J. R. Statist. Soc. A (2005)
168, *Part* 4, *pp.* 737–752

Modelling reporting bias: the operative mortality rate for ruptured abdominal aortic aneurysm repair

Dan Jackson,

University of Salford, UK

John Copas

University of Warwick, Coventry, UK

and Alex J. Sutton

University of Leicester, UK

[Received May 2004. Revised February 2005]

Summary. It is perhaps underappreciated that ruptured abdominal aortic aneurysm is a significant cause of mortality in the UK. The only curative treatment is an emergency operation and quantifying the success of this presents many difficulties. In particular, there is empirical evidence of reporting bias, suggesting that studies failing to report operating theatre mortality may be those where death in theatre is more common. We suggest a procedure for correcting for this bias and re-examine a recent meta-analysis of the available data. This casts considerable doubt on some conclusions from naïve analyses that do not take into account the potential bias. Perhaps most importantly, our procedure indicates a modest improvement in operating theatre mortality over the last 50 years, which is a trend that is not evident from the usual naïve analyses.

Keywords: Meta-analysis; Meta-regression; Non-ignorable selection; Reporting bias

1. Introduction

A meta-analysis of 50 years of ruptured abdominal aortic aneurysm (RAAA) repair has recently been performed (Bown *et al.*, 2002). This thorough review of the literature involved 171 studies and over 21 000 patients. It is perhaps underappreciated that RAAA is a significant cause of mortality in the UK; in 1989 it was reported that the number of deaths due to this in England and Wales was approximately 7500 per annum, and that this annual figure is now likely to be considerably higher (Bown *et al.*, 2002). The only curative treatment is an emergency operation, which has a high mortality rate, and quantifying the success of this is clearly of considerable importance. Bown *et al*. (2002) used meta-regression for this and concluded that the overall survival rate is improving over time. The mid-date is provided for each study, along with the number of patients and overall mortality, so that this type of analysis can be performed. Some studies also provide details of the number of deaths that occurred in the operating theatre (77 studies).

Although this analysis is of considerable interest, it is questionable whether the studies' results should be pooled in this way. In particular it should be noted that they use different definitions

Address for correspondence: Dan Jackson, Centre for Operational Research and Applied Statistics, Maxwell Building, University of Salford, Salford, M5 4WT, UK.
E-mail: d.jackson@salford.ac.uk

 0964–1998/05/168737

of overall mortality. Some define this as death in hospital and others use 30-day mortality (a few provide details of both of these; in such instances death in hospital is used in all the analysis that follows) whereas the rest fail to provide precise details of their definition. Although Bown *et al*. (2002) obtained similar conclusions using studies with a variety of definitions, this lack of consistency is far from desirable. It is therefore perhaps more meaningful to examine mortality in the operating theatre. By its very nature, this is clearly defined for all 77 studies that provide details of this. However, this also presents difficulties, as noted by Bown *et al*. (2000), due to the empirical evidence of reporting bias, suggesting that studies failing to report operating theatre mortality may be those where death in theatre is more common. This is an example of a general problem in which there are two events of interest (death in the operating theatre and overall mortality) which some studies effectively aggregate. Another medical example of this type of problem is provided by cardiology, where total coronary heart disease events (both fatal and non-fatal combined) may be provided but some studies may not distinguish between the two.

A taxonomy of the types of data which potentially could be missing from a study report, but required for a meta-analysis, is described by Sutton and Pigott (2005). Some such reporting shortfalls will be innocent and it may be reasonable to assume that biases are not present. If so, standard missing data methods could potentially be used to address such missingness (Little and Rubin, 2002). However, the assumption of no bias will not always be satisfied as less innocent data suppression mechanisms may be acting and hence such an assumption will be dangerous and produce misleading inferences.

A range of methods to address biases in meta-analysis have been described in the literature. These range from simple graphical plots, such as funnel plots (Sterne and Egger, 2001), through asymmetry-type tests based on similar premises to those of the funnel plot (Egger *et al.*, 1997), to modelling methods which attempt to adjust the analyses for biased sampling of studies (Sutton *et al.*, 2000; Copas and Shi, 2001). The majority of the existing methods primarily consider the suppression of whole studies, although sensitivity analysis methods to deal with incomplete reporting by outcome (Hutton and Williamson, 2000) and subgroup (Hahn *et al.*, 2000) have recently been considered. Such methods examine the robustness of meta-analysis results to potential extreme reporting bias scenarios which assume that only the significant fraction of outcomes, or only the most significant subgroup effects, are reported from each study. Although such methods are helpful, and can indicate whether the effort that is required to contact authors of the original studies for more details and possibly for their individual patient data will be worthwhile, they essentially assess the effect of the 'worst case scenario' and give little indication of the most likely effect of reporting biases.

We suggest a procedure which corrects for the bias in the reporting of operating theatre mortalities for the RAAA data set, which essentially imputes values of the event of interest but also allows for the possibility of reporting bias. The remainder of this paper is structured as follows. Section 2 describes the RAAA data set in detail and results of 'naïve' analyses ignoring the reporting bias problem. Funnel plots are used to assess informally whether biases are present and this raises questions concerning the validity of naïve inferences concerning the success of the emergency operation. Section 3 outlines our main model, which enables us to take into account the possibility of bias, and Section 4 describes how this is fitted to the RAAA data. Although there is some evidence of reporting bias, quantifying the severity of this presents difficulties, so analyses are carried out both with and without this type of bias. These results are then compared with the naïve analyses of Section 2, providing some contrasting results depending on the assumptions that are made. Since our procedure does not model the unconditional probability of death in theatre directly, Section 5 provides an approach for interpreting the results

in terms of this. Although we find some broad agreement between our model and simple naïve approaches, contrasting results are obtained concerning the time trend that is associated with the success of the emergency operation. Data imputation is used to explore this surprising and important finding. Section 6, the discussion, concludes the paper.

2. The ruptured abdominal aortic aneurysm repair meta-analysis

We re-examine the meta-analysis of Bown *et al.* (2002). Two main outcomes are of interest: mortality during surgery and the overall mortality of RAAA patients including the subsequent period in hospital. Both of these outcomes were assessed in the meta-analysis of Bown *et al.* (2002). Firstly, a meta-regression of overall mortality shows a small improvement over time. Explicitly, the model is $y_i \sim N(\alpha + \beta t_i, \sigma_i^2 + \tau^2)$, where y_i is the estimated log-odds of overall mortality in the ith study, t_i is the mid-date of the ith study centred at 1980, σ_i^2 is the within-study variance of y_i (which is assumed known but estimated in practice) and τ^2 is the between-study variance (random-effects variance). Conventionally τ^2 is estimated and then assumed known. Estimating τ^2 by maximum likelihood (effectively using method 3b of Thompson and Sharp (1999)) provides $\tau^2 = 0.21$. Estimates of α and β in this model are -0.07 and -0.014, with corresponding 95% confidence intervals $(-0.15, 0.01)$ and $(-0.022, -0.006)$. The meta-regression indicates a statistically significant decrease in mortality from about 59% in 1950 to 41% in 2000.

A useful diagnostic tool for assessing biases in meta-analysis is the funnel plot. Although variations exist, the convention that is adopted here is to plot study precision, defined to be the reciprocal of the within-study standard error σ_i, against study outcome. This places larger studies towards the top of the plot. Hence, in the absence of bias, the plot resembles a symmetrical inverted funnel as the outcomes from smaller studies scatter more widely towards the bottom of the plot. If studies with less promising results are censored, however, then the plot will appear asymmetric about the vertical. Typically the study outcomes are treatment effects from comparative studies and it is suspected that studies are censored on the basis of their statistical significance, giving rise to an asymmetric funnel plot. However, in the context of the non-comparative studies in this meta-analysis, if studies with large proportions of deaths are less likely to report this outcome then we can also expect an asymmetric funnel plot (with the right-hand side of the funnel omitted). This is because it is more likely that small studies have the extreme results due to sampling error and are thus suppressed. In particular, note that the established trim-and-fill (Duval and Tweedie, 2000a, b) method, which attempts to respond to the asymmetry of funnel plots caused by publication bias, assumes that studies with the most extreme outcomes are censored rather than those with the most statistically significant results.

An examination of the funnel plot for the individual study estimates of overall mortality, y_i, reveals no obvious asymmetry, suggesting that publication bias is not a problem in this analysis; see Fig. 1. Having fitted a linear regression on time it may be more appropriate to plot residuals against precision, but this also provides a plot with no obvious asymmetry.

Bown *et al.* (2002) also reported a similar analysis for the second outcome, mortality in theatre. Apart from the reduced number of studies reporting this outcome and hence the associated risk of reporting bias, their analysis also ignores the obvious connection between these two measures of mortality: that the probability of death in theatre is necessarily *less* than the probability of death overall. In our analysis reported below, we allow for this natural ordering by writing the probability of death in theatre as the product of two probabilities:

$$P(\text{death in theatre}) = P(\text{overall death})\, P(\text{death in theatre} | \text{overall death}). \quad (1)$$

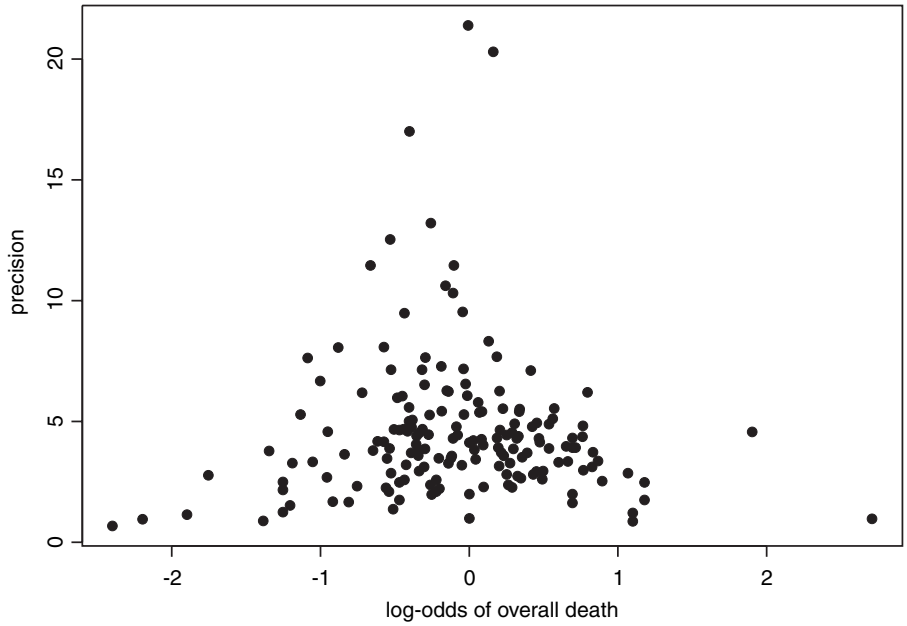

Fig. 1. Funnel plot for log-odds of overall death

The first probability, overall mortality, is modelled by the meta-regression just described. The second probability corresponds to the proportion of the overall deaths which took place during surgery, which is only reported for some of the studies. The aim of this paper is to propose a model for the analysis of this second (conditional) probability, which we present in Sections 3 and 4, and then to use equation (1) to bring the two analyses together in Section 5.

To examine the second term on the right-hand side of equation (1), the naïve approach would be to ignore the possibility of selection bias and to carry out a standard meta-regression of the conditional log-odds (measuring what proportion of the overall deaths occur in surgery) using data from only the 77 studies which report mortality in theatre. We find that the time trend is not significant, suggesting that a simple random-effects meta-analysis is appropriate. Estimating the random-effects variance τ^2 by maximum likelihood gives $\tau^2 = 0.20$. The method of DerSimonian and Laird (1986) gives a similar estimate of $\tau^2 = 0.21$. Both methods result in very similar 95% confidence intervals for the average conditional log-odds. With maximum likelihood, the estimate is -0.80 with confidence interval $(-0.94, -0.66)$. This means that between around 28% and 34% of overall deaths occur during surgery.

However, the funnel plot for this second meta-analysis is quite asymmetric (Fig. 2), suggesting that studies where a large proportion of the deaths occur during surgery are less likely to report this fact. If this is so then the analysis using only the available data will underestimate mortality. The model below is an attempt to quantify the likely size of this bias.

3. The model

3.1. Modelling mortality

For each study, let n be the total number of patients, d be the total number of deaths and d_1 be the number of deaths in theatre. Then necessarily $d_1 \leqslant d$. The empirical log-odds corresponding

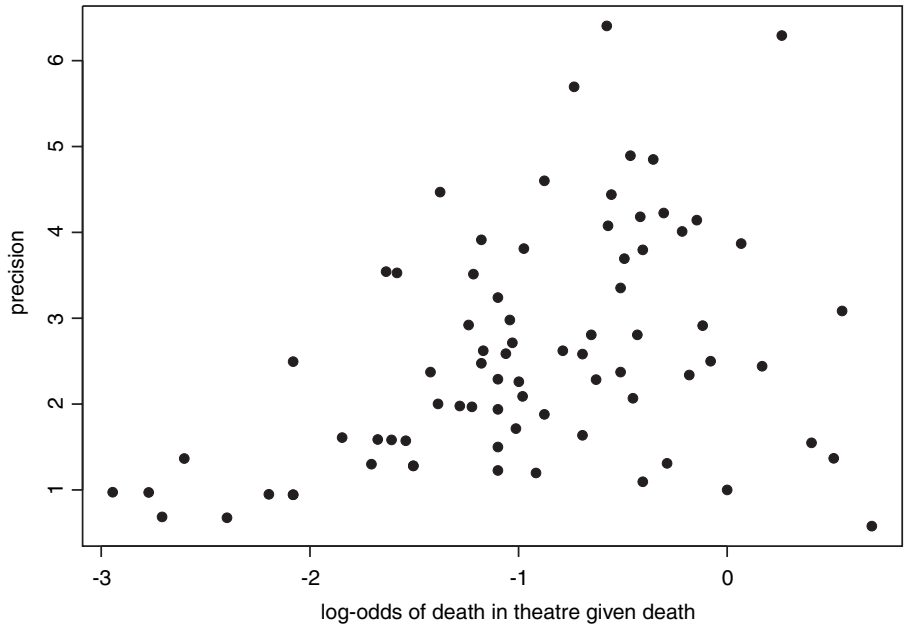

Fig. 2. Funnel plot for log-odds of the proportion of deaths in theatre

to the two probabilities on the right-hand side of equation (1) are

$$y = \log\left(\frac{d}{n-d}\right),$$

$$y_1 = \log\left(\frac{d_1}{d-d_1}\right).$$

We shall assume that, *for any given study*,

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} \Bigg| \begin{pmatrix} \theta \\ \theta_1 \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta \\ \theta_1 \end{pmatrix}, \begin{pmatrix} s^2 & 0 \\ 0 & s_1^2 \end{pmatrix}\right\}, \tag{2}$$

where θ and θ_1 are the study-specific underlying log-odds for overall mortality and conditional mortality in theatre respectively. The within-study variances s^2 and s_1^2 are assumed known but are estimated in practice, as is conventional in meta-analysis. Note the zero covariance in this conditional model. This reflects the sequential logic of equation (1). First, we can imagine dividing the patients into those who survived and those who died: y provides an estimate of the underlying probability. The patients who died are then divided into those who died in theatre and those who died in intensive care: y_1 provides an estimate of the probability underlying this second division. The value of y tells us about the *total* number of patients entering the second part of this process, but it tells us nothing about the *proportions* resulting from this second process.

This argument applies to a given study, i.e. it is conditional on the values of θ and θ_1. Undoubtedly there will be heterogeneity between studies, and so we model θ and θ_1 as joint random effects, which may be correlated. Assuming a bivariate normal model, suppose that the between-study variation follows

$$\begin{pmatrix} \theta \\ \theta_1 \end{pmatrix} \sim N\left\{\begin{pmatrix} \mu(t) \\ \mu_1(t) \end{pmatrix}, \begin{pmatrix} \tau^2 & \kappa\tau\tau_1 \\ \kappa\tau\tau_1 & \tau_1^2 \end{pmatrix}\right\}, \tag{3}$$

where $\mu(t)=\alpha+\beta t$, $\mu_1(t)=\alpha_1+\beta_1 t$ and t is the mid-date of the study (centred at 1980 as before). Then the marginal joint distribution of y and y_1 is

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mu(t) \\ \mu_1(t) \end{pmatrix}, \begin{pmatrix} s^2+\tau^2 & \kappa\tau\tau_1 \\ \kappa\tau\tau_1 & s_1^2+\tau_1^2 \end{pmatrix} \right\}.$$

Note that y and y_1 are now correlated, inheriting the dependence in the between-study random effects. Writing

$$\left.\begin{aligned} \sigma^2 &= s^2+\tau^2, \\ \sigma_1^2 &= s_1^2+\tau_1^2, \\ \rho &= \kappa\tau\tau_1/\sigma\sigma_1, \end{aligned}\right\} \tag{4}$$

this can be written more simply as

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mu(t) \\ \mu_1(t) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\sigma_1 \\ \rho\sigma\sigma_1 & \sigma_1^2 \end{pmatrix} \right\}. \tag{5}$$

3.2. Modelling the reporting process

Following the approach to publication bias that was developed by Copas and Shi (2000), we now imagine that the reporting process is triggered by a latent random variable, z say. Explicitly, suppose that d_1 (and hence y_1) is only observed if

$$z < a(t) + b\{y_1 - \mu_1(t)\} + c\{y - \mu(t)\},$$

where z is a standard normal latent variable that is independent of all the other variables in the model and $a(t)=a_1+a_2 t$. Note that, despite the obvious similarity with the model for publication bias that was proposed by Copas and Shi, the way in which the latent variable models the reporting process here is different; in particular it models the bias in a much more direct way. If R is the event that d_1 is reported, this model implies that

$$P(R|y, y_1) = \Phi[a(t) + b\{y_1 - \mu_1(t)\} + c\{y - \mu(t)\}], \tag{6}$$

where Φ is the standard normal cumulative distribution function.

We therefore allow the probability of reporting y_1 to depend on all the observables in the model that is described by equation (5) which permits a wide range of possibilities. Note that there are four parameters describing the reporting process. Parameters a_1 and a_2 allow the probability of R to depend directly on time, which is appropriate as reporting rates may change over the course of half a century. Parameter c allows the probability of R to depend on the value of y. The parameter b is of particular interest, as it allows the probability of R to depend on the possibly unobserved value of y_1. If $b<0$ then the type of reporting bias that is suspected from Fig. 2 is present, as this indicates that studies with high operative mortality rates are less likely to report y_1, and the size of b indicates the extent of this. This provides an explanation for the asymmetry in Fig. 2, as discussed in the previous section. Also note that the probability of R depends on the difference between the log-odds of the two types of mortality and the corresponding population mortality rates. This makes intuitive sense as, for example, if there is an improvement over time then a mortality rate which seemed good in 1950 may become disappointing 50 years later. Despite all these desirable features, it should be recognized that this is merely a plausible model for the reporting process.

3.3. The likelihood function

By first conditioning on y in model (5), we can write the joint distribution of y and y_1 as

$$\frac{1}{\sigma\sigma_1\sqrt{(1-\rho^2)}}\phi\left\{\frac{y-\mu(t)}{\sigma}\right\}\phi\left[(1-\rho^2)^{-1/2}\left\{\frac{y_1-\mu_1(t)}{\sigma_1}-\rho\frac{y-\mu(t)}{\sigma}\right\}\right], \tag{7}$$

where ϕ is the standard normal density function. A study in which d_1 is reported therefore contributes a factor

$$P(y, y_1, R) = P(R|y, y_1)\, P(y, y_1) \tag{8}$$

to the likelihood. This is just the product of expressions (6) and (7).

For a study in which d_1 is not reported, first note that

$$\begin{aligned} P(\bar{R}|y) &= P[z > a(t) + b\{y_1 - \mu_1(t)\} + c\{y - \mu(t)\}|y] \\ &= P\{z_1 > a(t)|y\}, \end{aligned} \tag{9}$$

where $\bar{R}$ denotes the complement of R and

$$z_1 = z - b\{y_1 - \mu_1(t)\} - c\{y - \mu(t)\}.$$

Since z, y and y_1 are jointly multivariate normal, the joint distribution of z_1 and y is bivariate normal and is

$$\begin{pmatrix} z_1 \\ y \end{pmatrix} \sim N\left\{\begin{pmatrix} 0 \\ \mu(t) \end{pmatrix}, \begin{pmatrix} 1+b^2\sigma_1^2+c^2\sigma^2+2bc\rho\sigma\sigma_1 & -(b\rho\sigma\sigma_1+c\sigma^2) \\ -(b\rho\sigma\sigma_1+c\sigma^2) & \sigma^2 \end{pmatrix}\right\}.$$

from which we can evaluate equation (9) as

$$P(\bar{R}|y) = 1 - \Phi\left[\{1+b^2\sigma_1^2(1-\rho^2)\}^{-1/2}\left\{a(t) + (b\rho\sigma_1 + c\sigma)\frac{y-\mu(t)}{\sigma}\right\}\right]. \tag{10}$$

The contribution to the likelihood for a study in which d_1 is not reported is

$$P(y, \bar{R}) = P(\bar{R}|y)\, P(y), \tag{11}$$

which is just the product of equation (10) and the marginal density of y from expression (5).

Using these formulae, we now have the log-likelihood function for the observed data as

$$\log(L) = \sum_{R} \log\{P(y, y_1, R)\} + \sum_{\bar{R}} \log\{P(y, \bar{R})\}, \tag{12}$$

the summations being over studies in which d_1 is, and is not, reported.

4. Fitting the model

4.1. Estimating the within-study variances

Assuming that the sample sizes are not too small, we follow the usual approximation of replacing the within-study variances by sample estimates. Overall mortality is always observed, so s^2 can be estimated in the usual way:

$$s^2 = \frac{1}{d} + \frac{1}{n-d}.$$

Similarly, when d_1 is observed, we can estimate

$$s_1^2 = \frac{1}{d_1} + \frac{1}{d-d_1}.$$

When d_1 is missing, we suggest replacing d_1 by its fitted value for that particular study, i.e. replace d_1 by $d\exp\{\mu_1(t)\}/[1+\exp\{\mu_1(t)\}]$, giving

$$s_1^2 = \frac{[1+\exp\{\mu_1(t)\}]^2}{d\exp\{\mu_1(t)\}}. \tag{13}$$

We now use expression (4) to write the marginal variances σ^2 and σ_1^2 in terms of the unknown parameters τ^2 and τ_1^2.

4.2. Simplifying the model

The full model has 11 parameters. The maximum likelihood estimates of these, found by numerical maximization of equation (12), are shown in the second column of Table 1. Standard errors, obtained from the observed information matrix, are shown in parentheses. Substituting these estimates into equation (12) achieves a log-likelihood of −339.51.

Five of these parameters (β, β_1, a_2, b and c) model time trends and biases that are included in the model but may not be needed to describe the data well. Their contribution to the model can be assessed by likelihood ratio (deviance) tests in which each one of these parameters is removed from the likelihood in turn: Table 2. It is clear that β is needed in the model, as we should expect from the previous meta-regression. It seems reasonable to simplify the model by setting β_1, a_2 and c to 0, although conclusions concerning b are less clear. Setting β_1 to 0 is consistent with the earlier finding that the time trend in the meta-regression for the conditional probability of death

Table 1. Maximum likelihood estimates†

Parameter	*Full model*	*With reporting bias ($b \neq 0$)*	*Without reporting bias ($b = 0$)*
α	−0.072 (0.041)	−0.072 (0.041)	−0.073 (0.041)
β	−0.014 (0.004)	−0.014 (0.004)	−0.014 (0.004)
α_1	−0.715 (0.111)	−0.725 (0.110)	−0.857 (0.071)
β_1	0.004 (0.007)	—	—
τ	0.458 (0.035)	0.458 (0.035)	0.458 (0.035)
τ_1	0.415 (0.064)	0.419 (0.063)	0.436 (0.064)
κ	0.425 (0.177)	0.438 (0.174)	0.527 (0.140)
a_1	−0.141 (0.101)	−0.139 (0.101)	−0.125 (0.096)
a_2	−0.003 (0.010)	—	—
b	−0.453 (0.293)	−0.423 (0.286)	—
c	0.088 (0.164)	—	—

†Standard errors are given in parentheses.

Table 2. Results obtained when removing one variable from the model

Variable removed	*Maximum log-likelihood*	*p-value*
β	−345.16	<0.01
β_1	−339.69	0.55
a_2	−339.56	0.75
b	−340.85	0.10
c	−339.66	0.59

in theatre is not significant. An investigation involving setting combinations of these parameters to 0 gave quite similar results. Because of our particular interest in parameter b and its role in modelling selection bias, we shall set β_1, a_2 and c to 0 and perform two analyses: one with $b=0$ and the other with $b\neq 0$. This will provide us with two sets of results: one assuming that data on the number of deaths in theatre are not subject to reporting bias; the other assuming that this type of bias has indeed occurred.

4.3. Results

Fitting the larger of these two reduced models by maximum likelihood produces the results that are shown in the third column of Table 1. The negative values of $\hat{b}$ in this and the full model confirm the impression from the funnel plot in Fig. 2, that studies with a high proportion of the deaths occurring in theatre are less likely to report this fact. The results if b is constrained to be 0 (no reporting bias) are shown in the fourth column of Table 1.

The three main parameters of interest are α and β, the intercept and slope of the logistic regression of overall mortality on time, and α_1, the logit of the proportion of deaths which occur in theatre (assumed constant over time). Inference for these quantities is summarized in Table 3. The appropriate estimate is given in the centre of each triple; the outer figures are the corresponding 95% confidence limits. The first row reports the results of the naïve analyses that were discussed in Section 2. The second and third rows report the analysis of the model with and without the constraint $b=0$. Here we show the maximum likelihood estimates with confidence limits obtained by inverting the corresponding likelihood ratio tests.

Table 3 shows that the inferences concerning α and β are almost identical for all three approaches. Since every study reports y, it is not surprising that modelling bias in y_1 makes little difference to our inference concerning these parameters. It is perhaps surprising that the naïve confidence intervals are not narrower for these parameters as these do not take into account the uncertainty concerning τ^2. With 171 studies this is evidently unimportant.

As expected, however, there are differences in the inference for α_1. Comparing the naïve analysis with the model with $b=0$, the intervals are roughly the same width, but the confidence interval for the model is shifted slightly downwards. This reflects the fact that the observed values of y tend to be slightly smaller in those studies where y_1 is reported ($\hat{c}>0$ in the full model), and that y and y_1 are positively correlated ($\hat{\kappa}>0$). But, if we allow for reporting bias ($b\neq 0$), the estimate of α_1 moves upwards, reflecting the negative value of $\hat{b}$ that was discussed earlier. More marked, however, is the doubling of the width of the confidence interval, reflecting the substantial uncertainty in the value of b. On the probability scale, the proportion of deaths which occur in theatre is estimated to be between 27% and 33% with the assumption of no reporting bias, but between 28% and 38% after correcting for this bias. This increase in uncertainty reflects a common problem in modelling publication bias in meta-analysis: we

Table 3. Comparison of the naïve analyses (Section 2) and those from the model†

Method	α	β	α_1
Naïve	(−0.15, −0.07, 0.01)	(−0.022, −0.014, −0.006)	(−0.94, −0.80, −0.66)
Model with $b=0$	(−0.15, −0.07, 0.01)	(−0.022, −0.014, −0.007)	(−1.00, −0.86, −0.72)
Model with $b\neq 0$	(−0.15, −0.07, 0.01)	(−0.022, −0.014, −0.007)	(−0.94, −0.73, −0.48)

†The three tabulated values are the lower limit of the 95% confidence interval, the point estimate and the upper limit of the 95% confidence interval.

suspect that there is non-ignorable selection but the data give us very little information about how severe the resulting bias may actually be. Evident from Table 1 is that very similar results concerning the key parameters shown in Table 3 are obtained when using the full model and the model with $b \neq 0$. Hence constraining β_1, a_2 and c to 0 makes very little difference to the resulting confidence intervals for these parameters. A further comment from our analysis is that all likelihood-based confidence intervals in Table 3 are remarkably symmetrical, suggesting that standard asymptotic inference using observed information matrices is adequate for models of this kind. This contrasts sharply with the analogous selection models for missing studies (Copas and Shi, 2000).

4.4. Illustrating the model fit

Our suspicion of reporting bias in this meta-analysis arose from the asymmetry of the funnel plot of the observed values of y_1 in Fig. 2. One way of illustrating a model of the selection process is to use it to impute the missing values in a funnel plot of this kind. If the model is appropriate, then the filled funnel plot should look nicely symmetrical.

In the notation of Section 3, the distribution that we need for imputing the missing values of y_1 is

$$P(y_1 | y, \bar{R}) = \frac{P(y, y_1)\{1 - P(R|y, y_1)\}}{P(y, \bar{R})}. \tag{14}$$

The numerator is given by formulae (7) and (6), and the denominator is the term (11) that is used in the likelihood function.

Imputation requires numerical values for all the parameters in the model. Fig. 3 is a filled funnel plot using simulated values of y_1 from equation (14) for the studies that fail to report this value assuming no reporting bias: here we have taken $b = 0$ and used the constrained maximum likelihood estimates of the remaining parameters from the fourth column of Table 1. Fig. 4 is

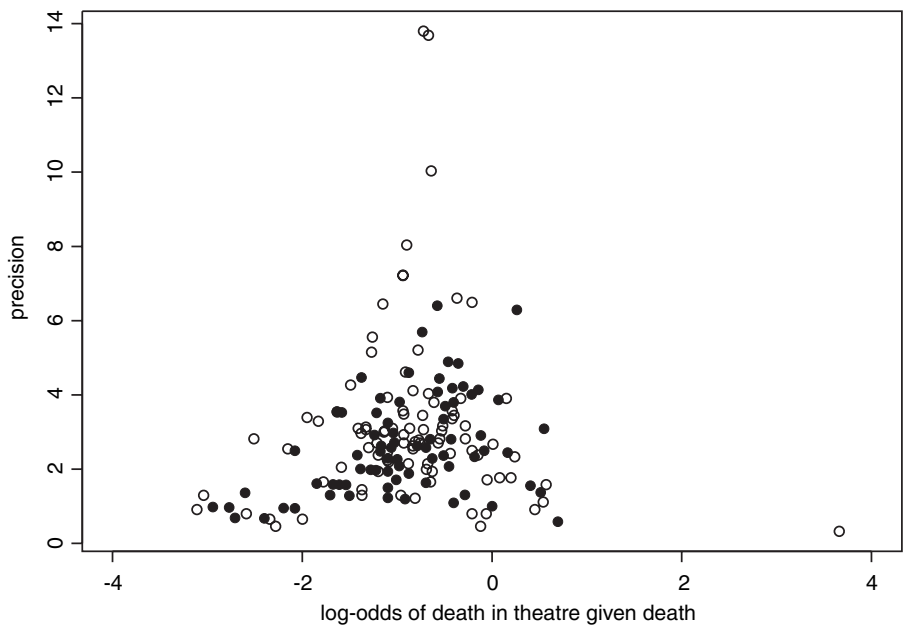

Fig. 3. Filled funnel plot with simulated values for $b = 0$: ○, imputed values; ●, observed values

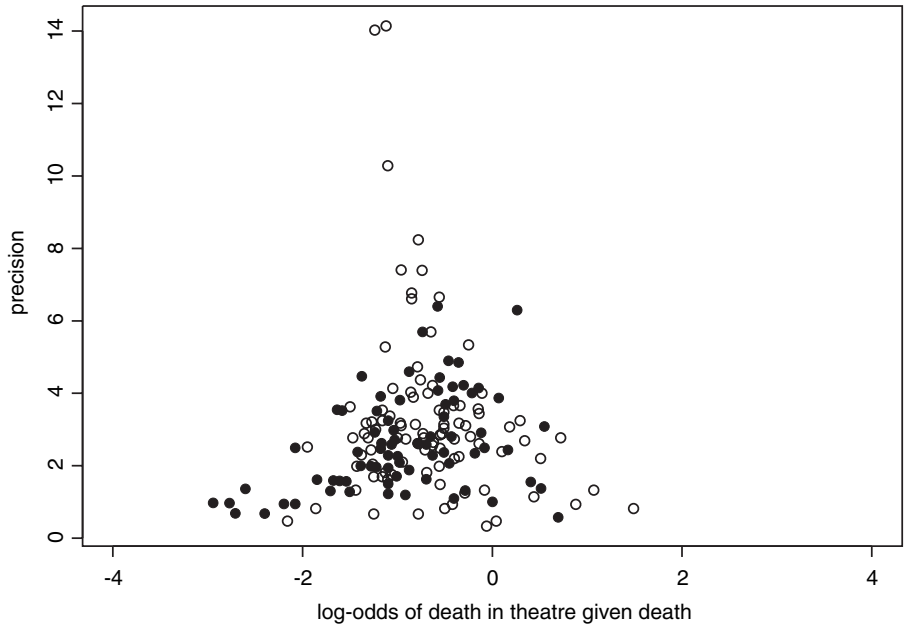

Fig. 4. Filled funnel plot with simulated values for $b \neq 0$: ○, imputed values, ●, observed values

a filled funnel plot for the model in which b is also estimated. In each plot the full circles are the observed values and the open circles are the imputed values of y_1. The plots are the result of just a single realization of this imputation procedure.

The filled plot with $b = 0$ still appears slightly asymmetric, although it is much more symmetrical than before. Note that by chance a very large value of y_1 has been produced by the simulation; this value would probably be considered an outlier if it were encountered in the original data set. The funnel plot for $b \neq 0$, however, appears more symmetrical, the kind of funnel plot that, had we seen this for the original data set, would reassure us that bias has had no obvious effect.

5. Modelling the unconditional probability of death in theatre

By studying the joint distribution of y and y_1 we now have a model for the two terms on the right-hand side of equation (1). Combining these into a single model for the unconditional probability of death in theatre gives a further way of discussing our analysis and allows us to compare this approach with the naïve meta-analysis of the number of deaths in theatre that was discussed in Bown *et al.* (2002).

If we ignore the problem of reporting bias, we could use standard meta-regression for the 77 studies in which d_1 is reported. This would fit the model

$$y_2 = \log\left(\frac{d_1}{n - d_1}\right) \sim N(\theta_2, s_2^2)$$

where s_2^2 is the estimated within-study variance of y_2 and θ_2 is a random effect with a possible time trend,

$$\theta_2 \sim N(\alpha_2 + \beta_2 t, \tau_2^2).$$

We find that the time trend is not significant, suggesting that we can set $\beta_2 = 0$. To assess whether the inference from our model and naïve analyses are broadly compatible, this simplification will be made here but the time trend will be reinstated for a more detailed comparison later. Assuming $\beta_2 = 0$ provides estimates of α_2 and τ_2 of -1.74 and 0.53 respectively. Converting back to the probability scale, the marginal probability of death in theatre is

$$f_1(\alpha_2) = E\left\{\frac{\exp(\theta_2)}{1+\exp(\theta_2)}\right\} = \int_{-\infty}^{\infty} \frac{\exp(\alpha_2+\tau_2 x)}{1+\exp(\alpha_2+\tau_2 x)}\phi(x)\,\mathrm{d}x. \tag{15}$$

Substituting parameter estimates and evaluating equation (15) by numerical integration gives 0.162 with standard error 0.010. The standard error is found by taking τ_2 as fixed, approximating $f_1(\alpha_2)$ as a linear function of α_2 and using the standard meta-analysis variance of $\hat{\alpha}_2$. The asymptotic 95% confidence interval for the probability of death in theatre is therefore about 14–18%.

Extending this argument to the model of Section 3.1, we now need to integrate over the joint random effects θ and θ_1 in distribution (3). Using equation (1), the marginal probability of death in theatre at time t is now

$$\begin{aligned} f_2(\alpha,\beta,\alpha_1,\beta_1,\tau,\tau_1,\kappa) &= E\left\{\frac{\exp(\theta)}{1+\exp(\theta)}\frac{\exp(\theta_1)}{1+\exp(\theta_1)}\right\} \\ &= \frac{\exp\{\mu(t)+\mu_1(t)\}}{\sqrt{(1-\kappa^2)}}\int_{-\infty}^{\infty}\frac{\exp(\tau x)}{1+\exp\{\mu(t)+\tau x\}}\phi(x) \\ &\quad\times\left[\int_{-\infty}^{\infty}\frac{\exp(\tau_1 y)}{1+\exp\{\mu_1(t)+\tau_1 y\}}\phi\left\{\frac{y-\kappa x}{\sqrt{(1-\kappa^2)}}\right\}\mathrm{d}y\right]\mathrm{d}x. \end{aligned} \tag{16}$$

Again this can be evaluated by numerical integration and asymptotic standard errors obtained from a linear approximation of the function f_2 with the appropriate observed information matrix.

Values of equation (16) for $t = -30, -20, \ldots, 20$ give estimates of mortality in theatre in 10-year intervals from 1950 to 2000. Because we are particularly interested in trends over time, we present the results using the maximum likelihood estimates for the full model of Section 3 without the simplifying restrictions of Section 4.2. In particular this reinstates the time trend in $\mu_1(t)$. The results are shown in Fig. 5. The points that are joined by dotted lines are the estimated probabilities of death in theatre with their upper and lower 95% confidence limits. The confidence limits from the naïve analysis using equation (15) are shown for comparison (the full lines). Although the results are broadly similar, the model does seem to have resulted in slightly larger estimates of the probability of death in theatre and the confidence intervals resulting from this are much wider than the interval from the naïve analysis. Owing to the difficulty in quantifying the degree of the reporting bias, this is exactly as expected. Similar results were also obtained for the two reduced models. The estimated probabilities from the model with $b \neq 0$ were generally slightly greater than those of the full model and, of course, slightly smaller estimates were produced with $b = 0$, but all three models produced results that were broadly consistent with the naïve analysis. The model with $b \neq 0$ produced narrower intervals than the full model, but these were still appreciably wider than those of models that ignore the possibility of reporting bias. The model with $b = 0$ produced the narrowest intervals, which were roughly the same width as the naïve analysis interval. Despite this, it should also be noted that all three models have one important thing in common: *all three indicate some improvement in operating survival rates over time*.

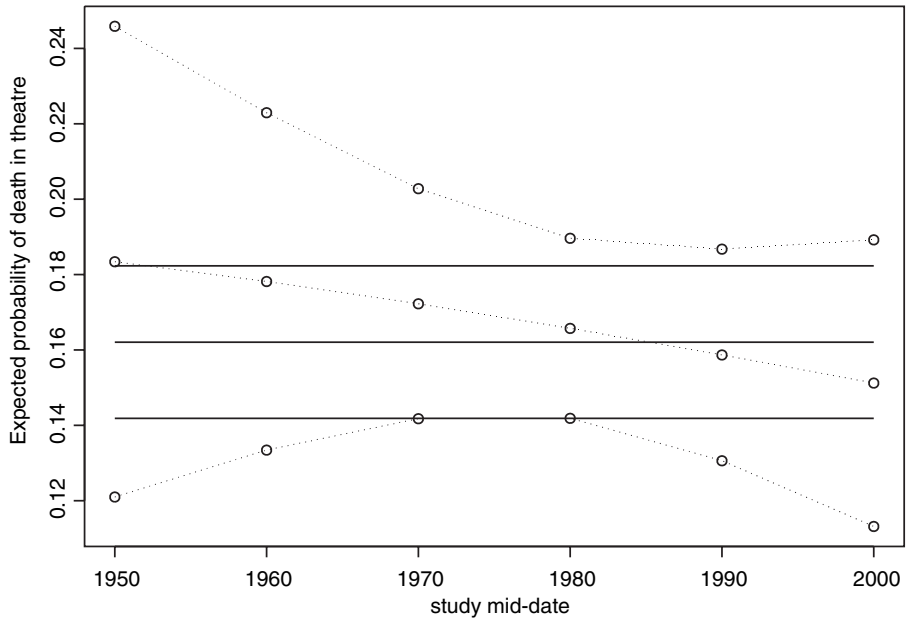

Fig. 5. Probability of death in theatre: ○, ·····, point estimates and limits of 95% confidence intervals for the years 1950, 1960,…, 2000; ———, results from a naïve meta-analysis

In particular, the full model indicates a modest improvement, by about 3 percentage points from 1950 to 2000. Interestingly, this downward trend is not at all evident if we analyse the data from the complete studies alone in the usual naïve way. Fig. 6 shows the observed values of y_2 plotted against time. Each of the 77 studies here is represented by a circle with area proportional to the within-study variance of y_2. In fact the trend is slightly *positive*; the point estimate of β_2 is 0.005 (with a standard error of 0.008).

To reconcile these findings we can use a procedure that is similar to Section 4.4 and use the model to impute the missing values of y_2, hence filling out the time plot of Fig. 6 to cover all 171 studies. This results in Fig. 7. The full circles are the observed values of y_2 copied from the previous graph; the open circles are the 94 imputed values with areas proportional to the corresponding estimated variances. These imputed values were obtained from the product in equation (1), taking the first term as the observed proportion of deaths overall, and the second term as the probability derived from the full 11-parameter model as used for Fig. 5. For the variance, we have simply used the standard formula for the variance of a log-odds. A few of these open circles are very large, indicating the substantial uncertainty in imputing data for some of the smaller studies (we have already seen this effect in Figs 3 and 4 with some of the imputed values being near the top of these plots). A standard meta-regression fitted to all 171 values of y_2, treating the imputed values as if they were genuine observations, gives the downward trend that is shown in Fig. 7. The line falls from a log-odds of -1.44 in 1950 down to -1.78 in 2000. On the probability scale, this means a decrease from about 19% to 14%, which is reasonably consistent with the analysis in Fig. 5.

To summarize, we find that taking into account reporting bias produces only slightly larger estimates of probabilities of death in theatre. Although the full lines in Fig. 5 represent the very simplest of naïve analyses, we can see from this that estimates from our full model are broadly similar, only differing from the naïve estimate by around 2% over the entire 50-year period. We

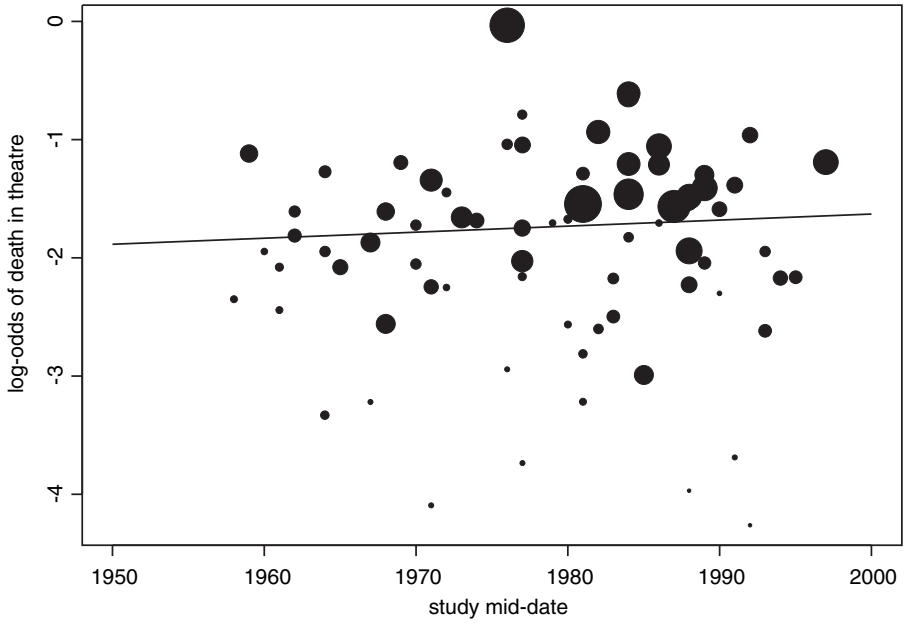

Fig. 6. Meta-regression for log-odds of death in theatre

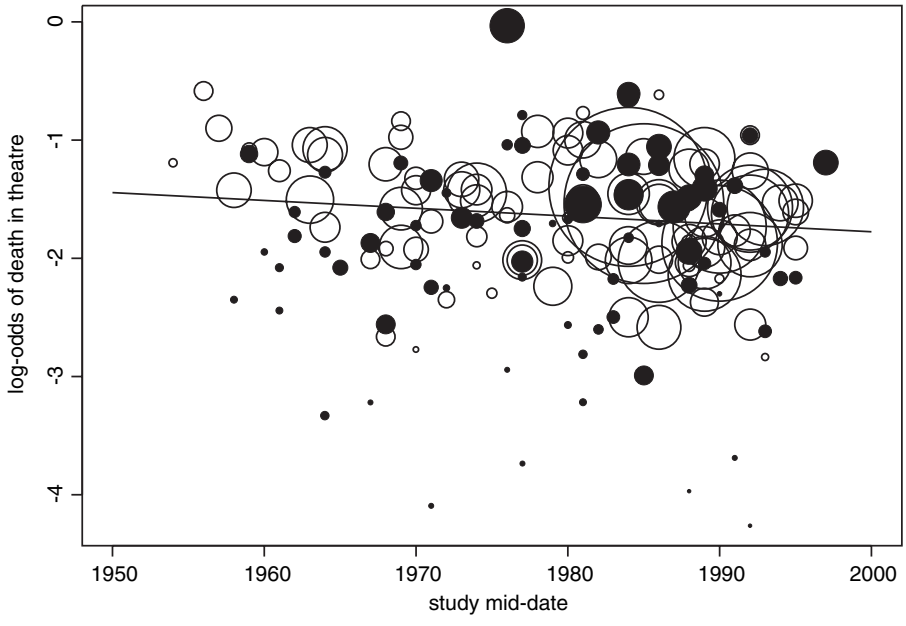

Fig. 7. Meta-regression for log-odds of death in theatre with imputed values: ○, imputed values; ●, observed values

also found broad agreement when using both reduced models. In part this is due to the wider confidence intervals resulting from our model; modelling reporting bias causes confidence intervals to widen, emphasizing that naïve methods have a tendency to provide confidence intervals that are artificially narrow. Our model also provides an explanation for the surprising positive time trend in the probability of death in theatre that is suggested by the naïve meta-regression.

All three models reverse this trend, reducing concerns that the emergency operation has become less effective over the last 50 years.

6. Discussion

This paper has presented a method for adjusting the inference from a meta-analysis in which the suppression of data would appear to be evident for one of the outcomes. Despite this, there was an absence of strong evidence for reporting bias. The results obtained therefore essentially provide a sensitivity analysis as, with a large proportion of studies failing to report a particular outcome, it is very difficult to make precise statements about the degree of bias and therefore certain parameters. For this data set, some of the inferences are not particularly robust and we are reminded of the importance of attempting to take into account the possibility of biases when performing meta-analyses. Of particular interest is that all three models that were used estimate an improvement in the operative mortality rate over time, which is in stark contrast with the naïve meta-regression. In this example, estimating mortality in theatre is of clinical importance since it is of interest to know whether improvements in overall mortality can be attributed to greater success in surgery or are due to post-surgical intensive care. Clearly, this picture can be distorted if we take the available data at 'face value'.

The model concerns the joint inference of log-odds of death, and conditional log-odds of death in the operating theatre, although the results can be related to and interpreted in terms of the corresponding unconditional probability as explained. This meta-analysis is perhaps atypical in the sense that non-comparative outcomes are of interest, and events in one outcome definition can be considered a subset of the events in the other. Meta-analyses on log-odds ratio scales comparing two different groups are more common. Modification and extension of the modelling ideas that are presented here so that they could be applied to other meta-analysis data sets are potentially of great interest. Another important feature of the data is that we know that both outcomes do exist for every study. In other applications, this may not be so; for example a depression rating questionnaire may be used in several trials to measure the effect of anti-depressant drugs but not in all. It is not clear whether modelling adjustments should always be made for trials that did not report an outcome on a particular scale, since they truly may not have used that instrument.

In the original meta-analysis of these data in Bown *et al.* (2002), the problem of outcome reporting bias was identified and the simple trim-and-fill (Duval and Tweedie, 2000a, b) method of adjusting asymmetric funnel plots by estimating and imputing the minimum number of studies required to make the funnel symmetric was applied. This analysis treated each outcome independently (i.e. in isolation), hence ignoring knowledge regarding the total number of studies and the relationships between outcomes. A direct comparison of the results of the model that is described here and trim and fill is difficult because trim and fill does not take into account moderator variables. However, the results are qualitatively similar with both methods suggesting that naïve analyses underestimate the operative mortality rate.

It has been observed that funnel asymmetry may be due to factors other than publication biases (Egger *et al.*, 1997). For example, if moderator variables exist which are related to both outcome and study precision that are not adjusted for in the analysis, such confounding may manifest itself as funnel plot asymmetry. Hence it is difficult to be certain that funnel asymmetry is truly publication bias. However, in this example, the fact that the funnel plot for total mortality is symmetrical, and that many studies do not report mortality in theatre, does add weight to the belief that the asymmetry is truly being caused by publication biases.

The model assumes that the reporting bias applies only to the number of deaths in theatre. Indeed, it is because of this assumption that good estimates of the selection parameters can be obtained. This is in contrast with previous modelling of missing studies (Copas and Shi, 2000, 2001). However, in other applications this may not be so, and data may be selectively missing at both the study and the outcome level. A method of modelling both selection mechanisms simultaneously is currently being investigated.

References

Bown, M. J., Sutton, A. J., Nicholson, M. L. and Bell, R. F. (2002) A meta-analysis of 50 years of ruptured aortic aneurism repair. *Br. J. Surg.*, **89**, 714–730.

Copas, J. and Shi, J. Q. (2000) Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, **1**, 247–262.

Copas, J. and Shi, J. Q. (2001) A sensitivity analysis for publication bias in systematic reviews. *Statist. Meth. Med. Res.*, **10**, 251–265.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Contr. Clin. Trials*, **7**, 177–188.

Duval, S. and Tweedie, R. (2000a) A nonparametric trim and fill method for accounting for publication bias in meta-analysis. *J. Am. Statist. Ass.*, **95**, 89–98.

Duval, S. and Tweedie, R. (2000b) Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **56**, 455–463.

Egger, M., Smith, G. D., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.*, **315**, 629–634.

Hahn, S., Williamson, P. R., Hutton, J. L., Garner, P. and Flynn, E. V. (2000) Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Statist. Med.*, **19**, 3325–3336.

Hutton, J. L. and Williamson, P. R. (2000) Bias in meta-analysis due to outcome variable selection within studies. *Appl. Statist.*, **49**, 359–370.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.

Sterne, J. A. C. and Egger, M. (2001) Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J. Clin. Epidem.*, **54**, 1046–1055.

Sutton, A. J. and Piggot, T. D. (2005) Bias in meta-analysis induced by incompletely reported studies. In *Publication Bias in Meta-analysis: Prevention, Assessments and Adjustments* (eds H. Rothstein, M. Borenstein and A. J. Sutton). Chichester: Wiley.

Sutton, A. J., Song, F., Gilbody, S. M. and Abrams, K. R. (2000) Modelling publication bias in meta-analysis: a review. *Statist. Meth. Med. Res.*, **9**, 421–445.

Thompson, S. G. and Sharp, S. J. (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Statist. Med.*, **18**, 2693–2708.