STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:2113–2126 (DOI: 10.1002/sim.1461)

Adjusting for publication bias in the presence of heterogeneity

Norma Terrin[1,*,†], Christopher H. Schmid[1], Joseph Lau[1] and Ingram Olkin[2]

[1]*Division of Clinical Care Research, New England Medical Center, Boston, U.S.A.*
[2]*Department of Statistics, Stanford University, Stanford, CA, U.S.A.*

SUMMARY

It is known that the existence of publication bias can influence the conclusions of a meta-analysis. Some methods have been developed to deal with publication bias, but issues remain. One particular method called 'trim and fill' is designed to adjust for publication bias. The method, which is intuitively appealing and comprehensible by non-statisticians, is based on a simple and popular graphical tool called the funnel plot. We present a simulation study designed to evaluate the behaviour of this method. Our results indicate that when the studies are heterogeneous (that is, when they estimate different effects), trim and fill may inappropriately adjust for publication bias where none exists. We found that trim and fill may spuriously adjust for non-existent bias if (i) the variability among studies causes some precisely estimated studies to have effects far from the global mean or (ii) an inverse relationship between treatment efficacy and sample size is introduced by the studies' *a priori* power calculations. The results suggest that the funnel plot itself is inappropriate for heterogeneous meta-analyses. Selection modelling is an alternative method warranting further study. It performed better than trim and fill in our simulations, although its frequency of convergence varied, depending on the simulation parameters.

KEY WORDS: simulation; trim and fill; funnel plot; selection model

1. INTRODUCTION

It is generally recognized that medical findings are more likely to be published, and published sooner, if they achieve statistical significance [1–4]. The resulting bias may lead to erroneous conclusions in meta-analysis that could seriously impact clinical practice.

The discussion of potential publication bias is becoming more frequent in published meta-analyses. Among articles that address the issue, there appears to be an increased reliance on the funnel plot and statistical methods based on it [5–13]. Studies have also used funnel plot methods to estimate the fraction of meta-analyses in the medical literature affected by publication bias [14–16]. It is therefore critical to examine the reliability of these methods.

* Correspondence to: Norma Terrin, New England Medical Center, 750 Washington Street, Box 63, Boston, MA 02111, U.S.A.
† E-mail: nterrin@tufts-nemc.org

Contract/grant sponsor: AHRQ; contract/grant number: R01 HS10254.

Received June 2001
Accepted November 2002

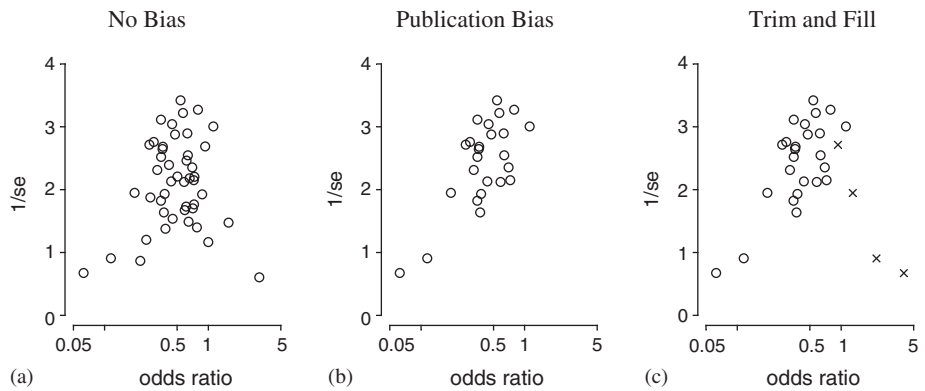

Figure 1. (*a*) Funnel plot of a simulated collection of trials with sample size 50–500 and true odds ratio 0.50. (*b*) A subset of the previous studies, with selection based on *p*-value and sample size. (*c*) Crosses were added to represent studies imputed by trim and fill.

The funnel plot is a scatter plot of the component studies in a meta-analysis, with the treatment effect on the horizontal axis, and a weight, such as the inverse standard error, or sample size, on the vertical axis. Light and Pillemer, originators of the funnel plot, explained, 'If all studies come from a single underlying population, this graph should look like a funnel, with the effect sizes homing in on the true underlying value as n increases… [If there is publication bias,] there should be a bite out of the funnel…' [17] (Figures 1(*a*) and (*b*)).

Although the simplicity of the funnel plot is appealing, it is not based on an underlying model that can be tested statistically. Consequently, various statistical methods have been developed to overcome the subjective visual interpretation of the plot. Rank correlation [18] and regression [14, 19] detect funnel plot asymmetry and 'trim and fill' [20] calculates a pooled treatment effect that adjusts for bias by imputing studies that make the funnel plot more symmetric (Figure 1(*c*)).

Meta-analysts frequently ignore Light's and Pillemer's caveat that the studies come from a single underlying population. For example, Boffeta *et al.* [9] found that the carcinogenic effect of diesel fuel differs for bus drivers, truck drivers and railroad workers. The authors correctly pooled the studies separately for the three different groups, but to test for publication bias, they applied a funnel plot regression method [14] to all studies together. In another meta-analysis, Fleischaur *et al.* [12] found that the protective effect of garlic differed for various types of cancer, but the funnel plot they used to check for publication bias included results for all cancers.

A collection of studies is called homogeneous if all studies estimate a single true effect, heterogeneous if they estimate a range of effects. Figures 2(*a*) and (*b*) show two ways heterogeneity can cause asymmetric funnel plots, even in the absence of publication bias. Figure 2(*a*) displays a heterogeneous collection of simulated studies, where the mean odds ratio is 0.50. There is no underlying relation between precision and benefit, but some precisely estimated odds ratios are considerably higher than the global mean, simply due to chance. The simulation in Figure 2(*b*) assumes that each of the studies has 80 per cent power to detect the true treatment effect associated with the study. Because the true effects cover a wide range, the sample size requirements under conditions where the treatment is highly effective are much

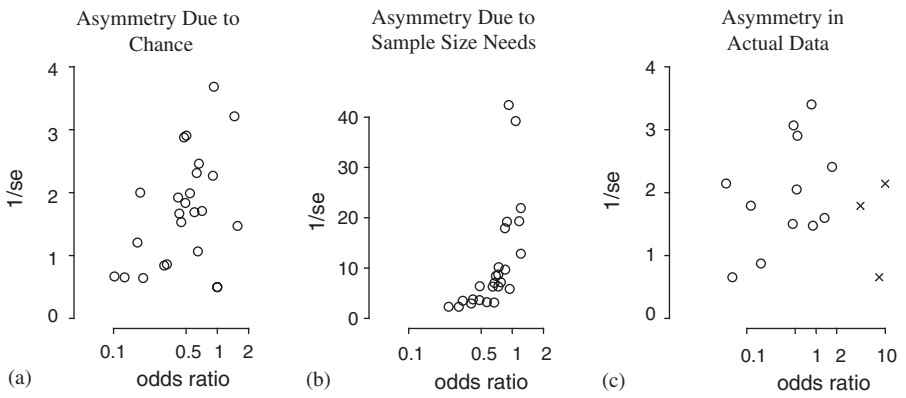

Figure 2. (*a*) Funnel plot of a heterogeneous collection of simulated trials, with true odds ratio 0.50 and no publication bias. (*b*) Funnel plot of a heterogeneous collection of simulated trials with mean odds ratio 0.80 and sample size selected for 80 per cent power against the true treatment effect. There is no publication bias. (*c*) Funnel plot of the effect of antibiotics on otitis media [21]. The crosses are studies imputed by trim and fill.

lower than where the treatment has minimal effect. In real meta-analyses, the relationship is not as extreme as in Figure 2(*b*), because investigators' ability to approximate treatment benefit prior to a study depends on the quality and availability of relevant prior research.

An example of funnel plot asymmetry that could have been caused by heterogeneity is shown in Figure 2(*c*), which displays 12 studies on the short term benefit of antibiotics for otitis media with effusion [21]. (Funnel plot methods were not used in the original article.) If the asymmetry were caused by publication bias alone, then the application of trim and fill would be appropriate, but trim and fill imputed unrealistically high odds ratios in this case (as high as ten in favour of not treating otitis media with antibiotics), suggesting that the asymmetry may be due to other causes. Heterogeneity is a plausible explanation (the odds ratios are heterogeneous by the Q-statistic [22], $p < 0.001$).

The study presented here is a large simulation study of the impact of heterogeneity on methods for publication bias. Simulations were needed because the methods can only be evaluated if the true extent of bias is known *a priori*. We focused on trim and fill [20] because (i) it has intuitive appeal, (ii) it calculates an adjusted pooled treatment effect (detection methods only warn that the estimate is suspect), and (iii) it was designed to allow for heterogeneity.

Simulation of the meta-analyses is described in Section 2. Section 3 contains a description of the trim and fill method and the results of applying trim and fill to the simulated meta-analyses. In Section 4 we explain and show results for selection modelling [23], an adjustment method that does not use the funnel plot. Section 5 is a discussion of results.

2. SIMULATION MODELS

We used fixed effects models to simulate homogeneous meta-analyses (Section 2.1) and random effects models to simulate heterogeneous meta-analyses (Section 2.2). The random effects

Table I. Sample size.

Number of studies per meta-analysis	Sample size distribution of the studies
10 and 25	*A*: 50 to 500, on the log scale*
	B: 100 to 1500, on the log scale*
	C: 100 to 10000, on the log scale*
	P: sample size chosen to attain 80 per cent power†

*For meta-analyses with sample size distributions *A*, *B* and *C*, sample size for each study was drawn from the uniform distribution on log min to log max (for example log 50 to log 500 for distribution *A*), then exponentiated.
†See Section 2.3.

Table II. Parameters for data drawn from a fixed effects model.

Parameter	Value
θ, the true odds ratio	0.5. 0.8, and 1.0
π_C, the true control rate	0.15 and 0.30

Table III. Parameters for data drawn from a random effects model.

Parameter	Value
μ_C, the mean of π_{C_i}	0.15 and 0.30
τ_C^2, the variance of π_{C_i}	0.005 for $\mu_C = 0.15$ and 0.02 for $\mu_C = 0.30$
The mean of θ_i	0.5, 0.8 and 1.0
β, the slope of the regression of log θ_i on π_{C_i}	-2 and 0
τ^2, the residual variance	0.01 and 0.15

models incorporated the effect of baseline risk on between-study variation. For the heterogeneous meta-analyses, studies were simulated to have either random sample size (Section 2.2) or sample sizes determined by power calculations (Section 2.3). The results of an extensive descriptive analysis of a database of 125 meta-analyses from leading journals and the Cochrane database [24, 25] were used to set the parameters for the simulations. Our aim was to make the simulated meta-analyses realistic by giving them the characteristics of meta-analyses normally found in practice. The parameters are provided in Tables I, II and III of this section, and the descriptive analysis is summarized in the Appendix.

At each configuration of the simulation parameters, we simulated 1000 meta-analyses without publication bias. The component studies of the meta-analyses were 2×2 tables of treatment group versus outcome, with equal sample sizes for the intervention and control groups. Table I shows the sample sizes.

2.1. Simulating homogeneous meta-analyses with fixed effects models

We chose to simulate the true odds ratio θ at three different values to represent large, moderate and no effect of treatment (Table II). We then chose two values for the true outcome rate in

the control group (control rate) π_C that frequently occur with the three chosen odds ratios in real data. The true treatment rate π_T is determined by θ and π_C. The number of events among the controls in the ith study, Y_{C_i}, was drawn from a binomial (n, π_C) distribution, where n is the number of patients per group. The number of events among the treated, Y_{T_i}, was drawn from a binomial (n, π_T) distribution. Sample size distributions A, B and C were used for data drawn using fixed effects (see Table I).

2.2. Simulating heterogeneous meta-analyses with random effects models

To simulate the heterogeneous meta-analyses, we used a published model [26] which had already been applied to the meta-analyses in the empirical database [24]. The model is hierarchical and incorporates the influence of the control rate on the outcome. It assumes the true control rate for the ith study, π_{C_i}, and the true log-odds ratio θ_i, conditioned on π_{C_i}, are normally distributed. Specifically

$$\pi_{C_i} \sim N(\mu_C, \tau_C^2)$$

and

$$\log(\theta_i|\pi_{C_i}) \sim N(\mu + \beta(\pi_{C_i} - \mu_C), \tau^2) \quad (1)$$

The parameter μ is the mean log odds ratio at the average control rate μ_C, τ_C^2 represents the variance of the control rates across studies, β is the slope of the regression of log θ_i on π_{C_i}, and τ^2 is the residual variance from the regression. When β is zero, we say there is no control rate effect, because there is no relation between the true log-odds ratio and the true control rate, although the sample log-odds ratio and sample control rate may be correlated.

The three values of θ and two values of π_C used in the simulation of homogeneous meta-analyses were used as the means of θ_i and π_{C_i} for the heterogeneous meta-analyses. The 2×2 tables were drawn using binomial distributions, as for fixed effects. The parameters of the simulations are displayed in Table III. Sample size distributions A, B and C were used.

2.3. Using power calculations to select sample size

Power considerations frequently play a role in researchers' decisions about how large a study should be. Therefore, we did additional simulations where sample size was determined by requiring 80 per cent power for the two-sided $\alpha = 0.05$ test of the null hypothesis that the true odds ratio is 1 against the alternative that the true odds ratio is θ_i. We call this method of assigning sample size 'P'. In practice, it is likely that sample size selection is neither independent of other parameters (as with A, B and C) nor determined by them (as with P), but is somewhere in between.

3. TRIM AND FILL

3.1. The trim and fill method

The idea behind the trim and fill method [20] of correcting for publication bias is to fill in the sparse corner of the funnel plot with imputed treatment effect estimates, and then to pool

Table IV. Trim and fill coverage probabilities for typical size heterogeneous meta-analyses. Standard errors range from 0.007 to 0.012.

Control rate effect	Residual variance	Mean odds ratio		
		0.5	0.8	1.0
No	Small	0.92	0.94	0.92
	Large	0.83	0.88	0.87
Yes	Small	0.94	0.95	0.93
	Large	0.89	0.87	0.90

all studies, actual and imputed. If a smaller value favours the treatment, as it does for the log-odds ratio, the trim and fill method imputes studies on the right side of the funnel plot.

First, the mean treatment effect is estimated using a random effects model. Then the studies are ranked based on their distance from this estimate. Next the number of unobserved studies is estimated as $L_0 = [4T_n - n(n+1)]/[2n-1]$, where T_n is the Wilcoxon statistic and n is the number of studies. The estimate is rounded up to L_0^+ to obtain an integer, and the mean effect is re-estimated without the L_0^+ most extreme studies on the left side of the funnel plot. All the studies are ranked again, based on their distance from the new estimate, and L_0^+ is recomputed. Iteration continues until L_0^+ stabilizes. Then L_0^+ studies are imputed on the right side of the plot by reflecting the L_0^+ studies furthest to the left around the new effect estimate (Figure 1(c)). Finally, actual and imputed studies are pooled using a random effects model. For this study we chose L_0^+ over other estimators for the number of unobserved studies because the authors indicated that L_0^+ was more robust in their simulations [20].

3.2. *Trim and fill results*

At each configuration of the simulation parameters, we calculated the coverage probability, which is the fraction of meta-analyses for which the 95 per cent confidence interval for the log-odds ratio contains the true mean log-odds ratio. Ideally, coverage probabilities should be 0.95. To economize on space, we show results only for meta-analyses that have (i) typical size: 10 studies and 50 to 500 subjects per study, (ii) large size: 25 studies and 100 to 10000 subjects per study, and (iii) 25 studies and number of subjects chosen to achieve 80 per cent power. The tables include only the results for mean control rate 0.15, because results with 0.30 were similar.

3.2.1. Homogeneous meta-analyses. The coverage probability for typical size fixed effects meta-analyses was 0.92 for odds ratio 0.5, and 0.95 for odds ratios 0.8 and 1.0. For large fixed effects meta-analyses, the coverage probability was 0.95 for all three odds ratios. The standard error, calculated as $\sqrt{\{p(1-p)/1000\}}$, was 0.007 for coverage probability $p = 0.95$ and 0.009 for $p = 0.92$.

3.2.2. Heterogeneous meta-analyses. For typical size heterogeneous meta-analyses with small residual variance ($\tau^2 = 0.01$), the results were similar to results for data generated by a fixed effects model (coverage near 0.95). The coverage probabilities were lower when the residual variance was large ($\tau^2 = 0.15$) (Table IV). For actual meta-analyses, $\tau^2 = 0.01$ is more typical

Table V. Trim and fill coverage probabilities for large heterogeneous meta-analyses. Standard errors range from 0.009 to 0.011.

Control rate effect	Residual variance	Mean odds ratio		
		0.5	0.8	1.0
No	Small	0.92	0.92	0.90
	Large	0.88	0.86	0.88
Yes	Small	0.90	0.89	0.89
	Large	0.90	0.92	0.90

Table VI. Trim and fill coverage probabilities for heterogeneous meta-analyses with sample size selected for 80 per cent power. Standard errors range from 0.010 to 0.016.

Control rate effect	Residual variance	Mean odds ratio	
		0.5	0.8
No	Small	0.81	0.42
	Large	0.27	0.12
Yes	Small	0.62	0.36
	Large	0.25	0.14

with larger odds ratios, and $\tau^2 = 0.15$ is more typical when odds ratios are around 0.5 (see Appendix). Coverage probabilities for larger meta-analyses are shown in Table V.

When there was no control rate effect, coverage probabilities differed from 0.95 because (i) trim and fill imputed studies when none was missing, and (ii) the sample log-odds ratios were not normally distributed. Chance variation also played a role. To see how much of the difference between 0.95 and the tabulated probabilities was caused by imputing studies when none was missing, we compared trim and fill to unadjusted pooling, which was affected by non-normality, but involved no imputation. When residual variance was large, the drop in coverage probability from unadjusted pooling to trim and fill was between 0.05 and 0.09 for typical size meta-analyses and between 0.04 and 0.06 for large meta-analyses. When residual variance was small, the difference between the two methods was smaller (less than 0.04 for typical size meta-analyses, and less than 0.02 for large meta-analyses).

When most of the between-study variation was due to the control rate effect (control rate effect = yes, residual variance = small), coverage probabilities were low for large meta-analyses, because estimation did not take the control rate effect into account. Schmid *et al.* [24] found that the odds ratio varies significantly with the control rate about 13 per cent of the time.

3.2.3. Heterogeneous meta-analyses, sample size adequate for 80 per cent power. The results for meta-analyses generated from the random effects models of Section 2.3 are displayed in Table VI. Here, sample size for each study was selected to attain 80 per cent power to detect the true treatment effect associated with the study. See Figure 2(*b*) for a typical example. Results are not shown for mean odds ratio 1.0, because researchers would not proceed with

a trial if they believed that the true odds ratio was greater than 1.0 (placebo or standard care superior to new treatment). Trim and fill did extremely poorly, both because it imputed studies when none were missing and because the largest weights went to studies with estimates far from the true mean odds ratio. The lowest coverage probabilities occurred when there was a large residual variance.

4. SELECTION MODELLING

4.1. The selection model

The selection process is modelled by assigning a weight to the estimated effect from each study [23]. The premise is that the probability a result is selected for inclusion in a meta-analysis depends only on the p-value. This assumption is not equivalent to the assumption made by funnel plot methods, that inclusion is based on the magnitude of treatment benefit, unless all studies have the same sample size. Selection methods estimate a pooled effect, adjusted for selection bias. A random effects model is assumed for the study effects. Each study's contribution to the likelihood is a weighted normal density.

There are several selection models in the literature [23, 27–34]. These can be categorized by: (i) the form of the weight function, which may be parametric or non-parametric; (ii) whether maximum likelihood or Bayesian estimation is used; (iii) whether covariates are incorporated in the model. The selection model [23] we chose uses maximum likelihood estimation, a non-parametric weight function, and does not incorporate covariates. The weight function on the p-values is a step function with discontinuities determined *a priori*. Because the number of studies per meta-analysis was relatively small in the simulations (10 or 25), we allowed the weight function to have only one cutpoint, at $p = 0.05$, to optimize the frequency of convergence for the algorithm. Computer code was made available by Vevea [29].

4.2. Selection model results

The selection model coverage probabilities for large heterogeneous meta-analyses and heterogeneous meta-analyses with sample size adequate for 80 per cent power are shown in Tables VII and VIII. Results for homogeneous and typical size heterogeneous meta-analyses are not shown, because in those cases the selection model estimates failed to converge most of the time. For the parameter configurations in Table VII, convergence occurred 71 to 86 per cent of the time when the residual variance was small, and over 96 per cent of the time when

Table VII. Selection model coverage probabilities for large heterogeneous meta-analyses. Standard errors range from 0.008 to 0.011.

Control rate effect	Residual variance	Mean odds ratio		
		0.5	0.8	1.0
No	Small	0.94	0.93	0.94
	Large	0.92	0.93	0.93
Yes	Small	0.88	0.88	0.88
	Large	0.92	0.93	0.94

Table VIII. Selection model coverage probabilities for heterogeneous meta-analyses with sample size selected for 80 per cent power. Standard errors range from 0.010 to 0.016.

Control rate effect	Residual variance	Mean odds ratio	
		0.5	0.8
No	small	0.94	0.74
	large	0.69	0.79
Yes	small	0.84	0.79
	large	0.70	0.82

the residual variance was large. For the parameter configurations in Table VIII, convergence occurred 81 to 98 per cent of the time, except when the odds ratio was 0.50 and there was no control rate effect and small residual variance. In that case convergence occurred 58 per cent of the time. The model's failure to converge in some cases meant that we could not evaluate it on all meta-analyses. To address the resulting potential for bias, we compared the trim and fill coverage probabilities based on all meta-analyses to trim and fill coverage probabilities based only on those meta-analyses for which selection modelling converged, and found the differences were about 0.01. There was one exception, when the odds ratio was 0.50 and there was no control rate effect and small residual variance. In that case the difference was 0.03. This analysis suggests that any bias may be small.

The coverage probabilities for large heterogeneous meta-analyses (Table VII) were generally higher than those for trim and fill (Table V), especially when there was no control rate effect and large residual variance. All numbers in Table VII were within 0.02 of the coverage probabilities for unadjusted pooling (results not shown). As with trim and fill, when most of the between-study variation was due to the control rate effect, coverage probabilities were low.

For meta-analyses with sample size selected for 80 per cent power, the selection model coverage probabilities (Table VIII) were much higher than those for trim and fill (Table VI), although they were substantially below 0.95. They were also much higher than those for unadjusted pooling, which gave the largest weights to studies with estimates far from the true mean odds ratio.

5. DISCUSSION

We found the trim and fill method of adjusting for publication bias spuriously adjusted the estimate of the global treatment effect when the studies were heterogeneous, if (i) the variability among studies caused some precisely estimated studies to have effects far from the global mean, simply due to chance, or (ii) an inverse relationship between treatment efficacy and sample size was introduced by the studies' *a priori* power calculations. In the second case, trim and fill performed extremely poorly. In both cases, performance of trim and fill was worse with larger random effects variances. The results suggest that the funnel plot itself is inappropriate for heterogeneous meta-analyses, because funnel plot interpreters may visually fill in the part of the plot where trim and fill imputes studies. Our findings support the

intention of the originators of the funnel plot that it be used only for studies from a single underlying population [17].

Selection modelling is another approach to publication bias adjustment which warrants further study. It performed much better than trim and fill in our simulations, although it frequently failed to converge. There may have been fewer numerical problems if the Newton–Raphson method in the program we used were replaced by the EM algorithm, or if a parametric weight function or Bayesian analysis were used. These are questions for further research. The selection model coverage probabilities were low when there was a control rate effect with small residual variance. Performance may be improved by incorporating the control rate effect in the selection model. Coverage probabilities were also low when sample size was selected for 80 per cent power, because in that case the within-study variance depended on the true treatment effect. This could be dealt with by modelling the within-study variance as a function of the true treatment effect, instead of assuming it is known.

Heterogeneity is common in meta-analysis [25] and examples can be found in which sample size decisions made by individual researchers may have led to funnel plot asymmetry. For example, Linde *et al.* [35] preferred selection modelling to funnel plot methods in a meta-analysis of homeopathy covering six clinical areas. As the authors explained, 'Since trials of more effective treatments need smaller sample sizes, the pattern of [the funnel] plot would mimic publication bias even when there was none.' Even when meta-analyses are focused on a specific clinical hypothesis, there may be considerable heterogeneity, and sample size requirements could vary among studies. For example, in a meta-analysis of the effect of dietary sodium reduction on blood pressure [36], the funnel plot asymmetry disappeared after controlling for intensity of the intervention [37]. This finding is consistent with researchers expecting greater benefit with more intense interventions, and selecting sample size accordingly.

There are other ways in which heterogeneity can lead to funnel plot asymmetry. Low quality studies may report exaggerated treatment benefit, and may also tend to be small. Studies of high risk patients, for whom the intervention may be most beneficial, may not be able to enroll many subjects. In either case, an inverse relation between treatment benefit and precision would result, even when there is no publication bias. Selection models which incorporate covariates could potentially adjust for quality and baseline risk.

The parameters in our simulations were carefully selected to be representative of a large database of actual meta-analyses. However, we do not know the extent to which sample sizes are affected by individual researchers' power calculations. We simulated the two extremes, where sample size is unrelated to any other characteristic of the study, and where sample size is determined in every study by the need to have 80 per cent power. In practice, power calculations may influence sample size for some studies, but not in a deterministic way.

Funnel plot shape may be influenced by factors other than publication bias and heterogeneity, including the measures of precision and effect used in the plot, and chance [38, 39]. In this study, we investigated only the impact of heterogeneity.

Our study focused on publication bias adjustment methods. Related results pertaining to the influence of heterogeneity on publication bias detection methods can be found in a recent study [40].

Estimating the number of unpublished studies is another likelihood based approach to handling publication bias [41–43]. We did not deal with it in this paper. A sensitivity approach to selection modelling has been suggested [44, 45], but because it requires visual inspection of plots, it was not included in our simulation study.

In summary, we found that the indiscriminate use of trim and fill, a publication bias method based on the funnel plot, may lead to spurious adjustment of the pooled estimate in meta-analysis, especially when there is heterogeneity among the individual studies. The results suggest that the funnel plot itself is inappropriate for heterogeneous meta-analyses. Selection modelling has the potential to perform better than trim and fill, if difficulties with convergence can be overcome.

APPENDIX: DESCRIPTIVE ANALYSIS OF AN EMPIRICAL DATABASE OF META-ANALYSES

The simulations were designed to reflect a database of 125 actual meta-analyses from seven major medical journals and the Cochrane database [25].

A1. Choosing the number of studies per meta-analysis

The median number of studies per meta-analysis in the database is ten, and only eight meta-analyses (6.4 per cent) have 25 or more studies.

A2. Choosing sample sizes for the component studies

The first three rows of Table A1 are represented in the simulations by sample size distributions *A*, *B* and *C* (see Table I).

A3. Choosing the mean control rate μ_C

The median and interquartile range of the estimated mean control rates for heterogeneous meta-analyses in our database are displayed in Table A2 for three odds ratio intervals. Meta-analyses were classified as heterogeneous if the p-value for the Q-statistic was less than 0.10. The three intervals are centred at 0.50, 0.80 and 1.0, the three choices of the mean of the odds ratio θ. Both 0.15 and 0.30, our choices for μ_C, are either within or nearly within the interquartile range of control rate for all three choices of odds ratio.

Table A1. Description of study sample sizes in the actual meta-analyses.

Number of meta-analyses	Sample size distribution within meta-analysis
54 (43%)	All studies have sample size <500.
40 (32%)	More than half of studies have sample size <500. At least one study has sample size >1000.
17 (14%)	More than half of studies have sample size >500. At least one study has sample size >1000.
14 (11%)	Some studies >500, but no studies >1000.

Table A2. Summary statistics for the mean control rate, by odds ratio interval.

Odds ratio	Median	Interquartile range
0.40 to 0.60	0.34	0.16 to 0.49
0.70 to 0.90	0.11	0.09 to 0.27
0.90 to 1.10	0.22	0.15 to 0.29

Table A3. Summary statistics for control rate variance, by control rate interval.

Control rate	Median	Interquartile range
0.10 to 0.20	0.008	0.004 to 0.010
0.25 to 0.35	0.023	0.022 to 0.024

Table A4. Summary statistics for the slope of the control rate, by odds ratio interval.

Odds ratio	Median	Interquartile range
0.40 to 0.60	−1.7	−2.7 to −0.9
0.70 to 0.90	0.0	−7.8 to 3.1
0.90 to 1.10	−1.1	−1.8 to 0.0

A4. Choosing the variance of the control rate τ_C^2

Table A3 shows the medians and interquartile ranges for estimates of the variance of the control rate in the heterogeneous meta-analyses for two control rate intervals. The two intervals are centred at 0.15 and 0.30, the choices for μ_C. The data in the table support our choices of $\tau_C^2 = 0.005$ when $\mu_C = 0.15$ and $\tau_C^2 = 0.020$ when $\mu_C = 0.30$.

A5. Choosing the slope β

Schmid *et al.* [24] have shown that the higher the control rate, the greater the treatment effect for a substantial fraction of meta-analyses. As can be seen from Table A4, the estimated slope β for regressing the log-odds ratio on control rate has considerable variability. We have chosen to simulate at $\beta = 0$ and $\beta = -2.0$. We chose $\beta = -2.0$ because it is within or nearly within the interquartile range of the estimated slopes for the three choices of odds ratio, and $\beta = 0$ because it represents no influence of control rate on the conditional mean odds ratio.

A6. Choosing the residual variance τ

Table A5 shows the median and interquartile range for estimates of τ^2, the residual variance of the log-odds ratio. Because of the wide range of possible values for τ^2, we have chosen to simulate at $\tau^2 = 0.01$ and $\tau^2 = 0.15$. Although $\tau^2 = 0.01$ is more typical for larger odds ratios, and $\tau^2 = 0.15$ is more typical for odds ratio around 0.5, we use both values of τ^2 for all three odds ratio choices, for the purpose of comparison.

Table A5. Summary statistics for the conditional variance of the log-odds ratio, by odds ratio interval.

Odds ratio	Median	Interquartile range
0.40 to 0.60	0.152	0.082 to 0.184
0.70 to 0.90	0.003	0.0005 to 0.015
0.90 to 1.10	0.015	0.006 to 0.060

REFERENCES

1. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials* 1993; Apr. 28:Doc. No. 50.
2. Stern MA, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* 1997; **315**:640–645.
3. Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Journal of the American Medical Association* 1998; **279**:281–286.
4. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Statistics in Medicine* 1987; **6**:11–29.
5. Delaney B, Moayyedi P, Deeks J. Innes M, Soo S, Barton P, Wilson S, Oakes R, Harris A, Raftery J, Hobbs R, Forman D. The management of dyspepsia: a systematic review. *Health Technology Assessment* 2000; **4**:1–189.
6. Kuznetsova T, Staessen JA, Wang JG, Gasowski J, Nikitin Y, Ryabikov A, Fagard R. Antihypertensive treatment modulates the association between the D/I ACE gene polymorphism and left ventricular hypertrophy: a meta-analysis. *Journal of Human Hypertension* 2000; **14**:447–454.
7. Priyadarshi A, Khuder SA, Schaub EA, Shrivastava S. A meta-analysis of Parkinson's disease and exposure to pesticides. *Neurotoxicology* 2000; **21**:435–440.
8. Tasche MJ, Uijen JH, Bernsen RM, de Jongste JC, van Der Wouden JC. Inhaled disodium cromoglycate (DSCG) as maintenance therapy in children with asthma: a systematic review. *Thorax* 2000; **55**:913–920.
9. Boffetta P, Silverman D, A Meta-analysis of bladder cancer and diesel exhaust exposure. *Epidemiology* 2001; **12**:125–130.
10. Oliver D, Hopper A, Seed P. Do hospital fall prevention programs work? A systematic review. *Journal of the American Geriatric Society* 2000; **48**:1679–1689.
11. McCullough ME, Hoyt WT, Larson DB, Koenig HG, Thoresen C. Religious involvement and mortality: a meta-analytic review. *Health Psychology* 2000; **19**:211–222.
12. Fleischauer AT, Arab L. Garlic and cancer: a critical review of the epidemiologic literature. *Journal of Nutrition* 2001; **131**:1032S–1040S.
13. Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K. Alchohol and coronary disease: a meta-analysis. *Addiction* 2000; **95**:1505–1523.
14. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997; **315**:629–634.
15. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal* 2000; **320**:1574–1577.
16. Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 2000; **53**:1119–1129.
17. Light RJ, Pillemer DB. *Summing Up*: *The Science of Reviewing Research*. Harvard University Press: Cambridge, MA, 1984.
18. Begg CB. Publication bias. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York, 1994.
19. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 2001; **20**:641–654.
20. Duval S, Tweedie R. A nonparametric 'Trim and Fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 2000; **95**:89–98.
21. Williams RL, Chalmers TC, Stange KC, Chalmers FT, Bowlin SJ. Use of antibiotics in preventing recurrent acute otitis media and in treating otitis media with effusion. A meta-analytic attempt to resolve the brouhaha. *Journal of the American Medical Asssociation* 1993; **270**:1344–1351.
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
23. Hedges LV. Modeling publication selection effects in meta-analysis. *Statistical Science* 1992; **7**:246–255.
24. Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine* 1998; **17**:1923–1942.

25. Engels E, Schmid C, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: and empirical study of 125 meta-analysis. *Statistics in Medicine* 2000; **19**:1707–1728.
26. McIntosh M. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* 1996; **15**:1713–1728.
27. Dear KBG, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science* 1992; **7**:237–245.
28. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. *Statistical Science* 1988; **3**:109–135.
29. Vevea JL, Hedges VH. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; **60**:419–435.
30. 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. *Statistical Science* 1997; **12**:221–250.
31. Silliman NP. Hierarchical selection models with applications in meta-analysis. *Journal of the American Statistical Association* 1997; **92**:926–936.
32. Smith DD, Givens GH, Tweedie RL. Adjustment for publication bias and quality bias in Bayesian meta-analysis. In *Meta-Analysis in Medicine and Health Policy*, Stangl DK, Berry DA (eds). Marcel Dekker: New York, 2000.
33. Patil GP, Taillie C. Probing encountered data, meta analysis and weighted distribution methods. In *Statistical Data Analysis and Inference*, Dodge Y (ed.). Elsevier Science Publishers: 1989.
34. Preston CL. Statistical models of publication bias in meta-analysis. *PhD Thesis*, Wolfson Institute of Preventive Medicine, Queen Mary and Westfield College, 2000.
35. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, Jonas WB. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997; **350**:834–843.
36. Midgley JP, Matthew AG, Greenwood CMT, Logan AG. Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. *Journal of the American Medical Asssociation* 1996; **275**: 1590–1597.
37. Greenwood CMT, Midgley JP, Mattew AG, Logan AG. Statistical issues in a metaregression analysis of randomized trials: impact on the dietary sodium intake and blood pressure relationship. *Biometrics* 1999; **55**: 630–636.
38. Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases In *Systematic Reviews in Health Care*: *Meta-Analysis in Context*, Egger M, Davey Smith G, Altman DG (eds). British Medical Journal Books: 2001.
39. Tang J-L, Liu JLY. Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology* 2000; **53**:477–484.
40. Schwarzer G, Antes G, Schumacher M. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* 2002; **21**:2465–2477.
41. Rosenthal R. The 'file-drawer problem' and tolerance for null results. *Psychological Bulletin* 1979; **86**:638–641.
42. Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. *Statistics in Medicine* 1996; **15**:2493–2507.
43. Eberly LE, Casella G. Bayesian estimation of the number of unseen studies in a meta-analysis. *Journal of Official Statistics* 1999; **15**:477–494.
44. Copas J. What works? Selectivity models and meta-analysis. *Journal of the Royal Statistical Society A* 1999; **162**:95–109.
45. Shi JQ, Copas J. Publiaction bias and meta-analysis for 2×2 tables: an average Markov chain Monte Carlo EM algorithm. *Journal of the Royal Statistical Society*, *Series B* 2002; **64**:221–236

