Notice: This Material May Be Protected by Copyright Law (Title 17 U.S. Code).

STATISTICS IN MEDICINE
Statist. Med. 2002; 21:1513–1524 (DOI: 10.1002/sim.1184)

June 15 (11)

Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research

Jonathan A. C. Sterne[1,*,†], Peter Jüni[2], Kenneth F. Schulz[3], Douglas G. Altman[4], Christopher Bartlett[1] and Matthias Egger[1]

[1] *MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, U.K.*
[2] *Department of Social and Preventive Medicine, University of Berne, Switzerland*
[3] *Family Health International and the Department of Obstetrics and Gynecology, School of Medicine, University of North Carolina at Chapel Hill, U.S.A.*
[4] *ICRF Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, U.K.*

SUMMARY

Biases in systematic reviews and meta-analyses may be examined in 'meta-epidemiological' studies, in which the influence of trial characteristics such as measures of study quality on treatment effect estimates is explored. Published studies to date have analysed data from collections of meta-analyses with binary outcomes, using logistic regression models that assume that there is no between- or within-meta-analysis heterogeneity. Using data from a study of publication bias (39 meta-analyses, 394 published and 88 unpublished trials) and language bias (29 meta-analyses, 297 English language trials and 52 non-English language trials), we compare results from logistic regression models, with and without robust standard errors to allow for clustering on meta-analysis, with results using a 'meta-meta-analytic' approach that can allow for between- and within-meta-analysis heterogeneity. We also consider how to allow for the confounding effects of different trial characteristics. We show that both within- and between meta-analysis heterogeneity may be of importance in the analysis of meta-epidemiological studies, and that confounding exists between the effects of publication status and trial quality. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; clinical trials; publication bias; language bias; trial quality

1. INTRODUCTION

Systematic reviews and meta-analyses of randomized trials provide the best evidence on the effects of medical interventions but they are not immune to bias [1]. Studies that find a

*Correspondence to: Jonathan Sterne, Department of Social Medicine, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, U.K.
†E-mail: jonathan.sterne@bristol.ac.uk

Contract/grant sponsor: U.K. National Health Service Health Technology Assessment Programme; contract/grant number: 97/18/05

Copyright © 2002 John Wiley & Sons, Ltd.

2

statistically significant effect of treatment are more likely to be published [2–4], published in English [5], cited by other authors [6] and to produce multiple publications [7, 8] than other studies. Such studies are also more likely to be identified and included in systematic reviews, which may therefore over estimate the beneficial effects of treatment [1]. The often-inadequate methodological quality of component trials is another important possible source of bias in systematic reviews. Clearly, if the 'raw material' is flawed, then the findings of reviews of this material may also be compromised [1].

Such biases may be examined by considering collections of meta-analyses in which component trials are classified according to characteristics such as study quality. For example, the study by Schulz *et al.* [9] demonstrated, for trials with binary outcomes included in meta-analyses from the Cochrane Pregnancy and Childbirth Database, that trials in which randomization was inadequately concealed or inadequately reported yielded exaggerated estimates of treatment effect in comparison with trials reporting adequate concealment, and found a similar (but smaller) effect for trials which were not double-blind. These observations have been replicated in other studies [10]. To ensure that treatment effects were compared only between studies in the same meta-analysis, these studies used a logistic regression approach in which the evidence for an interaction between the effects of trial quality and treatment group is examined, having controlled for the interaction between meta-analysis and treatment group. The analysis assumes that the effect of bias is constant across meta-analyses; if this assumption is false then standard errors of estimated differences will be too small. Further, the approach is computationally cumbersome as it requires that a large number of indicator variables be introduced into the model. Finally, meta-analyses based on treatment effect estimates other than odds ratios must be converted to odds ratios. This may not always be possible using published data.

In this paper we review methods for assessing the influence of trial characteristics on estimated treatment effects in data sets containing collections of meta-analyses, using a case study of the influence of unpublished trials and trials published in languages other than English on treatment effect estimates from meta-analyses. In Section 2 we introduce the data set that motivates the paper. In Section 3 we present the results of analyses using the fixed-effect logistic regression approach used in meta-epidemiological studies to date. In Section 4 we describe an alternative approach in which we combine estimated effects in a 'meta-meta-analysis'. In Section 5 we consider how to allow for the confounding effects of different trial characteristics.

2. DATA ON META-ANALYSES USING COMPREHENSIVE LITERATURE SEARCHES

To prevent publication and language biases, many systematic reviewers perform extensive literature searches which cover all relevant languages and, in addition to the major bibliographic databases, 'grey' literature such as conference abstracts, theses, book chapters or unpublished manuscripts. Such searches are time consuming, require the translation of articles and the contacting of experts, original investigators and pharmaceutical companies and thus add considerably to the costs of performing a review. Although excluding unpublished trials, or trials reported in languages other than English, may introduce bias and also reduce the precision of estimates of treatment effects, the importance and direction of these effects is unclear at present.

Copyright © 2002 John Wiley & Sons, Ltd.

3

The Research and Development Programme of the U.K. National Health Service commissioned research to clarify the importance of comprehensive literature searches. Briefly, we identified state-of-the-art meta-analyses that were based on comprehensive literature searches, in order to examine the contributions made by trials that were unpublished or published in languages other than English. A comprehensive literature search was defined as a search not restricted to the English-language literature, which covered (i) either the *Cochrane Controlled Trials Register* or at least two other electronic databases (such as MEDLINE or *Embase*) and (ii) at least one other source (for example, a search for unpublished material, a search of conference abstracts, theses or other grey literature, or a manual search of journals). Trials were classified as journal reports if they were published as full or short reports, editorials or letters in a regular issue or supplement of a journal. All other reports were classified as grey literature. We assessed language of publication for journal articles only. For each trial in each meta-analysis, we recorded the number of patients and events in each group, and used these to derive the log-odds ratio and standard error for the treatment effect. To obtain consistency across meta-analyses, endpoints were recoded if necessary, so that odds ratios below 1 indicated a beneficial effect of the intervention.

The data set analysed in Sections 3, 4 and 5 is restricted to 122 meta-analyses containing 1134 trials which employed comprehensive literature searches and were published in issue 1/1998 of the *Cochrane Database of Systematic Reviews* [11]. The median number of trials per meta-analysis was 7 (interquartile range 6 to 11). A total of 39 (32 per cent) meta-analyses contained unpublished trials (394 published and 88 unpublished trials) and 29 (24 per cent) contained trials published in languages other than English (297 English language trials and 52 non-English language trials). Treatment effects were measured as odds ratios in 105 (86 per cent) meta-analyses, and were mainly analysed using the Peto method [12]. The authors of only five (4 per cent) meta-analyses reported results from random-effects models.

3. ANALYSES USING FIXED-EFFECT LOGISTIC REGRESSION

Previous studies [9, 13, 14] have used standard logistic regression models to examine the effect of measures of trial quality on treatment effect estimates. In this section we use such models to estimate whether estimated treatment effects differ according to whether a trial is published or, where published, its language of publication. Suppose that we have data from M meta-analyses, containing a total of S studies. To estimate the effect of a binary study characteristic C (for example $C=1$ in published trials, 0 in unpublished trials) on estimated treatment effects we fit the model

$$\text{logit}(\pi) = \beta_0 + \beta_1 I_t + \beta_2 I_{tc} + \sum_{i=2}^{M} \gamma_i I_{tm_i} + \sum_{j=2}^{S} \delta_j I_{s_j}$$

where π is the probability that an (adverse) outcome event is observed, I_t, I_{tc}, $\{I_{tm_i}\}$ and $\{I_{s_j}\}$ are all indicator variables denoting, respectively, the effects of treatment ($I_t = 1$ in treated individuals, 0 otherwise) the treatment-characteristic interaction ($I_{tc} = 1$ in treated individuals in trials with $C = 1$, 0 otherwise), the treatment-meta-analysis interactions ($I_{tm_i} = 1$ in treated individuals in meta-analysis i, 0 otherwise) and study numbers ($I_{s_j} = 1$ for individuals in study j, 0 otherwise) and $\{\beta\}$, $\{\gamma\}$ and $\{\delta\}$ are the parameters of the logistic regression model. This

Copyright © 2002 John Wiley & Sons, Ltd.

4

Table I. Effects of publication status and language of trials on treatment effect estimates, using fixed-effect logistic regression models. Analyses based on 39 meta-analyses and 482 trials (394 unpublished versus 88 published) and 29 meta-analyses and 349 trials (52 non-English versus 297 English). OR = odds ratio.

	Unpublished versus published		Non-English versus English	
	Ratio of ORs (95% CI)	*P*	Ratio of ORs (95% CI)	*P*
Model-based standard errors	1.12 (0.99 to 1.26)	0.068	0.78 (0.67 to 0.91)	0.001
Robust standard errors	1.12 (0.95 to 1.31)	0.18	0.78 (0.62 to 0.98)	0.036

model allows the probability of the outcome event to vary according to treatment group, trial characteristic and trial, while the interaction terms $\{I_{\mathrm{tm}_i}\}$ mean that the effect of treatment is estimated separately in each meta-analysis. The estimated effect of the characteristic C on average treatment effects is then given by parameter β_2 (for the treatment-characteristic interaction), which estimates the log of the ratio of treatment odds ratios (ROR) in trials with and without the characteristic. This is assumed to be constant across meta-analyses. Note that only meta-analyses that contain trials with and without the characteristic contribute to this estimate. To ensure that RORs were estimable using logistic regression, meta-analyses were only included in the analyses presented here if at least one event was reported in both the treatment and control groups, after combining data from all trials with and all trials without the characteristic.

Table I (row 1) shows the estimated associations between treatment effects and publication status or language of publication. Since outcomes were adverse, treatment odds ratios less than one correspond to beneficial effects and so RORs less than one imply greater (more beneficial) treatment effects in the comparison group than in the reference group. Consistent with current understanding of mechanisms of publication bias, estimated treatment effects were less beneficial in unpublished trials (ROR 1.12, 95 per cent CI 0.99 to 1.26). Treatment effects were more beneficial in trials published in languages other than English (ROR 0.78, 95 per cent CI 0.67 to 0.91).

3.1. Use of robust standard errors

A simple way of examining the assumption that RORs are constant across meta-analyses is via 'robust' standard errors, which use the 'information sandwich' [15, 16] to estimate standard errors based on the regression residuals. Robust standard errors may also be estimated after allowing for clustering [17], providing that the number of clusters is at least 20 [18]. Table I (row 2) shows 95 per cent confidence intervals and Wald *P*-values derived using robust standard errors allowing for clustering on meta-analysis. The increased width of confidence intervals derived using robust standard errors suggests that RORs may differ between meta-analyses. Note that this approach corrects only the standard errors – parameter estimates are not affected.

With the logistic regression approach a total of $M + S + 1$ parameters must be estimated, where M is the number of meta-analyses and S is the number of studies. For example, the data on effect of publication status are derived from 39 meta-analyses containing 482 studies, so that a total of 522 parameters were estimated. Such numbers of parameters approach the limits of the dimensions of the design matrix, even for modern statistical software (for example,

Copyright © 2002 John Wiley & Sons, Ltd.

5

Table II. Effects of publication status and language of trials on treatment effect estimates, using fixed- and random-effects meta-analysis. Analyses based on 39 meta-analyses and 482 trials (394 unpublished versus 88 published) and 29 meta-analyses and 349 trials (52 non-English versus 297 English).

Within meta-analysis	Between meta-analysis: Fixed effect		Random effects			χ^2 statistic and P-value for between-meta-analysis heterogeneity
	Ratio of ORs (95% CI)	P	Ratio of ORs (95% CI)	P	$\hat{\tau}^{2*}$	
Unpublished versus published						
1. Fixed effect	1.12 (1.00 to 1.27)	0.056	1.13 (0.99 to 1.28)	0.062	0.013	$\chi^2=41.3$, 38 d.f., $P=0.33$
2. Random effects	1.15 (1.01 to 1.31)	0.036	1.14 (1.00 to 1.30)	0.59	0	$\chi^2=27.6$, 38 d.f., $P=0.89$
Non-English versus English						
3. Fixed effect	0.79 (0.68 to 0.92)	0.002	0.81 (0.65 to 1.00)	0.052	0.107	$\chi^2=44.4$, 28 d.f., $P=0.025$
4. Random effects	0.80 (0.67 to 0.96)	0.017	0.80 (0.67 to 0.96)	0.017	0	$\chi^2=27.7$, 28 d.f., $P=0.48$

*$\hat{\tau}^2$ is the DerSimonian and Laird estimate [19] of the between-meta-analysis variance in RORs.

the current maximum matrix size in Stata is 800). These difficulties are avoided by using the approach described in the next section.

4. ANALYSES USING A META-ANALYTIC APPROACH

Of the parameters of the logistic regression model described in Section 3, only the effect of the trial characteristic (β_2) is assumed to be constant across meta-analyses. An obvious alternative to this approach is to estimate the effect of the characteristic using a separate logistic regression in each meta-analysis (that is, fixed-effect within meta-analyses). Estimated RORs in each meta-analysis can then be combined using meta-analytic methods, using inverse-variance weighting and either fixed-effect or random-effects between meta-analyses. Random-effects analyses used the moment-based variance estimator ($\hat{\tau}^2$) proposed by DerSimonian and Laird [19]. Results of such 'meta-meta-analyses' are shown in Table II, rows 1 and 3. As expected, the fixed-effect analysis gives very similar results to the logistic regression method (see Table I, column 1).

If the effect of a trial characteristic varies between meta-analyses, then analyses based on the logistic regression approach (Section 2) will underestimate the uncertainty in estimated RORs. There was modest between-meta-analysis heterogeneity in the effect of publication status ($\hat{\tau}^2=0.013$) but more substantial heterogeneity in the effect of language of publication ($\hat{\tau}^2=0.107$), in the analyses using fixed-effects within meta-analyses. Table II shows that when estimates derived assuming fixed-effects within meta-analyses (rows 1 and 3) are combined using random-effects estimates, CIs are similar to those derived using robust standard errors in Table I. Figures 1 and 2 display the estimated differences in each meta-analysis (estimated

Copyright © 2002 John Wiley & Sons, Ltd.

6

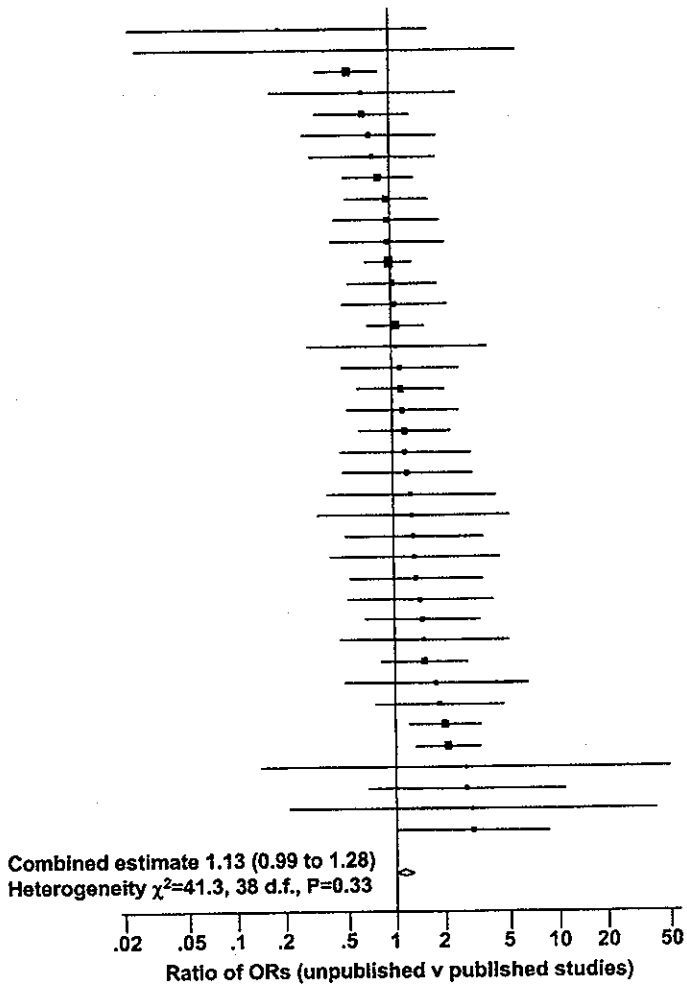

Figure 1. Ratios of ORs comparing unpublished with published trials in 39 meta-analyses published in the Cochrane Database of Systematic Reviews, together with the combined estimate of the ratio of ORs, derived using random-effects meta-analysis. Ratios of ORs greater than 1 imply that treatment effects were less beneficial in unpublished trials.

using logistic regression), together with the combined estimates (derived using random effects between meta-analyses).

The fixed-effect assumption may also be violated within-meta-analyses. This can be addressed by using random-effects meta-regression to allow for between-trial (within-meta-analysis) heterogeneity [20, 21]. Meta-regression examines associations between the estimated treatment effect (log OR) in each trial and one or more trial characteristics, allowing appropriately for the precision of the treatment effect via the standard error of the log OR in each trial. For a single meta-analysis, meta-regression estimates the same quantity (the ratio of ORs comparing trials with and without the characteristic) as is estimated using the logistic

Copyright © 2002 John Wiley & Sons, Ltd.

1

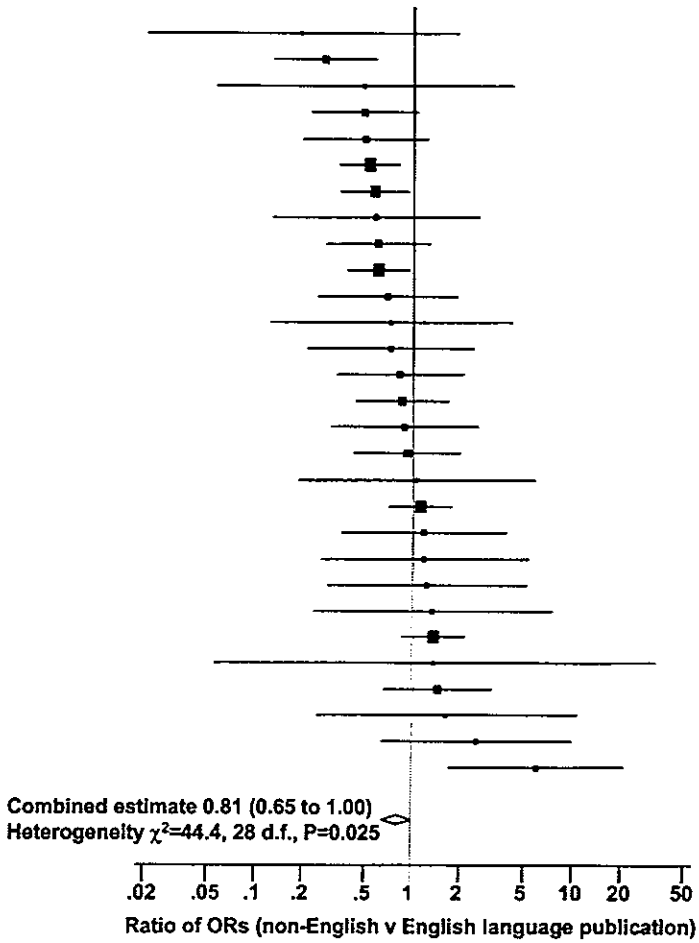

Figure 2. Ratios of ORs comparing trials published in a non-English language with trials published in English in 29 meta-analyses, published in the Cochrane Database of Systematic Reviews, together with the combined estimate of the ratio of ORs, derived using random-effects meta-analysis. Ratios of ORs less than 1 imply that treatment effects were more beneficial in trials published in a non-English language.

regression approach. In the absence of within-meta-analysis heterogeneity the ratios of ORs estimated using meta-regression and logistic regression will be similar.

Table II (rows 2 and 4) shows the effect of publication status and language of publication, estimated using meta-regression separately for each meta-analysis (random effects within meta-analyses, using restricted maximum-likelihood estimation), then combined using both fixed and random effects between-meta-analyses. Compared to analyses using fixed effects within meta-analyses, ratios of ORs and CIs were similar but there was less evidence of between-meta-analysis heterogeneity. For both publication status and language of publication, the estimated between-meta-analysis variance was zero.

Copyright © 2002 John Wiley & Sons, Ltd.

8

Table III. Effects of publication status and language of publication on treatment effect estimates, after controlling for trial quality. Analyses based on 39 meta-analyses and 482 trials (394 unpublished versus 88 published) and 29 meta-analyses and 349 trials (52 non-English versus 297 English).

	Unpublished versus published*			Non-English versus English*		
	Ratio of ORs (95% CI)	*P*	Heterogeneity *P*-value	Ratio of ORs (95% CI)	*P*	Heterogeneity *P*-value
Logistic regression						
Model-based standard errors	1.14 (1.01 to 1.28)	0.035		0.81 (0.70 to 0.95)	0.01	
Robust standard errors	1.14 (0.97 to 1.34)	0.12		0.81 (0.63 to 1.05)	0.12	
Meta-analytic approach, within-meta analysis differences estimated using logistic regression						
Fixed effect between meta-analyses	1.18 (1.03 to 1.34)	0.015		0.82 (0.68 to 0.98)	0.026	
Random effects between meta-analyses	1.21 (1.02 to 1.43)	0.029	0.084	0.88 (0.66 to 1.18)	0.40	0.003
Meta-analytic approach, within-meta analysis differences estimated using random-effects meta-regression						
Fixed effect between meta-analyses	1.18 (1.02 to 1.36)	0.028		0.76 (0.62 to 0.94)	0.011	
Random effects between meta-analyses	1.18 (1.02 to 1.38)	0.031	0.37	0.79 (0.62 to 1.02)	0.068	0.22

*Controlling for concealment of allocation and blinding.

5. CONFOUNDING

The different biases which affect systematic reviews and meta-analyses are unlikely to operate independently. For example, publication bias may lead to treatment effect estimates being smaller in unpublished trials, but such trials may also tend to be of lower methodological quality and therefore to overestimate treatment effects. The logistic regression approach discussed in Section 3 could obviously be used to control for such confounding by including the effects of more than one trial characteristic in the model. However this assumes that the effect of each characteristic is constant across meta-analyses. An alternative approach is to estimate effects controlling for confounding factors separately in each meta-analysis, using random-effects meta-regression, then combine these in the same way as in Section 4.

Information on concealment of allocation and blinding in each trial, provided by authors of meta-analyses, was extracted from the text of each Cochrane review but without referring to the original trial reports. For concealment of allocation we categorized trials as adequately concealed (central randomization, coded drug packs, assignment envelopes etc.), inadequately or unclearly concealed trials which either reported an inadequate approach (alternation, open random number tables etc.) or lacked a statement on concealment, or as unknown if the information was not reported in the review. For blinding we categorized trials as double-blind or assessor-blind, neither double-blind nor assessor-blind, or as unknown.

Table III shows estimated ratios of ORs for publication status, estimated first using fixed-effect logistic regression, secondly by combining the results of separate logistic regression

Copyright © 2002 John Wiley & Sons, Ltd.

-9-

analyses and thirdly by combining the results of separate meta-regression analyses. Differences between unpublished and published trials increased after controlling for trial quality, whereas differences between English language trials and trials published in languages other than English generally decreased.

6. DISCUSSION

Little consideration has been given to appropriate statistical methods for assessing the effects of trial characteristics such as methodological quality, publication status and language of publication in 'meta-epidemiological' research. We have compared results from fixed-effect logistic regression models, used in all previous meta-epidemiological studies, with a meta-analytic approach in which the effects of trial characteristics are estimated separately in each meta-analysis and then combined using standard meta-analytic methods. As expected, analyses using logistic regression give similar results to the meta-analytic approach assuming fixed effects. However our results suggest that such fixed-effect analyses are likely to underestimate standard errors because of the presence of within- and between-meta-analysis heterogeneity.

Although the effects of publication bias and poor methodological quality may lead to substantial overestimates of the beneficial effects of medical interventions, these effects cannot be estimated precisely in individual meta-analyses, which typically contain only small numbers of trials [22]. The imprecision of estimates from individual meta-analyses (see Figures 1 and 2) means that it is necessary to combine evidence from substantial numbers of meta-analyses to estimate the effects of factors such as publication status or language of publication on treatment effect estimates. The need to include an indicator variable for each trial and each meta-analysis in the data set means that estimation in the logistic regression models is slow and that limits on the dimension of the design matrix permitted in statistical packages may be reached. A 'meta-meta-analytic' approach, in which the effects of trial characteristics are estimated separately in each meta-analysis and then combined using standard meta-analytic techniques, avoids these computational difficulties and provides a natural way to examine the evidence for between-meta-analysis heterogeneity and possible explanations for this. For example, associations with trial quality may be larger for interventions in which subjectively assessed outcomes are measured, or in which there is a strong desire among investigators that a treatment be proved effective. Improved understanding of the circumstances in which bias is likely to undermine a systematic review should inform attempts to prevent bias in the future.

The analyses presented here suggest that previous meta-epidemiological studies may have overstated the evidence that trial quality is associated with treatment effects. We investigated this possibility for the study of Schulz *et al.* [9]. This study, of 250 trials contained in 33 meta-analyses, was the first to show that dimensions of trial quality, in particular the adequacy of allocation concealment, were associated with treatment effect estimates. It has been influential in motivating efforts to improve the quality of methods and reporting of randomised controlled trials [23, 24]. Schulz *et al.* briefly addressed the possibility of between-trial heterogeneity, using a quasi-likelihood approach to take account of overdispersion (extra-binomial variation). Their conclusions remained unchanged, although confidence intervals were somewhat wider [9].

Copyright © 2002 John Wiley & Sons, Ltd.

10

Table IV. Effects of dimensions of methodological quality on treatment effect estimates, using fixed-effect logistic regression models and fixed and random-effects meta-analysis. Data (33 meta-analyses containing 250 trials) from Schulz *et al.* [9].

	Concealed allocation versus other			Double-blind versus other		
	Ratio of ORs (95% CI)	*P*	Heterogeneity *P*-value	Ratio of ORs (95% CI)	*P*	Heterogeneity *P*-value
Logistic regression						
Model-based standard errors	0.66 (0.59 to 0.73)	<0.001		0.68 (0.60 to 0.78)	<0.001	
Robust standard errors	0.66 (0.55 to 0.78)	<0.001		0.68 (0.55 to 0.84)	<0.001	
Meta-analytic approach, within-meta analysis differences estimated using logistic regression						
Fixed effect	0.66 (0.60 to 0.74)	<0.001	0.005	0.68 (0.60 to 0.78)	<0.001	0.0001
Random effects	0.67 (0.57 to 0.78)	<0.001		0.67 (0.54 to 0.82)	<0.001	
Meta-analytic approach, within-meta analysis differences estimated using random-effects meta-regression						
Fixed effect	0.71 (0.63 to 0.80)	<0.001	0.19	0.72 (0.61 to 0.85)	<0.001	0.104
Random effects	0.67 (0.61 to 0.82)	<0.001		0.71 (0.57 to 0.87)	0.001	

In our reanalyses of these data, we have combined the 'unclearly concealed' and 'inadequately concealed' allocation categories used in the original paper. The results of these analyses, summarized in Table IV, show that there was evidence of some between-trial heterogeneity, so that confidence intervals that allow for this are wider than in the fixed-effect analyses reported in the original paper [9]. However, analyses that allow for between and within meta-analysis heterogeneity still show clear evidence of associations between trial quality and treatment effect estimates. Note that the trials analysed in this study were exclusively from one area of medicine (perinatal medicine), which may explain the relatively small amount of between-meta-analysis heterogeneity. Further reanalyses of previously published empirical studies [13, 14, 25] are planned (D. Moher, personal communication).

There is long-standing discussion in the literature on the merits and disadvantages of random-effects meta-analyses, which allow for between-trial heterogeneity in treatment effects. In recent years a consensus has emerged that in the presence of substantial heterogeneity the emphasis should not be on combining individual studies but on exploring reasons for between-study heterogeneity. We do not, in general, advocate analyses of meta-epidemiological studies using random-effects within *all* meta-analyses, for two main reasons. First, random-effects analyses will weight studies more equally in the presence of between-study variability. This is undesirable if the smaller studies are more prone to bias. Second, the numerator in the DerSimonian and Laird moment estimator of the between-study variance is the maximum of 0 and the χ^2 heterogeneity statistic Q minus its expected value (k-1, where k is the number of studies in the meta-analysis) [19]. If the true value of the between-study variance τ^2 is small there is a high probability (approaching 0.5 as the number of studies increases) that $Q < k - 1$ and hence that the estimated between-study variance $\hat{\tau}^2$ will be zero ($\approx \tau^2$). The rest of the time, $\hat{\tau}^2 \geqslant \tau^2$ so that, over many such meta-analyses, heterogeneity will tend to be overestimated. This may be a particular issue in meta-analyses that include small numbers

Copyright © 2002 John Wiley & Sons, Ltd.

11

of trials, so that the sampling variation in $\hat{\tau}^2$ is substantial. A possible solution is to use a two-stage strategy, using random-effects analyses in those meta-analyses for which the P-value for between-study heterogeneity is less than some threshold such as 0.1.

Two alternatives to the analyses presented here would be either to extend the logistic regression approach to allow for random effects between or within meta-analyses, or to use a multilevel approach based on summary treatment effect estimates from each trial [26]. These could have particular advantages in analyses that try to control for confounding effects of other trial characteristics, since the estimated effect of a characteristic in a particular meta-analysis would be shrunk towards the overall mean by an amount depending on the amount of information in the data for that meta-analysis. In analyses using the meta-analytic approach we have seen that the effect of trial characteristics is estimated very imprecisely, which will make it difficult to control for confounding effects.

In the analyses here we have ignored the analytic approach used by the authors of the original meta-analyses. For instance, those studies that use random-effects models may do so because of unexplained between-trial heterogeneity, and outcome measures (for example, odds ratios, risk ratios or hazard ratios) may also be selected because they are most appropriate in a particular context. Meta-analyses which report only summary statistics cannot be included in the logistic regression analyses; this includes meta-analyses using hazard ratios, which must be based on individual patient data. The numbers of patients and events in each trial were available for all Cochrane meta-analyses because these have to be entered in the Cochrane Collaboration's Review Manager software. It is common, however, that published meta-analyses do not provide this information. An alternative application of the meta-analytic approach is to estimate the between-trial comparisons of interest (for example, of publication status or language of publication) using the methods of the original meta-analysis. For example, for a meta-analysis based on risk ratios, we would estimate the ratio of the treatment risk ratio in unpublished trials compared to the published trials, using the estimation method of the original meta-analysis. These ratios can then be combined across meta-analyses using standard DerSimonian and Laird random-effects models. We took this approach in an analysis of the direction and impact of language bias using the data set on which the present study is based [27].

Kunz and Oxman [28] drew attention to the fact that bias in randomized trial research can go in either direction, masking effects or reversing the direction of effect, and argued that the effects of bias are ultimately unpredictable, although consistent findings as to the relevant importance of different types of bias are emerging when large numbers of meta-analyses are aggregated [10]. The methods discussed here should assist future researchers to adequately deal with unexplained and unpredictable heterogeneity and, more importantly, to develop a refined understanding of the mechanisms that introduce bias in different situations, thus hopefully reducing the unpredictability of bias effects in clinical trial research.

ACKNOWLEDGEMENTS

We thank Jos Kleijnen, Mark Starr, Carol Lefebvre, Guido Schwarzer, Deborah Tallon, Joanna Wardlaw and Lesley Stewart for help with assembling the data analysed in this paper. We also thank an anonymous referee for a number of helpful comments. PJ was funded by the Swiss National Science Foundation. The work was supported by the U.K. National Health Service Health Technology Assessment Programme (project number: 97/18/05). The views and opinions expressed are those of the authors and do not necessarily reflect those of the Department of Health.

Copyright © 2002 John Wiley & Sons, Ltd.

12

REFERENCES

1. Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In *Systematic Reviews in Health Care, Meta-Analysis in Context*, Egger M, Davey Smith G, Altman DG (eds). BMJ Books: London, 2000.
2. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**:867–872.
3. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *Journal of the American Medical Association* 1992; **263**:374–378.
4. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* 1997; **315**:640–645.
5. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; **350**:326–329.
6. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *British Medical Journal* 1992; **305**:15–19.
7. Tramer MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *British Medical Journal* 1997; **315**:635–640.
8. Gøtzsche PC. Multiple publication of reports of drug trials. *European Journal of Clinical Pharmacology* 1989; **36**:429–432.
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. *Journal of the American Medical Association* 1995; **273**:408–412.
10. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *British Medical Journal* 2001; **323**:42–46.
11. The Cochrane Database of Systematic Reviews. The Cochrane Library. Update Software, Oxford, 1998.
12. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Disease* 1985; **27**:335–371.
13. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**:609–613.
14. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; **356**:1228–1231.
15. Huber PJ. The behaviour of maximum likelihood estimates under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California: Berkeley, CA, 1967;221–233.
16. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980; **48**:817–830.
17. Rogers WH. sg17: Regression standard errors in clustered samples. *Stata Technical Bulletin* 1993; **3**:88–94.
18. Donner A. Some aspects of the design and analysis of cluster randomization trials. *Applied Statistics* 1998; **47**:95–113.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
20. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
21. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
22. 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.
23. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Journal of the American Medical Association* 1996; **276**:637–639.
24. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gøtzsche PC, Lang T. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine* 2001; **134**:663–694.
25. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology* 2000; **53**:964–972.
26. Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; **19**:3417–3432.
27. Jüni P, Holenstein F, Sterne JA, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology* 2002 (in press).
28. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *British Medical Journal* 1998; **317**:1185–1190.

Copyright © 2002 John Wiley & Sons, Ltd.

B