analysis Unknown method variables have potential to alter results of test
Fail-safe *N*	Allows assessment of effect of publication bias on results of meta-analysis	Plausibility of existence of certain number of unpublished studies is subjective judgement [59] Failure to locate all null studies never sole explanation of non-null pooled estimate [35] Primarily method to establish plausibility that publication bias explains all of observed association, therefore usefulness restricted Method assumes published and unpublished studies are of a similar size [12,13] Even in similar sized studies, method will be misleading if average effect of unpublished studies is in opposite direction to published studies [75]
Hackshaw's method	Allows assessment of effect of publication bias on results of meta-analysis	Usefulness restricted as primarily method to establish plausibility that publication bias explains all of observed association
Sugita's method	Can be used to correct for publication bias	Validity of assuming logarithm of relative risk is normally distributed is doubtful Should not be used when published relative risk estimates show significant heterogeneity
Givens' method	Can be used to correct for publication bias	Complex analysis involving numerous modeling assumptions and debatable choice of prior distributions For further comment on possible limitations see discussion of paper [14]

in four psychology journals and three medical journals [54]. The authors argued that, if publication bias did not exist, these high proportions implausibly suggested that "only studies with high power are performed and that the investigators formulate only true hypotheses."

An attempt to quantify publication bias [5] found that journals comprised only 7.6–12.3% of negative papers, compared to 20.5% for abstracts of papers presented at the American Psychological Association's 1962 annual meeting, and 30.2% for dissertation abstracts from the same year. An investigation into the subsequent publication of Ph.D. theses from 1956 found that while 12 of the 23 positive theses were published, only 2 of the 14 negative ones were. These statistically significant variations suggested that theses and dissertations were about three or four times more likely to be published if positive.

Other more recent studies have also found that published reports are more likely to be positive than unpublished ones [19,33,55]. One author concluded that it is difficult to give even a crude estimate of the size of the problem of publication bias from the information available, but using data from previous investigations, ratios of published to unpublished studies ranging from 128:1 to 1:1 were suggested, with most lying between 10:1 and 1:1 [19].

7.2. Funnel graphs

When diverse estimates of a value exist, some scatter around the underlying truth would be expected, the largest scatter being seen for estimates based on the smallest number of observations [29,56]. If the effect measure estimate is plotted against sample size the plot should, if there is no publication bias, resemble a funnel with a wide dispersion of results among small studies and a narrower range of results for large ones. Publication bias, however, will tend to skew the funnel shape, usually by excluding small studies with nonsignificant results, so that the lower left-hand region of the graph is missing or more sparsely occupied [11,13,29,56]. Funnel graphs have been used to demonstrate publication bias in the data relating passive smoking to heart disease [29], and to lung cancer in men [56], and also in obesity treatment trials [57].

Fig. 1 is a simulated example of a funnel plot, created by randomly drawing 100 samples of size varying from 50 to 2000 from an underlying normal distribution with a mean of 1 unit and standard deviation 10 units. The curves indicate the region within which 95% of samples of a given size are expected to fall. Closed circles indicate samples where the mean is significantly increased (above zero) at $P < 0.05$, open circles samples where it is not. For the full sample, the funnel shape is evident, but this would not be so if the open circles (or a proportion of them) were not included due to publication bias.

7.3 Statistical methods

7.3.1. Egger's method

Egger *et al*. [44] proposed that asymmetry in funnel graphs be tested for formally by carrying out a simple linear regression of y_i (the effect size in study i divided by its standard error) on x_i (the inverse of the standard error) and testing whether the intercept significantly differs (at $P < 0.1$) from zero. Based on data from eight meta-analyses, the authors suggested that the linear regression test of asymmetry may be more powerful than the rank correlation test described below.

7.3.2. Rank correlation

The rank correlation test, which has been described as a direct statistical analogue of the funnel graph, is used to detect correlation between effect size estimates, t, and sampling variance, v, among studies included in a meta-analysis, after first standardizing the effect sizes to stabilize the variance [58]. As variance is approximately inversely proportional to sample size in many applications, the test is similar to correlating effect size with sample size.

It should be noted that publication bias is not the only reasons why the effect size and sampling variance might be correlated. For example, large studies, despite their smaller sample variance, may, for reasons of cost, collect less detailed data than small studies, so allowing less control of confounding and hence a differing estimate of effect.

7.3.3. Begg's method

Begg's method [23] for estimating publication bias assumes that one has an estimate, y, of the true effect size, μ, based on a sample of n subjects, and that the number of subjects, N, with unpublished data is also known, or can be approximately guessed. The main method assumes that there are $N/n = m$ studies of a similar size to the published study, that μ is the same in all the $m + 1$ studies and that the observed value of y is the largest of the $m + 1$ studies, although the model, based on order-statistics, also allows one to assume the published study has a lower ranking.

7.3.4. Truncated sampling

If it is assumed that only studies with significant results are published, the bias in any single study can be determined by comparing the expected results conditional on a significant result with the expected results in the absence of this condition [13].

7.3.5. Weighted distribution theory

Weighted distribution theory, which is a generalization of truncated sampling, is based on the premise that a study is included in the analysis with a probability that is determined by its outcome, with the selection probabilities being related to different possible outcomes via a weight function [59].

7.3.6. Maximum likelihood

Using a maximum likelihood approach, Rust *et al*. [60] estimated the extent of publication bias given that all studies are published if the effect size is greater than a threshold but only a proportion of studies are published if the effect size is below it. Using their method, the authors demonstrated the existence of some publication bias in meta-analyses of consumer experiments and econometric advertising models, but none was found in a meta-analysis of proprietary research data.

7.3.7. Fail-safe N

Rosenthal's "fail-safe N" [2], N_{FS}, is the number of unpublished null studies needed to remove the significance from the findings of a meta-analysis. The method involves computing the standardized normal deviate Z_i associated with each published study and then calculating a combined deviate Z_c. The values of N_{FS} required to bring the new overall P-value to any desired level can then be calculated, an implausibly high value being regarded as evidence against the file-drawer hypothesis. It has been suggested that N_{FS} should be presented for all meta-analyses, as an aid in the assessment of the degree of confidence that can be placed in the results [61]. However, it has recently been demonstrated that N_{FS} may differ considerably from alternative estimates of the actual number, N, of unreported studies [62], N being estimated using the P-values reported in published studies, using two models, both of which assume the null hypothesis being tested is true. The first model assumed the P-values observed are the k smallest among the $N + k$

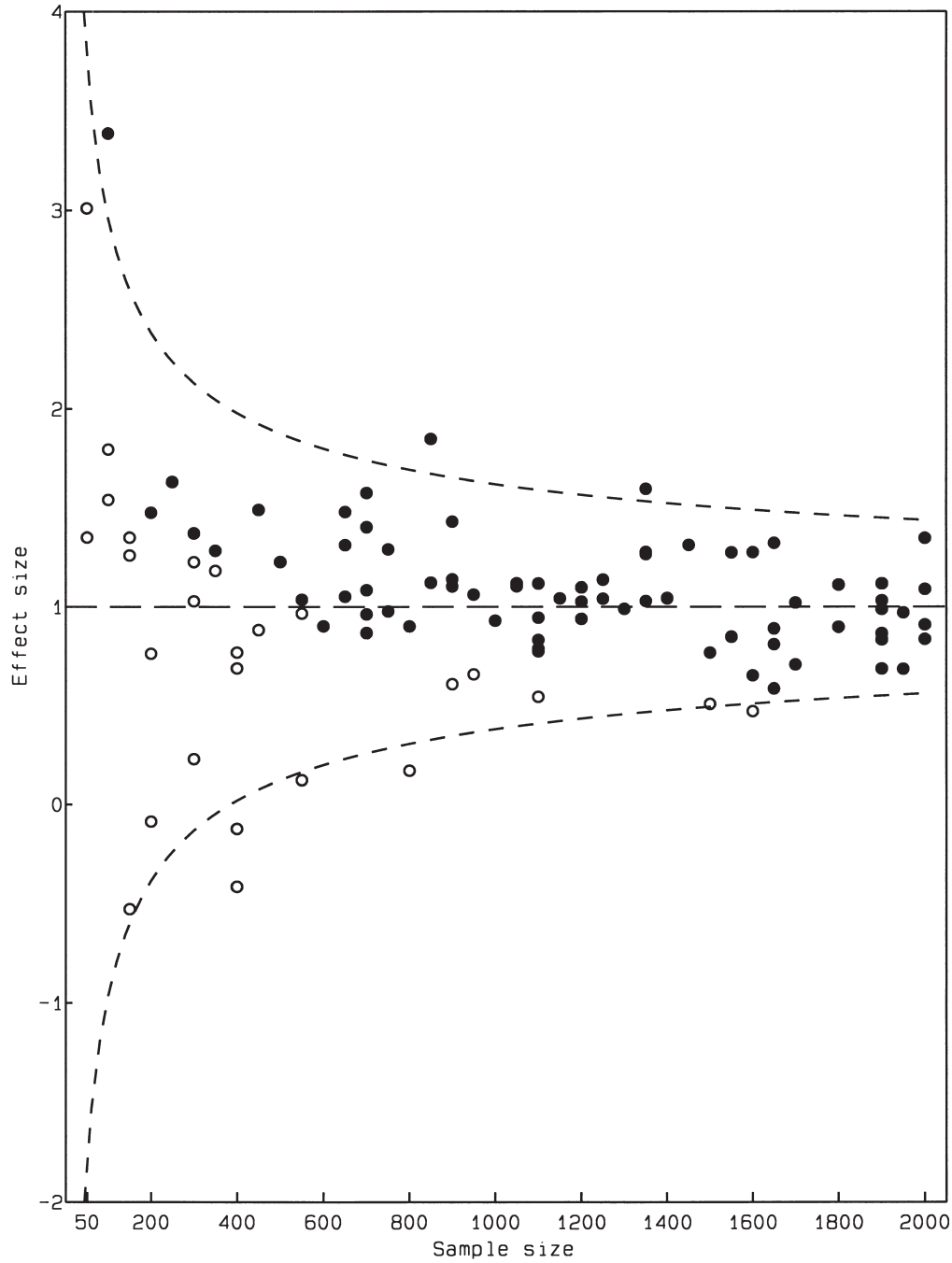

Fig. 1. Simulated funnel plot. (●) Effect size significantly increased ($P < 0.05$). (○) Effect size not significant. (— —) Expected value of effect size. (– –) Expected 95% confidence region for samples.

reported and unreported studies, while the second model assumed the probability that a study is reported is a function *g*(P), of the attained P-value.

7.2.8. Hackshaw's method

In this method [63], which is similar to the fail-safe *N*, the number of published studies actually available (39 in their example) is compared with the number one would expect in the absence of a true association, given the significance level and the number of studies significant at this level (280 based on 7 significant increases at $P < 0.05$, a chance of 1 in 40).

7.3.9. Sugita's method

This method [64] assumes that there is one unpublished study and the overall distribution of the logarithm of the relative risks of all studies, published and unpublished, is ex-

actly normal. By using well-known properties of the normal distribution, the relative risk and confidence limits of the unpublished study are calculated, and the revised meta-analysis estimate is then computed as usual. A slightly amended method [65] allows calculation of the publication probability of a study according to its odds ratio value. In one application of the method [66], based on data on passive smoking and lung cancer [67], a significant relative risk of 1.19 was reduced to a nonsignificant 1.11 after correction for publication bias.

7.3.10. Given's method

Givens *et al*. [14] used a Bayesian model to augment observed data by simulating the outcomes for missing studies, thereby creating a "complete" data set for meta-analysis. In this model, the observed data **Y** can be thought of as a partial realization of the random variable $X = (YZ)$, where a complete realization **X** is called the complete data, and a realization **Z** of Z is called the missing or latent data. The assumption that the distribution of X depends on parameters of interest θ through the family $p(\mathbf{X}|\theta)$ gives a marginal distribution $p(\mathbf{Y}|\theta)$ for the observed data. In publication bias, both the number and outcomes of unpublished studies are treated as latent data to augment the observed study outcomes. The authors describe how the random-effects model may be extended to account for publication bias, assuming that in addition to the n observed studies there are a further m studies that are not observed. The number m and relative risks found from these studies are unknown and must be estimated, and uncertainties about these estimates are reflected in the final meta-analysis inference by treating them as parameters in a Bayesian analysis. The method uses a fixed set of intervals to stratify P-values, but a more flexible, data-based determination of how the probability of publication depends on P-value may be obtained from methods that estimate the end points and number of such intervals rather than fixing them in advance.

As noted in Table 1, most of the methods of detecting, or correcting for, publication bias have specific drawbacks in their design. Additionally, all of them are based on strong assumptions that may, in practice, not be true. A further problem is the impossibility of rigorously testing the methods, due to the lack of a "gold standard" set of studies, which would include detected unpublished studies. The efficacy of each method in detecting, or correcting for, publication bias therefore remains purely theoretical.

8. Methods of preventing publication bias

8.1. Registries

Identifying published trials through the use of literature searches and computer databases is relatively straightforward, but information on unpublished trials is not as readily available. The use of registries has been advocated to overcome this, and registries already exist in the fields of perinatal medicine, cancer and acquired immunodeficiency syndrome treatment, and antithrombotic trials [19,33,68]. As registration usually occurs before results are known a complete database of all trials will be built up, thereby minimizing any possible effects of publication bias [18–20,33,68].

The value of registries is illustrated by data from trials evaluating the effect of chemotherapy on survival in advanced ovarian cancer. Although most published trials did not show a statistically significant benefit from chemotherapy, the trend usually favored this treatment and a pooled analysis showed a clear, but modest, enhancement in survival. However, pooled analysis of trials registered by 1983 failed to show a statistically significant survival advantage for chemotherapy, although there was still a small trend in favor of the treatment [69].

Obtaining regular information on all relevant trials being conducted is, however, difficult and time consuming. Thus, many current registries are limited by a failure to include an accurate and up-to-date log of all trials [13,68]. Furthermore, reviews may exclude trials that are published but not registered, producing bias if the results of registered and unregistered trials differ [69]. It may be possible to collect information through the research ethical committees from which researchers are required to seek approval before a study can commence, or through central funding agencies [13,66]. Additionally, major journals recently asked for details of all unpublished trials to be submitted for registration, with the intention of displaying the information on a computer Web site [70]. However, the incentive for registration is less than that for publication, so it is unlikely that this attempt to identify all trials will totally succeed [71]. It may also be difficult to obtain an accurate registry of observational studies, as these do not normally need approval from ethics committees.

8.2. Editorial policy

In the absence of registries, an editorial policy that undertakes to publish all studies of a high quality, regardless of results, is one way of overcoming publication bias [5,8,13,19,26,54,57]. A process of blinding both the scientist writing the paper and the journal refereeing it to a study's results could be used [12], in which two papers would be written, each with common introduction, methods, and results sections, but with no data entered. Only the discussion section would vary, reflecting the different conclusions drawn depending on the study outcome. The journal, which would be sent both versions, must agree to accept both or neither and on acceptance the data would be examined and the tables completed. Though this process involves additional effort and is perhaps unlikely to become popular, it would avoid the dependence of publication on outcome, and hence publication bias. Negative studies may not even have to be published in full; abstracts or titles with details of the source from which full reports could be obtained would be sufficient [5,13]. Additionally, journal editors and reviewers should insist on high standards in the conduct of research, including the clarification of the key aspects of a

study known to influence publication bias, such as proof of registration, whether the study was confirmatory, whether randomization took place, and any other known correlates of publication bias [13,26]. There is also some speculation as to whether the advent of electronic publishing will eliminate publication bias, as online journals will not be constrained by the same limits on space as conventional printed journals [72].

9. Conclusions

Publication bias appears to be a widespread problem in the scientific literature, and has been demonstrated in many fields of research. Various aspects of the design and execution of both single studies and meta-analyses may increase the probability of bias of this type, and its occurrence may seriously distort any attempts to derive valid estimates by pooling data from a group of studies, skewing the outcome towards positive results. Although various methods have been proposed for determining the presence of publication bias, and even correcting for it, all have their limitations. Therefore, the best option may be to prevent it from occurring in the first place, either by registering the existence of every trial undertaken, or by publishing all studies, regardless of their outcome. However, constraints of time and space may make these ideals hard to achieve. Until the problem of publication bias has been overcome, all reviewers and readers should be aware that they may be viewing a biased sample of experimental results and should moderate the strength of their conclusions accordingly. This is especially true when studying weak associations using the meta-analysis method, where the calculation of an overall estimate already endows the review with a semblance of accuracy that may not always be warranted.

Acknowledgments

We thank Mrs P.J. Wassell and Mrs D.P. Morris for their assistance in the typing of this manuscript, and Mrs B.A. Forey for preparing the simulated funnel plot. Financial support was provided by Philip Morris Europe, to whom we are also grateful.

References

[1] Last JM, editor. A Dictionary of Epidemiology. Oxford: Oxford University Press, 1983. p. 12.
[2] Rosenthal R. The “file drawer problem” and tolerance for null results. Psychol Bull 1979;86:638–41.
[3] Smith MB. Editorial. J Abnorm Social Psychol 1956;52:1–4.
[4] Sterling T. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. Am Stat Assoc J 1959;54:30–4.
[5] Smart RG. The importance of negative results in psychological research. Can Psychol 1964;5:225–32.
[6] Tweedie RL, Scott DJ, Biggerstaff BJ, Mengersen KL. Bayesian meta-analysis, with application to studies of ETS and lung cancer. Lung Cancer 1996;14(suppl. 1):S171–94.
[7] Angell M. Negative studies. N Engl J Med 1989;321:464–6.
[8] Newcombe RG. Towards a reduction in publication bias. BMJ 1987; 295:656–9.
[9] Pocock SJ, 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.
[10] Green SJ, Fleming TR, Emerson S. Effects on overviews of early stopping rules for clinical trials. Stat Med 1987;6:361–9.
[11] Felson DT. Bias in meta-analytic research. J Clin Epidemiol 1992;45: 885–892.
[12] Lee PN. Problems in interpreting epidemiological data. In: Mohr U, Bates DV, Dungworth DL, Lee PN, McClellan RO, Roe FJC, editors. Assessment of Inhalation Hazards. Berlin: Springer-Verlag, 1989. pp. 49–50.
[13] Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Stat Soc A 1988;151:419–63.
[14] Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis: A Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. Stat Sci 1997;12:221–50.
[15] Engler RI, Covell JW, Friedman PJ, Kitcher PS, Peters RM. Misrepresentation and responsibility in medical research. N Engl J Med 1983;317:1383–9.
[16] Rosenthal R, Fode KL. Psychology of the scientist: V. Three experiments in experimenter bias. Psychol Rep 1963;12:491–511.
[17] Rosenthal R. Experimenter outcome-orientation and the results of the psychological experiment. Psychol Bull 1964;61:405–12.
[18] Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. Stat Med 1987; 6:315–28.
[19] Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990;263:1385–9.
[20] Kleijnen J, Knipschild P. Review articles and publication bias. Arzneimittelforschung 1992;42:587–91.
[21] Minerva. Re publication bias. BMJ 1992;304:264.
[22] Light RJ. Accumulating evidence from independent studies: what we can win and what we can lose. Stat Med 1987;6:221–31.
[23] Begg CB. A measure to aid in the interpretation of published clinical trials. Stat Med 1985;4:1–9.
[24] Bulpitt CJ. Meta-analysis. Lancet 1988;ii:93–4.
[25] Koren G, Shear H, Graham K, Einarson T. Bias against the null hypothesis: the reproductive hazards of cocaine. Lancet 1989;ii:1440–2.
[26] Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. JAMA 1990;263:1392–5.
[27] Charlton B. Think negative: science needs its failures. New Scientist 1987;29Oct:72.
[28] Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cog Ther Res 1977;1:161–75.
[29] LeVois ME, Layard MW. Publication bias in the environmental tobacco smoke/coronary heart disease epidemiologic literature. Regul Toxicol Pharmacol 1995;21:184–91.
[30] Bero LA, Galbraith A, Rennie D. Sponsored symposia on environmental tobacco smoke. JAMA 1994;271:612–7.
[31] L’Abbé KA, Detsky AS, O’Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987;107:224–33.
[32] Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). Annu Rev Public Health 1996;17:1–23.
[33] Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991;337:867–72.
[34] Blair A, Burg J, Foran J, Gibb H, Greenland S, Morris R, Raabe G, Savitz D, Teta J, Wartenberg D, Wong O, Zimmerman R. Guidelines for application of meta-analysis in environmental epidemiology. Regul Toxicol Pharmacol 1995;22:189–97.
[35] Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 1987;9:1–30.

[36] Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450–5.
[37] Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McIlroy W, Oxman AD. Should unpublished data be included in meta-analyses? JAMA 1993;269:2749–53.
[38] Kilpatrick SJ. The epidemiology of environmental tobacco smoke (ETS) and the weight of evidence argument. Int Surg 1992;77:131–3.
[39] Pignon J-P, Arriagada R. Meta-analyses of randomized clinical trials: how to improve their quality? Lung Cancer 1994;10(suppl. 1):S135–41.
[40] Wald NJ, Nanchahal K, Thompson SG, Cuckle HS. Does breathing other people's tobacco smoke cause lung cancer? BMJ 1986;293: 1217–22.
[41] Jones DR. Meta-analysis of observational epidemiological studies: a review. J R Soc Med 1992;85:165–8.
[42] DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clin Trials 1986;7:177–88.
[43] Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996;347:363–6.
[44] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629–34.
[45] Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995;48:159–63.
[46] Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. Controlled Clin Trials 1998;19:159–66.
[47] Gøtzsche PC. Reference bias in reports of drug trials. BMJ 1987;295: 654–6.
[48] Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ 1997; 315:635–40.
[49] Morris RD. Meta-analysis in cancer epidemiology. Environ Health Perspect 1994;102(suppl. 8):61–6.
[50] Cappelleri JC, Ioannidis JPA, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 1996;276:1332–8.
[51] Helfenstein U, Steiner M. Fluoride varnishes (Duraphat): a meta-analysis. Community Dent Oral Epidemiol 1994;22:1–5.
[52] Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol 1994;140:290–6.
[53] Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. Stat Med 1990;9:657–71.
[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] Chan WC, Fung SC. Lung cancer in non-smokers in Hong Kong. In: Grundmann E, editor. Cancer Campaign Vol 6. Cancer Epidemiology. Stuttgart/New York: Gustav Fischer Verlag, 1982. pp. 199–202.
[56] Vandenbroucke JP. Passive smoking and lung cancer: a publication bias? BMJ 1988;296:391–92.
[57] Allison DB, *et al.* Publication bias in obesity treatment trials? Int J Obes Relat Metab Disord 1996;20:931–7.
[58] Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088–1101.
[59] Begg CB. Publication bias. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994. pp. 399–409.
[60] Rust RT, Lehmann DR, Farley JU. Estimating publication bias in meta-analysis. J Market Res 1990;XXVII:220–6.
[61] Persaud R. Misleading meta-analysis: "fail-safe N" is a useful mathematical measure of the stability of results [Letter]. BMJ 1996;312: 125.
[62] Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. Stat Med 1996;15:2493–2507.
[63] Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. BMJ 1997;315:980–8.
[64] Sugita M, Kanamori M, Izuno T, Miyakawa M. Estimating a summarized odds ratio whilst eliminating publication bias in meta-analysis. Jpn J Clin Oncol 1992;22:354–8.
[65] Sugita M, Yamaguchi N, Izuno T, Kanamori M, Kasuga H. Publication probability of a study on odds ratio value circumstantial evidence for publication bias in medical study areas. Tokai J Exp Clin Med 1994;19:29–37.
[66] Sugita M, Izuno T, Kanamori M. Recalculation of summarised odds ratios for the relationship between passive smoking and lung cancer based on data in the EPA report. Indoor Environment 1995;4:177–81.
[67] US Environmental Protection Agency. Respiratory health effects of passive smoking: lung cancer and other disorders. EPA/600/6-90-006F. Washington DC, 1992.
[68] Easterbrook P. Reducing publication bias [Letter]. BMJ 1987;295: 1347.
[69] Simes RJ. Confronting publication bias: a cohort design for meta-analysis. Stat Med 1987;6:11–29.
[70] Smith R. Misconduct in research: editors respond. The Committee on Publication Ethics (COPE) is formed [Editorial]. BMJ 1997;315: 201–2.
[71] Naylor CD. Meta-analysis and the meta-epidemiology of clinical research [Editorial]. BMJ 1997;315:617–9.
[72] Berlin JA. Will publication bias vanish in the age of online journals? [Editorial]. Online J Curr Clin Trials 1992; Doc No 12.
[73] Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased [Letter]. BMJ 1998;316:470.
[74] Vandenbroucke JP. Bias in meta-analysis detected by a simple, graphical test. Expert's views are still needed [Letter]. BMJ 1998; 316:469–70.
[75] Evans S. Misleading meta-analysis: statistician's comment. BMJ 1996;312:125.