11 Investigating and dealing with publication and other biases

JONATHAN A C STERNE, MATTHIAS EGGER, GEORGE DAVEY SMITH

Summary points

- Asymmetrical funnel plots may indicate publication bias, or be due to exaggeration of treatment effects in small studies of low quality.
- Bias is not the only explanation for funnel plot asymmetry. Funnel plots should be seen as a means of examining "small study effects" (the tendency for the smaller studies in a meta-analysis to show larger treatment effects) rather than a tool to diagnose specific types of bias.
- When markers of adherence to treatment or of the biological effects of treatment are reported, these may be used to examine bias without assuming a relationship between treatment effect and study size.
- Statistical methods may be used to examine the evidence for bias, and to examine the robustness of the conclusions of the meta-analysis in sensitivity analyses. "Correction" of treatment effect estimates for bias should be avoided, since such corrections may depend heavily on the assumptions made.
- Multivariable models may be used, with caution, to examine the relative importance of different types of bias.

Studies that show a statistically significant effect of treatment are more likely to be published,[1,2,3] more likely to be published in English,[4] more likely to be cited by other authors[5,6] and more likely to produce multiple publications[7,8] than other studies. Such "positive" studies are therefore more likely to be located for and included in systematic reviews, which may introduce bias. Trial quality has also been shown to influence the size of treatment effect estimates, with studies of lower methodological quality showing the larger effects.[9,10] These biases, reviewed in detail in Chapters 3 and 5, are more likely to affect small rather than large studies. The smaller

a study, the larger the treatment effect necessary for the results to be declared statistically significant. In addition, the greater investment of money and time in larger studies means that they are more likely to be of high methodological quality and published even if their results are negative. Bias in a systematic review may therefore become evident through an association between treatment effect and study size.

In this chapter we examine how we may check a meta-analysis for evidence of such bias, using graphical and statistical methods. We also examine methods for quantifying the possible impact of bias on overall treatment effect estimates, and for correcting effect estimates for bias.

Funnel plots

First used in educational research and psychology,[11] funnel plots are simple scatter plots of the treatment effects estimated from individual studies on the horizontal axis against some measure of study size on the vertical axis. The name "funnel plot" is based on the fact that the precision in the estimation of the underlying treatment effect will increase as the sample size of component studies increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias, the plot will resemble a symmetrical inverted funnel (see Figure 11.1(a)).

Choice of axes

Relative measures of treatment effect (risk ratios or odds ratios) are plotted on a logarithmic scale. This is important to ensure that effects of the same magnitude but opposite directions, for example risk ratios of 0·5 and 2, are equidistant from 1·0.[12] There are a number of possible choices for the measure of study size to be used as the vertical axis in funnel plots. Treatment effects have generally been plotted against sample size, or log sample size. However, the statistical power of a trial is determined both by the total sample size and the number of participants developing the event of interest. For example, a study with 100 000 patients and 10 events is less powerful than a study with 1000 patients and 100 events. Measures based on the standard error or variance of the effect estimate (or their inverse) rather than total sample size, have therefore been increasingly used in funnel plots. Plotting against standard error may generally be a good choice because it emphasizes differences between studies of smaller size for which biases are most likely to operate. In contrast, plotting against precision (1/standard error) will emphasize differences between larger studies. Using standard error is also consistent with statistical tests for funnel plot asymmetry,[13,14] discussed below, which look for associations between the treatment effect size and its standard error. A disadvantage of using standard error is that the

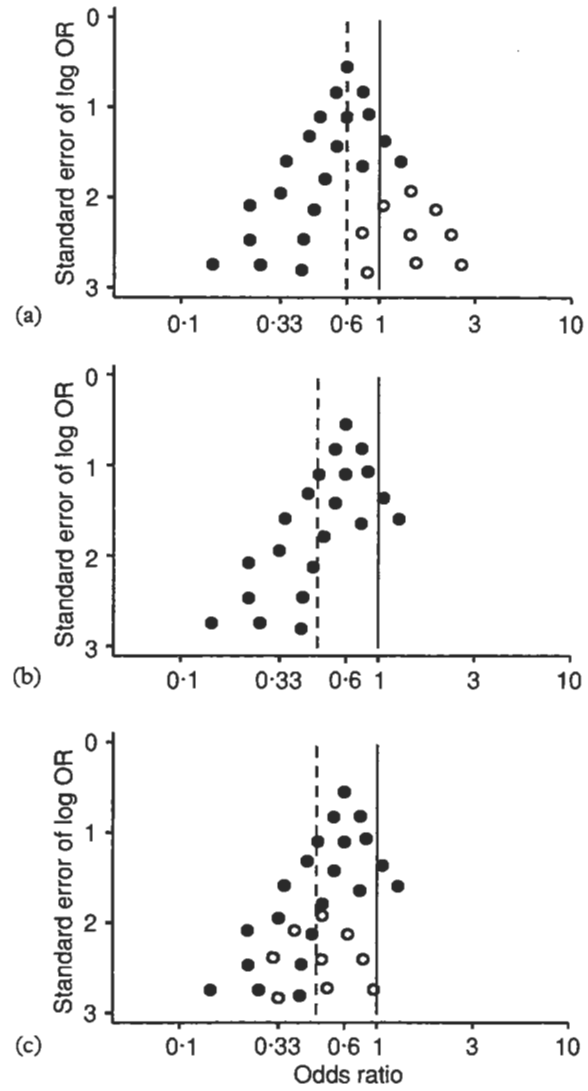

Figure 11.1 Hypothetical funnel plots: (a) symmetrical plot in the absence of bias (open circles indicate smaller studies showing no statistically significant effects); (b) asymmetrical plot in the presence of publication bias (smaller studies showing no statistically significant effects are missing); (c) asymmetrical plot in the presence of bias due to low methodological quality of smaller studies (open circles indicate small studies of inadequate quality whose results are biased towards larger effects). The dashed line is the pooled odds ratio, the solid line is the null effect. The estimated treatment effects are exaggerated in the presence of bias.

vertical axis must be inverted (smallest standard errors at the top) in order to produce the conventional inverted funnel graph.

Bias as a source of funnel plot asymmetry

Bias may lead to asymmetry in funnel plots. For example, if smaller studies showing no statistically significant effects (open circles in Figure 11.1) remain unpublished, then such publication bias[1,2,15] will lead to an asymmetrical appearance of the funnel plot with a gap in the right bottom side of the graph (Figure 11.1(b)). In this situation the combined effect from meta-analysis will overestimate the treatment's effect.[13,16] The more pronounced the asymmetry, the more likely it is that the amount of bias is substantial.

Trials of lower quality also tend to show larger effects. In particular, studies with inadequate concealment of treatment allocation or studies which are not double blind have been shown to result in inflated estimates of treatment effect [9,10] (see also Chapter 5). Smaller studies are, on average, conducted and analysed with less methodological rigour than larger studies. Trials that, if conducted and analysed properly, would have given no evidence for a treatment effect may thus become "positive", as shown in (Figure 11.1(c)), again leading to asymmetry. Thus the funnel plot should be seen as a generic means of examining "small study effects" (the tendency for the smaller studies in a meta-analysis to show larger treatment effects) rather than a tool to diagnose specific types of bias.

Asymmetry is not proof of bias: alternative sources of funnel plot asymmetry

The trials displayed in a funnel plot may not always estimate the same underlying effect of the same intervention and such heterogeneity between results may lead to asymmetry in funnel plots if the true treatment effect is larger in the smaller trials. For example, if a combined outcome is considered then substantial benefit may be seen only in patients at high risk for the component of the combined outcome which is affected by the intervention.[17,18] A cholesterol-lowering drug which reduces coronary heart disease (CHD) mortality will have a greater effect on all cause mortality in high risk patients with established cardiovascular disease than in young, asymptomatic patients with isolated hypercholesterolaemia.[19] This is because a consistent relative reduction in CHD mortality will translate into a greater relative reduction in all-cause mortality in high-risk patients in whom a greater proportion of all deaths will be from CHD. Trials conducted in high-risk patients will also tend to be smaller, because of the difficulty in recruiting such patients and because increased event rates mean that smaller sample sizes are required to detect a given effect.

Small trials are generally conducted before larger trials are established. In

the intervening years standard (control) treatments may have improved, thus reducing the relative efficacy of the experimental treatment. Changes in standard treatments could also lead to a modification of the effect of the experimental treatment. Such a mechanism has been proposed as an explanation for the discrepant results obtained in clinical trials of the effect of magnesium infusion in myocardial infarction.[20] It has been argued that magnesium infusion may not work if administered after reperfusion has occurred. By the time the ISIS-4 trial[21] (which gave no evidence of a treatment effect) was performed, thrombolysis had become routine in the management of myocardial infarction. However this argument is not supported by subgroup analysis of the ISIS-4 trial, which shows no effect of magnesium even among patients not receiving thrombolysis.[22]

Some interventions may have been implemented less thoroughly in larger trials, thus explaining the more positive results in smaller trials. This is particularly likely in trials of complex interventions in chronic diseases, such as rehabilitation after stroke or multifaceted interventions in diabetes mellitus. For example, an asymmetrical funnel plot was found in a meta-analysis of trials examining the effect of inpatient comprehensive geriatric assessment programmes on mortality.[13,23] An experienced consultant geriatrician was more likely to be actively involved in the smaller trials and this may explain the larger treatment effects observed in these trials.[13,23]

Odds ratios are more extreme (further from 1) than the corresponding risk ratio if the event rate is high. Because of this, a funnel plot which shows no asymmetry when plotted using risk ratios could still be asymmetric when plotted using odds ratios. This would happen if the smaller trials were consistently conducted in high-risk patients, and the large trials in patients at lower risk, although differences in underlying risk would need to be substantial. Finally it is, of course, possible that an asymmetrical funnel plot arises merely by the play of chance. Mechanisms which can lead to funnel plot asymmetry are summarised in Table 11.1.

Table 11.1 Potential sources of asymmetry in funnel plots.

1. *Selection biases*
Publication bias and other reporting biases (see Chapter 3)
Biased inclusion criteria
2. *True heterogeneity*: size of effect differs according to study size
Intensity of intervention
Differences in underlying risk
3. *Data irregularities*
Poor methodological design of small studies (see Chapter 5)
Inadequate analysis
Fraud
4. *Artefact*: heterogeneity due to poor choice of effect measure (see Chapter 16)
5. *Chance*

Funnel plot asymmetry thus raises the possibility of bias but it is not proof of bias. It is important to note, however, that asymmetry (unless produced by chance alone) will always lead us to question the interpretation of the overall estimate of effect when studies are combined in a meta-analysis.

Other graphical methods

Examining biological plausibility

In some circumstances, the possible presence of bias can be examined via markers of adherence to treatment, such as metabolites of a drug in patients' urine, or of the biological effects of treatment such as the achieved reduction in cholesterol in trials of cholesterol-lowering drugs, which, as discussed in Chapter 9, predicts the reduction in clinical heart disease[24,25] and mortality.[25]

If patients' adherence to an effective treatment were measured (for example as the percentage of patients actually taking the assigned medication), and varied across trials, then this should result in corresponding variation in treatment effects. Scatter plots of treatment effect (vertical axis) against adherence (horizontal axis) can be a useful means of examining this relationship. The scatter plot should be compatible with there being no treatment effect at 0 per cent adherence, and so a simple linear regression line should intercept the *y*-axis at zero treatment effect (Figure 11.2(a)). If a scatter plot indicates a treatment effect even when no patients adhere to treatment then bias is a possible explanation (Figure 11.2(b)). Similar considerations apply to scatter plots of treatment effect against change in biological markers believed to be closely associated with effects on clinical outcome. The advantage of such plots is that they provide an analysis that is independent of study size.

In a meta-analysis of trials examining the effect of reducing dietary sodium on blood pressure, Midgley *et al.*[26] plotted reduction in blood pressure (clinical outcome) against reduction in urinary sodium (biological marker) for each study and performed a linear regression analysis (Figure 11.3). The plot of difference in diastolic blood pressure (treatment effect) against change in urinary sodium (marker) suggests the possibility of bias. However, the assumption that the marker fully captures the treatment's effect on the clinical outcome may not always be appropriate: effects of the intervention not captured by the marker may account for the residual effect.[27,28] For example, dietary changes leading to a reduction in sodium intake may also lead to weight loss and hence to a reduction in blood pressure.

It should be noted that error in estimating the effect of the treatment on

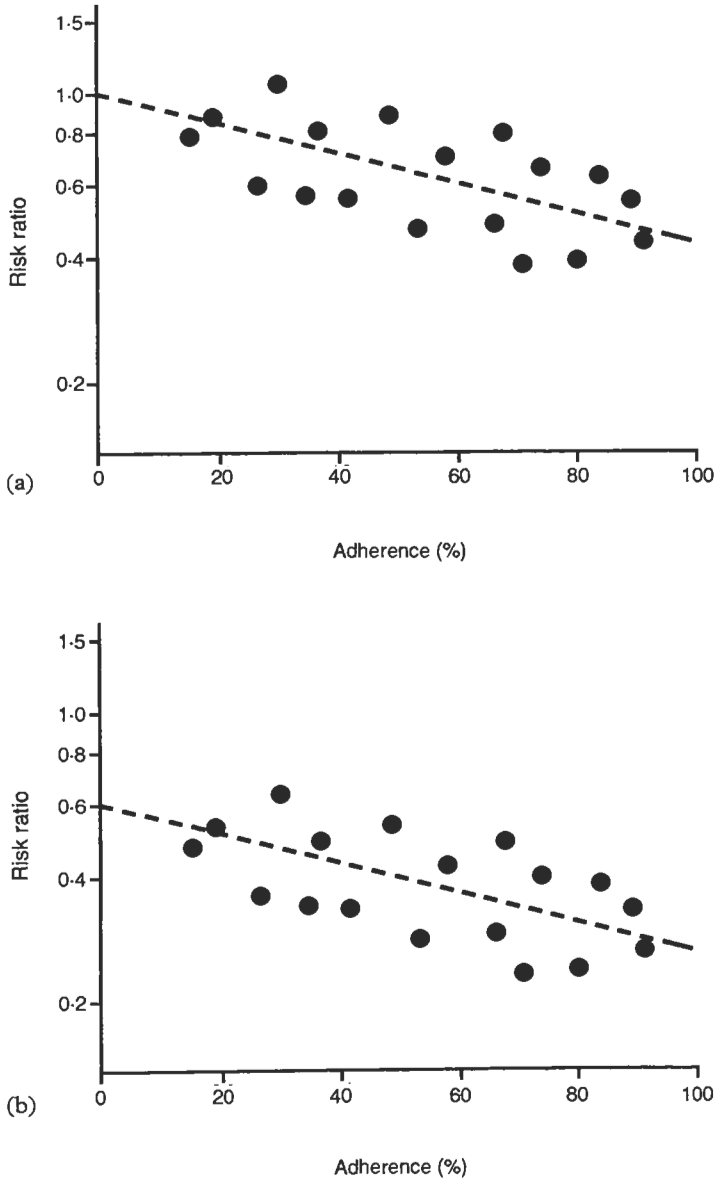

Figure 11.2 Linear regression plot of treatment effects from 18 hypothetical trials against the proportion of patients adhering to the experimental treatment. In the absence of bias the regression line intercepts the vertical axis at zero treatment effect (a). If the plot indicates a treatment effect even when no patients adhere to treatment (b) then bias is a likely explanation.

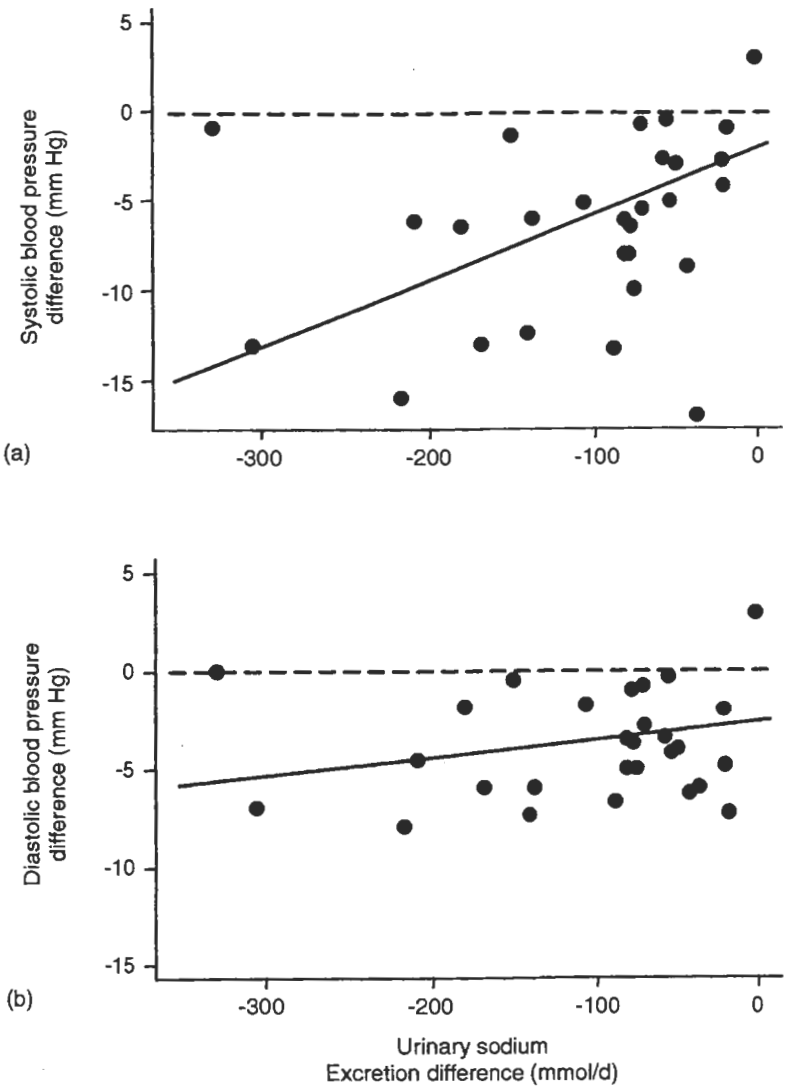

Figure 11.3 Regression lines, adjusted for the number of sodium excretion measurements, of the predicted change in blood pressure for a change in urinary sodium excretion from randomised controlled trials of dietary sodium reduction. Note that intercepts indicate a decline in blood pressure even if the diets in intervention and control groups were identical, which may indicate the presence of bias. Modified from Midgley *et al.*[26]

the marker could lead both to underestimation of its association with treatment, and to the estimated intercept being biased away from zero. In this situation a non-zero intercept could be misinterpreted as evidence of bias. We discuss how to use regression models to overcome this problem below.

Statistical methods to detect and correct for bias

Fail-safe *N*

Rosenthal[29] called publication bias the "file drawer problem", whose extreme version he described as "that the journals are filled with the 5% 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." Rosenthal proposed that the potential for publication bias to have influenced the results of a meta-analysis can be assessed by calculating the 'fail-safe *N*': the number of 'negative' studies (studies in which the treatment effect was zero) that would be needed to increase the P value for the meta-analysis to above 0·05. Iyengar and Greenhouse[30] noted that the estimate of fail-safe *N* is highly dependent on the mean treatment effect that is assumed for the unpublished studies. The method also runs against the widely accepted principle that in medical research in general, and systematic reviews in particular, one should concentrate more on the size of the estimated treatment effect and the associated confidence intervals, and less on whether the strength of the evidence against the null hypothesis reaches a particular, arbitrary, threshold.

Selection models

A number of authors have proposed methods to detect publication bias, based on the assumption that an individual study's results (for example the P value) affect its probability of publication. These methods model the selection process that determines which results are published and which are not, and hence are known as "selection models". However, as explained earlier, publication bias is only one of the reasons that will lead to associations between treatment effects and study size (small study effects). Since an unexpected distribution of study results is likely to imply an association between treatment effect size and study size, selection models should perhaps be seen as examining small study effects in general rather than publication bias in particular.

Iyengar and Greenhouse[30] assumed that publication was certain if the study P value was <0·05 (i.e. "statistically significant"). If the study P value was >0·05 ("non-significant") then publication probability might

be a constant (less than 1) or might decrease with decreasing treatment effect. Dear and Begg[31] and Hedges[32] extended this approach by assuming that different ranges of study P value (for example 0·01 to 0·05, 0·005 to 0·01 and so on) correspond to different publication probabilities. The observed distribution of P values is compared to the expected distribution assuming no publication bias, so that a reduced proportion of P values in (for example) the range 0·1 to 1 provides evidence of publication bias. These latter methods avoid strong assumptions about the nature of the selection mechanism but require a large number of studies so that a sufficient range of study P values is included.

Figure 11.4 (adapted from Dear and Begg[31]) shows the estimated publication probabilities from a meta-analysis of studies of the effect of open versus traditional education on creativity. Note that the apparent reduction in the probability of publication bias does not appear to coincide with the traditional cutoff of P = 0·05. These methods can be extended to estimate treatment effects, corrected for the estimated publication bias.[33] This approach was recently used in a meta-analysis of placebo-controlled trials of homoeopathy, an example discussed in more detail below.[34] A Bayesian approach in which the number and outcomes of unobserved studies are simulated has also been proposed as a means of correcting treatment estimates for publication bias.[35] For a meta-analysis examining

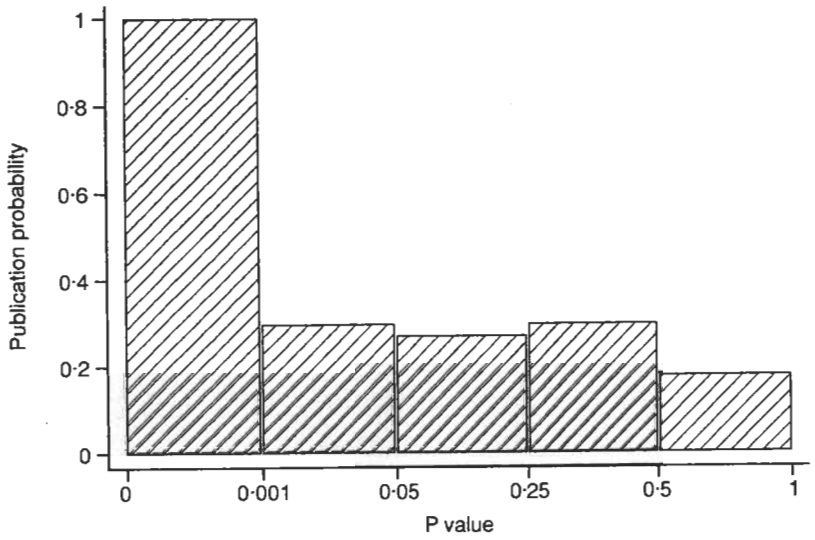

Figure 11.4 Publication probabilities according to study P value, estimated from a meta-analysis of 10 studies using a semi-parametric selection model (modified from Dear and Begg[31]).

the association between passive smoking and lung cancer, the relative risk was 1·22 (95% confidence interval 1·08 to 1·37) before and 1·14 (1·00 to 1·28) after allowing for publication bias.

The complexity of the statistical methods, and the large number of studies needed, probably explain why selection models have not been widely used in practice. In order to avoid these problems, Duval and Tweedie[36–38] have proposed "trim and fill": a method based on adding studies to a funnel plot so that it is symmetrical. The method works by omitting small studies until the funnel plot is symmetrical (trimming), using the trimmed funnel plot to estimate the true "centre" of the funnel, and then replacing the omitted studies and their missing "counterparts" around the centre (filling). As well as providing an estimate of the number of missing studies, an adjusted treatment effect is derived by performing a meta-analysis including the filled studies. Like other selection models, the method depends on the assumption that the association between treatment effect and trial size arises only because of publication bias, and not for the other reasons listed earlier in this chapter. Sutton *et al.*[39] used the trim and fill method to assess publication bias in 48 meta-analyses from the *Cochrane Database of Systematic Reviews*. They found that 56% of Cochrane meta-analyses had at least one missing study and may therefore be subject to publication bias, while in 10 the number of missing studies was statistically significant. However simulation studies have found that the trim-and-fill method detects "missing" studies in a substantial proportion of meta-analyses, even in the absence of bias.[40] There is thus a danger that uncritical application of the method could mean adding and adjusting for non-existent studies in response to funnel plot asymmetry arising from nothing more than random variation.

The 'correction' of effect estimates when publication bias is assumed to be present is problematic and a matter of ongoing debate. Results may be heavily dependent on the modelling assumptions used. Many factors may affect the probability of publication of a given set of results and it will be difficult if not impossible to model these adequately. It is therefore prudent to restrict the use of statistical methods which model selection mechanisms to the identification of bias rather than correcting for it.[41]

Copas[42] developed a model in which the probability that a study is included in a meta-analysis depends on its standard error. Because it is not possible to estimate all model parameters precisely, he advocates sensitivity analyses in which the value of the estimated treatment effect is computed under a range of assumptions about the severity of the selection bias. Rather than a single estimate treatment effect "corrected" for publication bias, the reader can see how the estimated effect (and confidence interval) varies as the assumed amount of selection bias increases. Application of the method to epidemiological studies of environmental tobacco smoke and

lung cancer suggests that publication bias may explain some of the association observed in meta-analyses of these studies.[43]

Statistical analogues of the funnel plot

An alternative approach, which does not attempt to define the selection process leading to publication or non-publication, is to use statistical methods to examine associations between study size and estimated treatment effects, thus translating the graphical approach of the funnel plot into a statistical model. Begg and Mazumdar[14] proposed an adjusted rank correlation method to examine the association between the effect estimates and their variances (or, equivalently, their standard errors). Egger *et al.*[13] introduced a linear regression approach in which the standard normal deviate (θ/s) is regressed against precision ($1/s$). This latter approach can be shown to correspond to a weighted regression of effect size θ on standard error s ($\theta = b_0 + b_1 s$), where the weights are inversely proportional to the variance of the effect size. The greater the value of the regression coefficient b_1, the greater the evidence for small study effects. Because each of these approaches looks for an association between treatment effect (e.g. log odds ratio) and its standard error in each study, they can be seen as statistical analogues of funnel plots of treatment effect against standard error. Both methods have been implemented in the statistical package Stata (see Chapter 18).

Sterne *et al.*[44] used simulation studies to investigate the sensitivity of the two methods (i.e. their ability to detect small study effects). The sensitivity of the methods was low in meta-analyses based on less than 20 trials, or in the absence of substantial bias. The regression method appeared more sensitive than the rank correlation method.

It has been claimed that the methods may give evidence of bias when bias is not in fact present (false-positive test results).[45] Sterne *et al.* found that the methods gave false-positive rates which were too high when there were large treatment effects, or few events per trial, or all trials were of similar sizes.[44] They concluded that the weighted regression method is appropriate and reasonably powerful in the situations where meta-analysis generally makes sense – in estimating moderate treatment effects, based on a reasonable number of trials – but that it should only be used if there is clear variation in trial sizes, with one or more trials of medium or large size.

Meta-regression

An obvious extension to the methods described above is to consider a measure of study size (for example the standard error of the effect estimate) as one of a number of different possible explanations for between-study heterogeneity in a multivariable 'meta-regression' model (see also Chapters 8 to 10 for a discussion of the use of regression models in meta-analysis).

For example, the effects of study size, adequacy of randomisation and type of blinding might be examined simultaneously. Thompson and Sharp[46] recently reviewed different methods for meta-regression. These have been implemented in Stata (see Chapter 18).

Three notes of caution are necessary. Users of standard regression models know that it is unwise to include large numbers of covariates, particularly if the sample size is small. In meta-regression the number of data points corresponds to the number of studies, which is usually less than 50 and often less than 10.[44] Thus tests for association between treatment effect and large numbers of study characteristics may lead to "overfitting" and spurious claims of association. Secondly, all associations observed in such analyses are observational, and may therefore be confounded by other unknown or unmeasured factors. Thirdly, regression analyses using averages of patient characteristics from each trial (such as the mean age of all the patients) can give a misleading impression of the relation for individual patients. As discussed in Chapter 9, there is potential for the so-called ecological fallacy,[47] whereby the relation with treatment benefit may be different across trials as compared to within trials.

Meta-regression could be used to examine associations between clinical outcomes and markers of adherence to treatment or of the biological effects of treatment, weighting appropriately for study size. As discussed above, the intercept (coefficient of the constant term) should be zero if there is no biological effect so a non-zero intercept may be evidence of bias, or of a treatment effect which is not mediated via the marker. Unless the error in estimating the effect of treatment on the marker is small, this error must be incorporated in models of the association between the treatment effect and the change in the surrogate marker. Daniels and Hughes[48] discuss this issue and propose a Bayesian estimation procedure. This method has been applied in a study of CD4 cell count as a surrogate endpoint in HIV clinical trials.[49]

The case study illustrates the use of some of the methods described in this chapter, using the example of a widely cited meta-analysis of placebo-controlled trials of homoeopathy.

Case study: is the effect of homoeopathy due to the placebo effect?

The placebo effect is a popular explanation for the apparent efficacy of homoeopathic remedies.[50,51,52] Linde *et al.* addressed this question in a fascinating systematic review and meta-analysis of placebo-controlled trials of homoeopathy, in which all trials, independent of clinical condition and outcomes, were included.[34] The authors performed an extensive literature search, without language restrictions, covering a wide range of bibliographic databases and complementary medicine registries. Linde *et al.* included 89 published and unpublished reports of randomised placebo-controlled trials. Quality assessment covered the dimensions of internal validity that are

known to be associated with treatment effects[9,10] (see Chapter 5): concealment of the allocation of homoeopathic remedies or placebo, blinding of outcome assessment and handling of withdrawals and dropouts.

The funnel plot of the 89 homoeopathy trials is clearly asymmetrical (Figure 11.5(a)) and both the rank correlation and the weighted regression

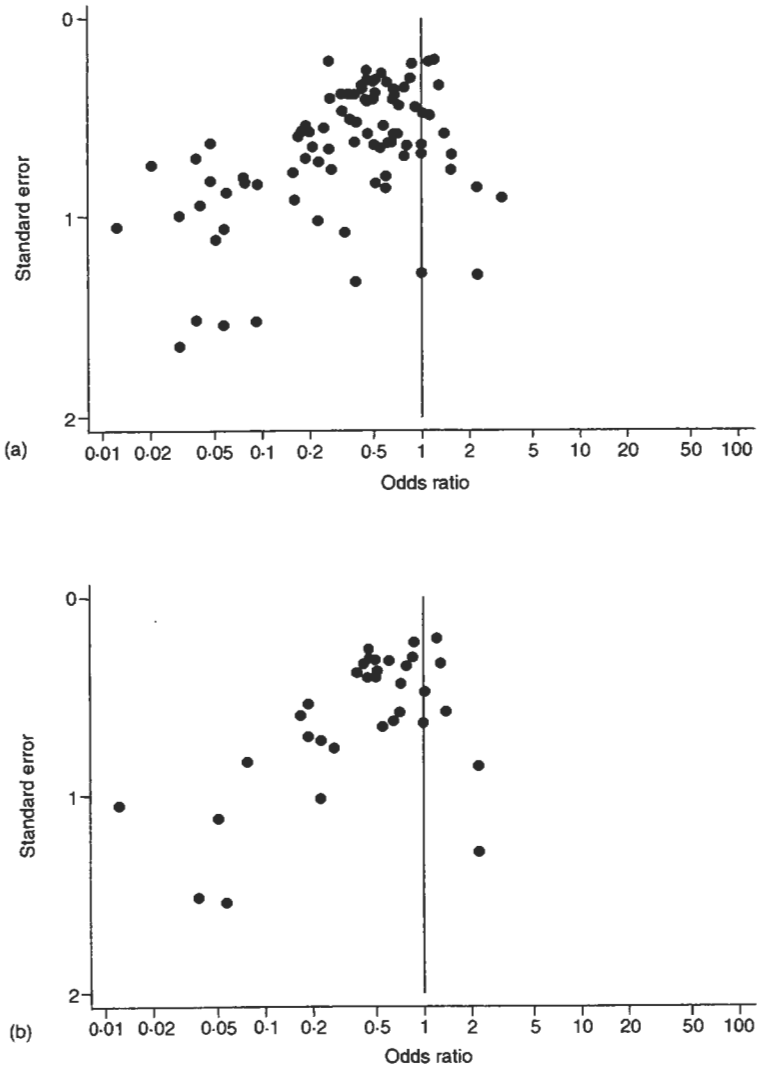

Figure 11.5 Funnel plot of all 89 randomised controlled trials comparing homoeopathic medicine with placebo identified by Linde *et al.*[34] (a) and 34 trials of high methodological quality (b).

test indicated clear asymmetry (P<0·0001). This asymmetry remained (Figure 11.5(b)) when the plot was restricted to 34 double-blind trials with adequate concealment of treatment allocation (P = 0·014 with rank correlation and <0·001 with regression method). The authors used a selection model (assuming that the likelihood that a study was reported depended on the one-tailed P value)[32,33] to correct for publication bias, and found that the odds ratio was increased from 0·41 (95% confidence interval 0·34 to 0·49) to 0·56 (0·32 to 0·97) after correcting for bias. Similar results are obtained with the fill and trim method (Figure 11.6). To make the funnel plot symmetric, 16 studies are added. The adjusted odds ratio (including the filled studies) is 0·52 (0·43 to 0·63).

Linde *et al.* therefore concluded that the clinical effects of homoeopathy are unlikely to be due to placebo.[34] However the method they used does not allow simultaneously for other sources of bias, and thus assumes that publication bias is the sole cause of funnel plot asymmetry. Table 11.2 shows the results from meta-regression analyses of associations between trial characteristics and the estimated effect of homoeopathy. Results are presented as ratios of odds ratios (ORs) comparing the results from trials with to trials without the characteristic. Thus ratios below 1 correspond to a smaller treatment odds ratio for trials with the characteristic, and hence a larger apparent benefit of homoeopathic treatment. For example, in univariable analysis the odds ratio was reduced by factor 0·24 (ratio of ORs

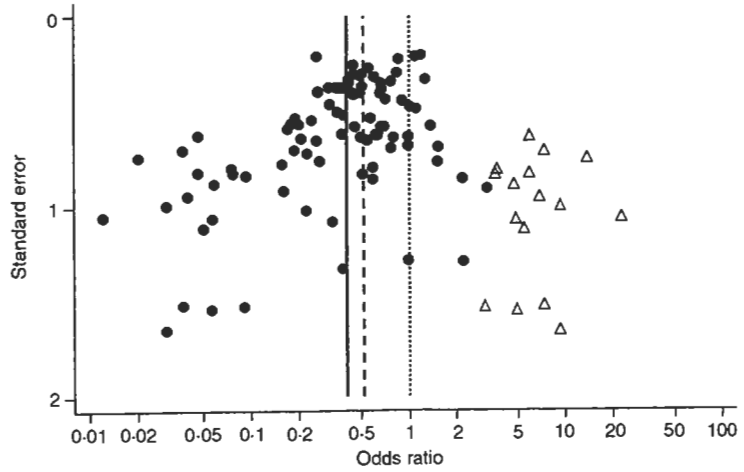

Figure 11.6 Application of the "trim and fill" method to the funnel plot of 89 trials comparing homoeopathic medicine with placebo. The solid circles represent the 89 trials, while the open triangles are the "filled" studies. The solid line is the original (random-effects) estimate of the pooled odds ratio, the dashed line is the adjusted estimate (including the filled studies) and the dotted line is the null value (1).

Table 11.2 Meta-regression analysis of 89 homoeopathy trials.

Study characteristic	Univariable analysis		Controlling for all variables	
	Ratio of odds ratios* (95% CI)	P	Ratio of odds ratios* (95% CI)	P
Unit increase in standard error of log OR	0·18 (0·10 to 0·34)	<0·001	0·20 (0·11 to 0·37)	<0·001
Language (Non-English vs. English)	0·73 (0·51 to 1·06)	0·097	0·73 (0·55 to 0·98)	0·038
Study quality				
Allocation concealment (not adequate vs. adequate)	0·70 (0·49 to 1·01)	0·054	0·98 (0·73 to 1·30)	0·87
Blinding (not double-blind vs. double-blind)	0·24 (0·12 to 0·46)	<0·001	0·35 (0·20 to 0·60)	<0·001
Handling of withdrawals (not adequate vs. adequate)	1·32 (0·87 to 1·99)	0·19	1·10 (0·80 to 1·51)	0·56
Publication type (not MEDLINE-indexed vs. MEDLINE-indexed)	0·61 (0·42 to 0·90)	0·013	0·91 (0·67 to 1·25)	0·57

*Odds ratio with characteristic divided by odds ratio without characteristic. Ratios below 1 correspond to a smaller treatment odds ratio for trials with the characteristic, and hence a larger apparent benefit of homoeopathic treatment.

0·24, 95% confidence interval 0·12 to 0·46) if outcome assessment was not blinded. Inadequate concealment of allocation, and publication in a non-MEDLINE-indexed journal were also associated with greater estimated benefits. Univariable analyses provided strong evidence of small study effects: trials with larger standard errors had substantially greater estimated benefits of homoeopathic treatment. When the effect of each variable was controlled for all others there remained strong associations with standard error (ratio of ORs 0·20, 95% confidence interval 0·11 to 0·37), and inadequate blinding of outcome assessment (ratio of ORs 0·35, 95% confidence interval 0·20 to 0·60).

Meta-regression analysis and funnel plots indicate that the treatment effects seen are strongly associated with both methodological quality and study size. The two largest (i.e. with the smallest standard error) trials of homoeopathy which were double blind and had adequate concealment of randomisation (on 300 and 1270 patients) show no effect. This is consistent with the intercept from the regression of effect size on standard error, which can be interpreted as the estimated effect in large trials (OR = 1·02, 95% confidence interval 0·71 to 1·46). The evidence is thus compatible with the hypothesis that the clinical effects of homoeopathy are completely due to placebo. This example illustrates that publication bias is only one of

the reasons why the results from small studies may be misleading. However, we emphasise that our results cannot *prove* that the apparent benefits of homoeopathy are due to bias.

Conclusions

Prevention is better than cure. In conducting a systematic review and meta-analysis, investigators should make strenuous efforts to ensure that they find all published studies, and to search for unpublished work, for example in trial registries or conference abstracts (see Chapter 4). The quality of component studies should also be carefully assessed (Chapter 5). Summary recommendations on examining for and dealing with bias in meta-analysis are shown in Box 11.2. Selection models for the process of publication bias are likely to be of most use in sensitivity analyses in which the robustness of a meta-analysis to possible publication bias is assessed. Funnel plots should be used in most meta-analyses, to provide a visual assessment of whether treatment effect estimates are associated with study

Box 11.1 Summary recommendations on investigating and dealing with publication and other biases in a meta-analysis

Examining for bias

- Check for funnel plot asymmetry using graphical and statistical methods.
- Consider meta-regression to look for associations between key measures of trial quality and treatment effect size.
- Consider meta-regression to examine other possible explanations for heterogeneity.
- If available, examine associations between treatment effect size and changes in biological markers or patients' adherence to treatment.

Dealing with bias

- If there is evidence of bias, report this with the same prominence as any combined estimate of treatment effect.
- Consider sensitivity analyses to establish whether the estimated treatment effect is robust to reasonable assumptions about the effect of bias.
- Consider excluding studies of lower quality.
- If sensitivity analyses show that a review's conclusions could be seriously affected by bias, then consider recommending that the evidence to date be disregarded.

size. Statistical methods which examine the evidence for funnel plot asymmetry are now available, and meta-regression methods may be used to examine competing explanations for heterogeneity in treatment effects between studies. The power of all of these methods is limited, however, particularly for meta-analyses based on a small number of small studies, and the results from such meta-analyses should therefore always be treated with caution.

Statistically combining data from new trials with a body of flawed evidence will not remove bias. However there is currently no consensus on how to guide clinical practice or future research when a systematic review suggests that the evidence to date is unreliable for one or more of the reasons discussed in this chapter. If there is clear evidence of bias, and if sensitivity analyses show that this could seriously affect a review's conclusions, then reviewers should not shrink from recommending that some or all of the evidence to date be disregarded. Future systematic reviews could then be based on new, high-quality evidence. Important improvements such as better conduct and reporting of trials, prospective registration, easier access to data from published and unpublished studies and comprehensive literature searching are being made to the process of assessing the effect of medical interventions. It is to be hoped that these will mean that bias will be a diminishing problem in future systematic reviews and meta-analyses.

Acknowledgements

We are grateful to Klaus Linde and Julian Midgley who provided unpublished data from their meta-analyses.

1 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;**337**:867–72.
2 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 1992;**263**:374–8.
3 Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;**315**:640–5.
4 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.
5 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;**295**:654–6.
6 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;**305**:15–19.
7 Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;**315**:635–40.
8 Gøtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;**36**:429–32.
9 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.

10 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.
11 Light RJ, Pillemer DB. *Summing up: the science of reviewing research.* Cambridge: Harvard University Press, 1984.
12 Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med* 1988;7:889–94.
13 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.
14 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;**50**:1088–101.
15 Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Statist Soc A* 1988;**151**:419–63.
16 Villar J, Piaggio G, Carroli G, Donner A. Factors affecting the comparability of meta-analyses and the largest trials results in perinatology. *J Clin Epidemiol* 1997;**50**:997–1002.
17 Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. *BMJ* 1994;**308**:72–4.
18 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;**311**:1356–9.
19 Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ* 1993;**306**:1367–73.
20 Baxter GF, Sumeray MS, Walker JM. Infarct size and magnesium: insights into LIMIT-2 and ISIS-4 from experimental studies. *Lancet* 1996;**348**:1424–6.
21 ISIS-4 (Fourth International Study of Infact Survival) 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**:669–85.
22 Collins R, Peto R. Magnesium in acute myocardial infarction. *Lancet* 1997;**349**:282
23 Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 1993;**342**:1032–6.
24 Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* 1993;**2**:173–92.
25 Holme I. Cholesterol reduction and its impact on coronary artery disease and total mortality. *Am J Cardiol* 1995;**76**:10C–17C.
26 Midgley JP, Matthew AG, Greenwood CMT, Logan AG. Effect of reduced dietary sodium on blood pressure. A meta-analysis of randomized controlled trials. *JAMA* 1996;**275**:1590–7.
27 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; **8**:431–40.
28 Fleming TR, Demets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;**125**:605–13.
29 Rosenthal R. The "file drawer" problem and tolerance for null results. *Psychol Bull* 1979;**86**:638–41.
30 Iyengar S, Greenhouse JB. Selection problems and the file drawer problem. *Stat Sci* 1988;109–35.
31 Dear KBG, Begg CB. An approach to assessing publication bias prior to performing a meta-analysis. *Stat Sci* 1992;7:237–45.
32 Hedges LV. Modeling publication selection effects in meta-analysis. *Stat Sci* 1992;7:246–55.
33 Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995;**60**:419–35.
34 Linde K, Clausius N, Ramirez G, *et al.* Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.
35 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.
36 Taylor SJ, Tweedie RL. Practical estimates of the effect of publication bias in meta-analysis. *Austral Epidemiol* 1998;5:14–17.

37 Duval SJ, Tweedie RL. Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;**56**:455–63.
38 Duval SJ, Tweedie RL. A non-parametric "trim and fill" method of assessing publication bias in meta-analysis. *J Am Stat Assoc* 2000;**95**:89–98.
39 Sutton AJ, Duval SJ, Tweedie RL, *et al.* Empirical assessment of effect of publication bias on meta-analyses. *BMJ* 2000;**320**:1574–7.
40 Sterne JAC, Egger M. High false positive rate for trim and fill method (electronic letter). http://www.bmj.com/cgi/eletters/320/7249/1574
41 Begg CB. Publication bias in meta-analysis. *Stat Sci* 1997;**12**:241–4.
42 Copas J. What works?: selectivity models and meta-analysis. *J R Statist Soc A* 1999;**162**:95–109.
43 Copas JB, Shi JQ. Reanalysis of epidemiological evidence on lung cancer and passive smoking. *BMJ* 2000;**320**:417–18.
44 Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53**:1119–29.
45 Irwig L, Macaskill P, Berry G, Glasziou P. Graphical test is itself biased. *BMJ* 1998;**316**:470–1.
46 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;**18**:2693–708.
47 Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;**72**:1336–44.
48 Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* 1997;**16**:1965–82.
49 Hughes MD, Daniels MJ, Fischl MA, Kim S, Schooley RT. CD4 cell count as a surrogate endpoint in HIV clinical trials: a meta-analysis of studies of the AIDS Clinical Trials Group. *AIDS* 1998;**12**:1823–32.
50 Maddox J, Randi J, Stewart WW. "High-dilution" experiments a delusion. *Nature* 1988;**334**:287–91.
51 Gøtzsche PC. Trials of homeopathy. *Lancet* 1993;**341**:1533.
52 Vandenbroucke JP. Homoeopathy trials: going nowhere. *Lancet* 1997;**350**:824.

Part III: Systematic reviews of observational studies