A Sensitivity Analysis for Publication Bias in Systematic Reviews

J. B. Copas*
and
J. Q. Shi
University of Warwick, UK

SUMMARY

There is no simple method of correcting for publication bias in systematic reviews. We suggest a sensitivity analysis in which different patterns of selection bias can be tested against the fit to the funnel plot. Publication bias leads to lower values, and greater uncertainty, in treatment effect estimates. Two examples are discussed. An appendix lists the *S-plus* code needed for carrying out the analysis.

**Address for correspondence*: Department of Statistics, University of Warwick, Coventry CV4 7AL, UK, email: jbc@stats.warwick.ac.uk

1 Introduction

Publication bias is increasingly recognised as a major difficulty in systematic reviews of medical studies. Research journals cannot possibly publish all submitted manuscripts and so editors and referees have to be selective. Inevitably, it is the papers with interesting and statistically significant results, rather than studies with inconclusive outcomes, which tend to be the ones accepted for publication. Thus, when the published literature is reviewed, the studies which can be identified are a biased selection of the research that has been done in that particular area.

The problem is made worse by the fact that many medical studies look at multiple outcomes, and there is a tendency for only those suggesting a significant effect to be mentioned when the study is written up. Outcomes which show no clear treatment effect are often ignored, and so will not be included in any later review of studies looking at those particular outcomes. This is another cause of publication bias (Hutton and Williamson[1]).

A well known example is the review of studies of the effect of magnesium in reducing risk of heart disease, Yusuf *et al.*[2]. Most studies found a positive effect, which led to the reviewers' conclusion that "(magnesium is) effective, safe, simple and inexpensive". A subsequent large multi-centre clinical trial contradicted this result and showed that magnesium is in fact completely ineffective[3]. Egger and Smith[4] demonstrated that the meta-analysis was wrong because the authors failed to allow for what seemed in this case to be very substantial publication bias. Sutton *et al.*[5] suggest that about one half of all published reviews show clear signs of the same problem.

However, how to allow for publication bias in a review is far from obvious. Egger *et al.*[6] suggested a test for the presence of publication bias based on testing for a trend in the funnel plot. This is the plot of study outcome against study size (or some measure of accuracy of that outcome). Looking at the funnel plot should surely be the first step in any meta-analysis, but the danger of the testing approach is the temptation to assume that, if the test is not significant, there is no problem and hence the possibility of publication bias can be ignored. We suggest this is an illusion — publication bias is endemic to all empirical research and so is likely to be there even if it is not obvious from the plot. Many meta-analyses have only a small number of studies, and the statistical power of such tests can be very low.

Several methods have been suggested for "correcting" for publication bias, as recently reviewed by Sutton *et al.*[7]. Since it is impossible to know exactly how many studies have actually been done but not written up or published, all these methods have to make assumptions which cannot be tested, and so at best can only be regarded as tentative or suggestive. These methods are somewhat *ad hoc* and can come to quite different conclusions. It is best to regard these as an informal sensitivity analysis – we try different approaches and monitor how sensitively the results depend on which assumptions we make. The purpose of this paper is to show that this can be done much more directly — we use a *model* to describe the process of study selection, estimate the treatment effect of interest for different parameter choices within this model, and then check the fit of each estimate with the evidence in the funnel plot.

A more technical specification of the method we describe is in Copas and Shi[8], following earlier ideas of sample selection models in Copas[9] and Copas and Li[10]. In this paper we aim to give a rather simpler description of the model, explain how the sensitivity analysis works, and look at two contrasting examples. In the Appendix we list the *S-Plus* routines which we have used for carrying out these calculations. These routines could be translated into forms suitable for other software packages, assuming that suitable mathematical routines are available or can be imported.

The paper is organised as follows. Section 2 sets out the underlying statistical model and Section 3 discusses inference aspects such as estimates and confidence intervals. Then Section 4 sets out the details of the sensitivity analysis and describes the *S-plus* functions. Section 5 illustrates their use on two examples, with some concluding comments in Section 6.

2 The model

The aim of meta-analysis is to combine together estimates of a quantity of interest, y say, across several different studies. The most direct example is a series of randomised clinical trials in which the same treatment or intervention is being compared with the same control, and where y is the log-odds ratio using the same binary outcome. Often there will be confounding factors or differences in study design which may effect the outcome — we assume that each trial has made an appropriate correction to the value of y to allow for such differences. We also assume that y is measured in such a way that the larger is y the more effective is the treatment.

The model is based on two ideas.

Study Population. There is a population of studies which have been conducted in the area of interest. Each has its outcome y measured with standard deviation σ. Our aim is to learn about the average value of y across this population.

Study Selection. The studies in the meta analysis are a selection from the studies in this population. To be selected, a study has to be written up, published, and selected by the reviewer. Each of these stages involves uncertainty, which we model as a probability, p say, that any particular study in the population appears in the meta analysis. Bias arises when p is correlated with y. We make two assumptions about p:

- Very large studies have p close to 1. This reflects the fact that sponsors of large and expensive studies will always require the results to be disseminated, and journal editors will be keen to publish what will be regarded as the definitive study in the particular area. For small studies, on the other hand, p will be less than 1, possibly quite small.
- For a given size of study, the selection probability will be an increasing function of y. Thus the more effective the study claims the treatment to be (the more highly significant the result), the more likely it is that that study will be selected.

Each part is modelled separately as follows:

Population Model. The ith study in the population has estimate y_i with

$$y_i \sim N(\mu_i, \sigma_i^2)$$

and

$$\mu_i \sim N(\mu, \tau^2).$$

This is the standard random effects model, such as in Whitehead and Whitehead[10]. The parameters are μ, the overall mean effect, σ^2, the within study variance, and τ^2, the between study variance measuring the heterogeneity between the different studies.

Selection Model. Suppose that each study in the population has treatment effect estimate y with *reported* standard deviation s (which may of course be different from the true standard deviation σ). Then, amongst studies reporting the same value of s, we expect a proportion

$$P(\text{select}|s) = \Phi(a + \frac{b}{s}) \tag{1}$$

to be selected, where Φ is the standard normal cumulative distribution function. Parameter a controls the overall proportion of studies selected, parameter b (assumed positive) controls how fast this proportion increases as s becomes smaller (study size increases).

Equation (1) can be written in an entirely equivalent way: a study is selected if and only if the value of the random quantity

$$z = a + \frac{b}{s} + \delta$$

is positive, where

$$\delta \sim N(0, 1).$$

We can interpret z as the *propensity for selection*: the larger is z the more likely the study is to be selected.

Combined Model. Combining these two models, we have for the ith study in the population

$$\begin{aligned} y_i &= \mu_i + \sigma_i \epsilon_i & (2)\\ \mu_i &\sim N(\mu, \tau^2) & (3)\\ z_i &= a + \frac{b}{s_i} + \delta_i, \end{aligned}$$

where standard normal residuals (ϵ_i, δ_i) are jointly normal with correlation

$$\text{corr}(y_i, z_i) = \rho.$$

These residuals are independent across different values of i. The studies in the meta analysis consist of those ones with data (y_i, s_i) for which $z_i > 0$. Note that the model has introduced a correlation parameter ρ. If ρ is large and positive, then selected studies will have $z_i > 0$, hence are more likely to have positive δ_i, hence are more likely to have positive ϵ_i, and hence are more likely to have y_i overestimating the true mean μ. This is just what we would expect of publication bias.

Some Mathematical Consequences. To set out some properties of the above model, define the notation

$$u = a + \frac{b}{s}\ , \quad \lambda(u) = \frac{\phi(u)}{\Phi(u)}\ , \quad \tilde{\rho} = \frac{\sigma}{\sqrt{\tau^2 + \sigma^2}}\rho$$

where ϕ is the standard normal density function. Then for a typical study in the population it can be shown that

$$P(\text{select}|y, s) = p(a, b, s, y) = \Phi\left(\frac{u + \tilde{\rho}\frac{y-\mu}{\sqrt{\tau^2+\sigma^2}}}{\sqrt{1 - \tilde{\rho}^2}}\right). \tag{4}$$

This formula confirms the dependence of the selection probability on both s and y. Averaging this over y for a given value of s gives

$$P(\text{select}|s) = \Phi(u)$$

as before.

It can also be shown that

$$\mathrm{E}(y|s, \text{select}) = \mu + \rho\sigma\lambda(u). \tag{5}$$

This confirms that the expected value of y for a selected study (a study actually in the meta-analysis) is an increasing function of σ, the smaller the study the larger the observed outcome y. This is the dependence between y and s which we often observe in the funnel plot.

Similarly

$$\mathrm{Var}(y|s, \text{select}) = \sigma^2 \left(1 - \lambda(u)(u + \lambda(u))\right) \tag{6}$$

showing that the standard deviations s of the selected studies tend to be somewhat smaller than the standard deviations σ in the population as a whole.

These formulae are derived from the assumption that the conditional distribution of ϵ and δ given s is bivariate normal. This is exactly true if y is the mean of a normal sample, as y and s are then independent, but is only an approximation in other cases. See Copas and Shi[7] for discussion of this and other technical aspects.

3 Sensitivity analysis: theory

Using these formulae we can construct the likelihood function, using the fact that the distribution of y in a selected study is given by the conditional distribution of y given that $z > 0$. By Bayes theorem, the likelihood is the product of the marginal distribution of y times the conditional distribution of selection given s and y divided by the marginal probability of selection. The log likelihood $L(\mu, \rho, \tau, a, b)$ is therefore the sum over the studies in the meta analysis of

$$-\frac{1}{2}\log(\sigma_i^2 + \tau^2) - \frac{(y_i - \mu)^2}{2(\sigma_i^2 + \tau^2)} + \log p(a, b, s_i, y_i) - \log \Phi(u_i). \tag{7}$$

We assume that the number of patients in the ith study is sufficiently large for s_i^2 to be an accurate estimate of the left hand side of (6), so that the value of σ^2 can be deduced by equating s_i^2 with the right hand side of (6).

It is clearly impossible to estimate a and b since we do not know how many unselected studies there are in the population. But for any *given* pair (a, b) this likelihood function can be maximized to give estimates of μ, τ and ρ. The first stage of the sensitivity analysis is to examine the dependence of $\hat{\mu}$ on a and b. The fitted model can be judged by substituting these estimates into formula (5) — evaluating this for each study gives the fitted or expected values of y_i.

These expected values give a good fit to the data if they satisfactorily predict any observed trend in the funnel plot. Copas and Shi[8] show how the fit of the model can be tested by refitting the same model but with a linear term in s added to the expected value of y. In other words we extend the model (2) to

$$y_i = \mu_i + \beta s_i + \sigma_i \epsilon_i$$

but keep the other features of the model exactly as before. Then if we accept the null hypothesis that $\beta = 0$ then we accept that the selection model has satisfactorily explained any apparent linear relationship between y and s. By comparing the maximum of the likelihood with the β term added with the maximum of the original likelihood with $\beta = 0$, we can calculate a P-value by the likelihood ratio test in the standard way. Only if this P-value is sufficiently large, say over 5%, can we accept the selection model as giving a reasonable explanation of what we have observed.

Often it will be of interest to test the overall significance of the treatment effect by testing the null hypothesis that $\mu = 0$. This is done by another likelihood ratio test, this time comparing the maximum value of $L(\mu, \rho, \tau, a, b)$ with its maximum when $\mu = 0$. This gives a P-value for the significance of μ and hence, assuming asymptotic normality, an associated confidence interval.

4 Sensitivity analysis: application

The first step in the sensitivity analysis is to identify a range of selection models which cover all reasonable possibilities. We assume that

$$0.01 \leq P(\text{select}|s) \leq 0.99 \tag{8}$$

for all the values of s observed in the meta analysis, constrained by the requirement that the selection probabilities cannot decrease as s decreases. This can be converted into a range of values of a and b by solving equation (1) for the two end points of the interval in (8) and for the smallest and largest observed value of s, only retaining those pairs (a, b) with $b \geq 0$.

This region of values of a and b is now covered by a grid of individual (a, b) pairs, and at each grid point we can calculate the quantities discussed in the last section. It is also useful to add the point $(a = 10, b = 0)$ to represent the usual random effects model with no allowance for selection. We can now plot the values of $\hat{\mu}$ against the corresponding P-values for the fit to the funnel plot. Of course this will not give an exact curve, but will indicate how different patterns of selection affect these two quantities jointly. The basic inference is that any value of $\hat{\mu}$ associated with a P-value exceeding 5% (say) can be taken as a reasonable estimate of the treatment effect in the light of the available evidence.

Examples show that this plot has a characteristic shape. The estimates $\hat{\mu}$ tend to decrease as the P-value increases. The standard random effects estimate gives the largest $\hat{\mu}$ and the smallest P-value. If there is a significant trend in the funnel plot (as judged by the Egger test, for example) then the P-value for the standard estimate will likely be small and thus be discredited by the sensitivity analysis. Finding the largest value of $\hat{\mu}$ with $P \geq 0.05$ is a good way of describing the *minimum* downwards adjustment to the treatment effect which is needed to allow for publication bias. If there is no clear trend in the funnel plot, then all values of $\hat{\mu}$ in this plot are likely to be retained as reasonable possibilities.

The fit of the model at each grid point in the (a, b) range can also be illustrated by calculating fitted values from equation (5) and superimposing on the funnel plot. In the examples below we plot just two different model fits, the case $(a = 10, b = 0)$ giving the horizontal line at the standard random effects estimate, and a case in the middle of the (a, b) range, giving a curve which hopefully gives a nice representation of the trend in the funnel plot.

As well as looking at these plots, a concise summary of the sensitivity analysis in tabular form is also useful. We suggest taking five or six "representative" pairs from the (a, b) range, at one extreme the point $(a = 10, b = 0)$ and at the other extreme a point in which all the selection probabilities are small. Then for each we can summarize the inference by tabulating the following eight quantities:

1. $\hat{\mu}$
2. P-value for testing $H_0 : \mu = 0$
3. lower limit of the 95% confidence interval for μ
4. upper limit of the 95% confidence interval for μ
5. P-value for fit to the funnel plot
6. $P(\text{select}|s_{min})$
7. $P(\text{select}|s_{max})$
8. the estimated number of published and unpublished studies given by

$$\sum_i \{P(\text{select}|s_i)\}^{-1}.$$

The last quantity gives an indication of the overall severity of the selection model. With no selection (e.g. at $a = 10$ and $b = 0$) this will just be the number of studies in the meta analysis. If there is substantial selection then this number will be much larger. In the terminology of Section 2, this is an estimate of the size of the study population from which the observed studies are assumed to be a sample. This therefore gives us an idea of the overall sampling fraction involved.

The Appendix lists some *S-plus* code which can be used to carry out all of these calculations. To save space we have added only brief comment lines to these routines, but as far as possible have tried to use variable names to correspond to the above notation. We end this section with a commentary on each of the routines, explaining what each does and how they are used.

There are two main *S-plus* functions, which call on four subroutines. To use these it is necessary to input the values of y_i and s_i into vectors y and s respectively. We also need to set a 4-component starting vector $start0$, for example

$$start0 = (0, 0.5, 0.05, 0),$$

and to set some values of the selection probability pairs

$$\{P(\text{select}|s = s_{min}), \quad P(\text{select}|s = s_{max})\} \tag{9}$$

into a 2-column matrix *sp*. These define the selection of (a, b) pairs used in the tabular summary of the sensitivity analysis. In our examples we use five pairs

$$(.99, .80), (.8, .5), (.6, .3), (.4, .1) \text{ and } (.2, .01)$$

but for a more detailed analysis more pairs could be chosen.

The four subroutines are

- *tran* — this converts a 2-column matrix of pairs such as (9) into the corresponding 2-column matrix of values of a and b.
- *lik* — this calculates the likelihood function for parameter vector

  $$par = (\mu, \rho, \tau, \beta),$$

  with an argument *pind* which has two binary elements controlling whether parameters μ and/or β are set to zero.
- $maxlik1$ — this finds maximum likelihood estimates at a given pair (a, b), using $start0$ as the starting value in the optimization algorithm. This function also outputs the value of the likelihood, the value of (9) and the sum of the recipricols of $P(\text{select}|s_i)$.
- $maxlik2$ — this takes a 2-column matrix of values of a and b and runs $maxlik1$ for all these pairs, first with $\mu = \beta = 0$, then with $\beta = 0$, and thirdly with no restrictions on the parameter estimation. The output is a matrix containing all of these results.

The two main functions are

- $scode1$ — this plots the funnel plot and superimposes the two lines discussed earlier in this section. The numerical output from this function is the table summarizing the sensitivity analysis. The first row is the fit of the standard random effects model (no allowance for publication bias) and the subsequent rows correspond to the choices of (9) given in the input matrix *sp*. The columns are the eight statistical summary measures described above.
- $scode2$ — this divides each of the ranges (8) for $s = s_{min}$ and $s = s_{max}$ into n equal parts, and finds the resulting 2-column matrix of all the possible values of a and b. As default we take $n = 10$ but a larger value of n can be taken at the expense of greater run time. This function then plots $\hat{\mu}$ against the P-values measuring quality of fit to the funnel plot. This plot adds the point corresponding to the standard random effects analysis (with a different plotting character) and draws a vertical line at $P = 5\%$.

5 Examples

5.1 Passive smoking and coronary heart disease

He *et al.*[12] reviewed 10 cohort studies and 8 case-control studies, each designed to estimate the increase in risk of coronary heart disease resulting from prolonged exposure to environmental tobacco smoke (passive smoking). Here we take y to be the log relative risk. Obviously there are substantial questions here about the comparability of the studies, consistency of the exposure measures and so on, but we follow these authors in arguing that the 18 estimates in this review are sufficiently comparable for a meta analysis to be meaningful. See the original paper for a discussion of these and related issues. Ignoring the possibility of publication bias, as these authors do, gives the standard random effects estimate $\hat{\mu} = 0.25$, an excess risk of 28%.

Using the *S-plus* coding in the Appendix, function *scode1* gives the funnel plot in Figure 1. Clearly there is a strong positive trend in the funnel plot in this meta analysis, well fitted by the dashed line corresponding to the fitted model for an (a, b) pair near the centre of the range. There is an outlier towards the top right of the plot, but in fact this study is so much smaller than the others that it is given very little weight in any of the analyses. The numerical output from *scode1* gives the entries in Table 1. As the number of unpublished studies increases (column 8, reading downwards) the estimate of log relative risk (column 1) falls sharply. However the precision of these studies is sufficiently strong for the overall evidence of a positive increase in risk (column 2) to remain clear over all of the selection possibilities considered in the sensitivity analysis. The first two rows of the table can be discounted on the grounds of a poor fit to the funnel plot (column 5), suggesting that the standard estimate of $\hat{\mu} = 0.25$ is an overestimate, possibly a substantial overestimate, of the risk.

Function *scode2* gives the plot in Figure 2. The slightly odd shape of these points is an artifact of the regular grid over the range being considered — increasing n in the function would produce a more uniform cloud of points. The vertical line at $P = 0.05$ divides off the points with the higher estimates of μ — only the points with $\hat{\mu}$ less than about 0.20 can be considered as reasonably consistent with the data.

The conclusion from the sensitivity analysis is that a 28% excess risk is too high as an estimate from these studies, but that the contention that there is *at least some* increase in risk seems robust to publication bias. An extreme model (the last row of Table 1), reflecting very strong publication bias, gives an estimated risk only half as high as the superficial analysis, but the P-value remains less than 0.01.

5.2 RCTs for the effectiveness of antibiotic prophylaxis

D'Amico *et al.*[13] discuss a number of different meta analyses of randomized clinical trials on the effectiveness of antibiotic prophylaxis in critically ill adults. All of these reviews provide good illustrations of the method of this paper, some showing substantial evidence of publication bias, others less so. As a contrasting example to the passive smoking review, we reanalyse these authors' meta analysis of estimates of the reduction in mortality. Here y is the log odds ratio calculated from the number of deaths and the number of patients in treatment and control groups, estimated in 17 different studies.

Figures 3 and 4 and Table 2 are the exact analogues of the graphs and table referred to under the first example. The funnel plot in Figure 3 shows no clear trend, providing little or no evidence for the presence (or absence) of publication bias. All of the models forming the points in Figure 4 give a very acceptable fit to the funnel plot (no small P-values) and so $\hat{\mu}$ can be anywhere between 0.17 and the conventional estimate of 0.22. The main effect of the possibility of publication bias is shown in Table 2, and that is an *increase in uncertainty.* The lower confidence limit for μ crosses zero between the second and third rows, so although the conventional analysis reported by these authors gives a significant reduction in mortality ($P = 0.016$), only two or three unpublished papers are needed to induce sufficient publication bias for the overall reduction in mortality to be no longer significant. The numerical estimate of μ does not change substantially, but its uncertainty does.

6 Comments

All statistical models make assumptions, and even a sensitivity analysis makes assumptions which need to be recognised and assessed. The important assumptions of our model are equations (2) and (3). The first equation asserts that any trend in the funnel plot is the result of study selection, and not of any intrinsic link between y and s. This could be mistaken, for example smaller studies might be conducted more carefully and so produce an enhanced treatment effect. The opposite effect could also be envisaged.

We have assumed that the results of each study are effectively summarized by a single estimate y and its standard error. Often there will be corresponding observations on covariates or confounding factors. We have assumed that y has been appropriately corrected for such values, but in principle the model can be extended by replacing μ_i in (2) by a linear regression on covariates. Such extended models have been called "meta regression". If the selection equation (1) still applies then the maximum likelihood analysis for given values of a and b extends in the straightforward way. However if selection also depends on covariates, so that the linear term $c^T x$ is added to the probit in (1), then the sensitivity region of (a, b) would become a region of values of (a, b, c), which would be more complicated. If there are many covariates, then for this to be a practical procedure some restrictions on the possible values of c would be needed.

Equation (3) asserts that the population of studies, assuming they could all be observed, have true values symmetrically distributed about the overall value we are trying to estimate. This is the standard random effects model for heterogeneity, but it raises the question of whether we really want to estimate an average effect or whether it is to assess the effectiveness of the treatment if adopted under a particular set of circumstances, such as in a particular country or a particular clinic. The presence of heterogeneity means that these are not the same thing. As we usually have a relatively small number of studies in a meta analysis, the assumption of normality in (3) is arguably less important.

Assumption (1) seems a very special model, but trying alternatives to this specification suggest that the important thing is that we allow for differential selection between studies of different sizes, and not that we do so in this particular way. For example, within a broader setting, Copas and Li [10] discuss a number of different formulations of the selection model.

We would argue that a sensitivity analysis is useful even if these assumptions are not entirely plausible. The standard random effects analysis makes some very strong assumptions, and if the sensitivity analysis shows that our conclusion could be materially different if some of these assumptions are relaxed, than that is reason enough for recommending that the standard analysis should be viewed with caution. The question of whether the class of models underlying the sensitivity analysis contains the "true" model is not the central issue.

The objective of statistical analysis is not to produce definitive estimates and P-values, but to describe and interpret the evidence in the data which is available to us. As we have no control on the sampling of the studies under review, it is impossible to estimate a single model without bringing in assumptions from outside the data. Publication bias is perhaps the clearest example of this dilemma. We claim no special optimality for the sensitivity analysis suggested here, only that it makes explicit the uncertainty inherent in systematic reviews, an uncertainty which in our view is glossed over by any attempt to overcome the problem of publication bias by ignoring it, correcting for it, or testing for it.

Acknowledgement

One of us (JQS) is supported by a grant from the *ESRC*.

References

1. Hutton JL, Williamson PR. Bias in meta analysis due to outcome variable selection within studies. *Applied Statistics*, **49**, to appear.

2. Yusuf S, Koon T, Woods K. Intravenous magnesium in acute myocardial infarction: an effective, safe, simple and inexpensive intervention. *Circulation*, 1993; **87**: 2043-2046.

3. ISIS-4 Collaborative Group. A randomized factorial trial assessing early oral captopril, oral mononitrate and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet*, 1995; **345**: 669-685.

4. Egger M, Smith GD. Misleading meta-analysis. *British Medical Journal*, 1995; **310**: 752-754.

5. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. The impact of publication bias on meta-analyses within the cochrane database of systematic review. *Technical Report, Dept of Epidemiology and Public Health, U. of Leicester, UK.* 1999.

6. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 1997; **315**: 629-634.

7. Sutton AJ, Song F, Gilbody SM, Abrams KR. Modelling publication bias in meta-analysis: a review. *Technical Report, Dept of Epidemiology and Public Health, U. of Leicester, UK.* 2000.

8. Copas JB, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, 2000 (in press).

9. Copas JB. What works?: selectivity models and meta-analysis. *J. R. Statist. Soc. A*, 1999; **162**: 95-109.

10. Copas JB, Li HG. Inference for non-random sample (with discussion). *J. R. Statist. Soc. B*, 1997; **59**: 55-95.

11. Whitehead A, Whitehead J. A general parametric approach to the meta analysis of randomized clinical trials. *Statistics in Medicine*, 1991; **10**: 1665-1677.

12. He J, *et al.* Passive smoking and the risk of coronary heart disease – A meta analysis of epidemiologic studies. *The New England J. of Medicine*, 1999; **340**: 920-926.

13. D'Amico R, *et al.* Effectiveness of antibiotic prophylaxis in critically ill adult patients: systematic review of randomised controlled trails. *British Medical Journal*, 1998; **316**: 1275-1285.

Appendix: Splus codes

```
—scode1
function()
{
# This function plots the funnel plot and calculates the summary table
ab ← rbind(c(10, 0), tran(ps))
da ← maxlik2(ab)
mu ← da[, 10]
plot(s, y)
lines(c(min(s), max(s)), c(mu[1], mu[1]), lty = 2)
a1 ← ab[4, 1] + ab[4, 2]/s
a2 ← dnorm(a1)/pnorm(a1)
a3 ← da[4, 11]
bb ← cbind(s, mu[4] + (a3 * a2 * s)/sqrt(1 - a3∧2 * a2 * (a1 + a2)))
cc ← bb[order(bb[, 1], bb[, 2]), 1:2]
lines(cc[, 1], cc[, 2], lty = 4)
lik1 ← 2 * (da[, 1] - da[, 9])
lik2 ← 2 * (da[, 9] - da[, 17])
temp ← (abs(mu) * 1.96)/sqrt(lik1)
cbind(mu, 1 - pchisq(lik1, 1), mu - temp, mu + temp, 1 - pchisq(lik2, 1), da[, 6], da[, 7], da[, 8])
}
```

```
—scode2
function()
{
# This function plots the μ̂ against the P-value
n ← 10
temp ← seq(0.01, 0.99, length = n)
ab ← cbind(rep(temp, rep(n, n)), rep(temp, n))
ab ← rbind(c(10, 0), tran(ab[ab[, 1] >= ab[, 2], ]))
da ← maxlik2(ab)
mu ← da[, 10]
temp ← 1 - pchisq(2 * (da[, 9] - da[, 17]), 1)
plot(temp, mu, xlab = "P-value", ylab = "mu", xlim = c(0, 1))
points(temp[1], mu[1], pch = 1)
lines(c(0.05, 0.05), c(min(mu), max(mu)), lty = 2)
}
```

```
—maxlik1
function(ab, pind)
{
# This function calculates individual maximum likelihood estimates
temp ← nlminb(start = start0, objective = lik, upper = c(Inf, 1, Inf, Inf), lower = c(-Inf, 0, -Inf,
        -Inf), gamma = ab, pind = pind)
p ← pnorm(ab[1] + ab[2]/s)
c(temp$objective, temp$parameters, max(p), min(p), sum(1/p) %/% 1)
}

—maxlik2
function(ab)
{
#This function calculates the whole set of MLEs
c1 ← apply(ab, 1, maxlik1, pind = c(0, 0))
c2 ← apply(ab, 1, maxlik1, pind = c(1, 0))
c3 ← apply(ab, 1, maxlik1, pind = c(1, 1))
t(rbind(c1, c2, c3))
}

—lik
function(par, gamma, pind)
{
# This function defines the log likelihood function
u ← gamma[1] + gamma[2]/s
mu ← par[1] * pind[1]
rho ← par[2]
tau ← par[3]
beta ← par[4] * pind[2]
lambdu ← dnorm(u)/pnorm(u)
rhos ← rho∧2 * lambdu * (u + lambdu)
sigma ← s∧2/(1 - rhos)
rhoi ← (rho * sqrt(sigma))/sqrt(sigma + tau∧2)
v ← (u + (rhoi * (y - mu - beta * s))/sqrt(tau∧2 + sigma))/sqrt(1 - rhoi∧2)
sum(0.5 * log(tau∧2 + sigma) + ((y - mu - beta * s)∧2)/(2 * (tau∧2 + sigma)) - log(pnorm(v)))
}

—tran
function(p)
{
# This function converts selection probabilities into values of a and b
q1 ← qnorm(p[, 1])
q2 ← qnorm(p[, 2])
b ← (q1 - q2)/(1/min(s) - 1/max(s))
a ← q1 - b/min(s)
cbind(a, b)
}
```

Table 1. Passive smoking and coronary heart disease data: output of *scode*1

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]	[, 6]	[, 7]	[, 8]
[1,]	0.2498851	3.293093e-06	0.14459620	0.3551740	0.0005001727	1.00	1.00	18
[2,]	0.2071571	5.090268e-06	0.11813767	0.2961765	0.0069888939	0.99	0.80	20
[3,]	0.1649526	2.553092e-04	0.07654346	0.2533618	0.4610171523	0.80	0.50	31
[4,]	0.1356235	4.036167e-03	0.04317569	0.2280714	0.8794559445	0.60	0.30	49
[5,]	0.1313016	4.576821e-03	0.04053970	0.2220636	0.6673465736	0.40	0.10	125
[6,]	0.1314307	4.630813e-03	0.04045928	0.2224021	0.9418290413	0.20	0.01	912

Table 2. Antibiotic prophylaxis data: Output of *scode*1

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]	[, 6]	[, 7]	[, 8]
[1,]	0.2213620	0.01560341	0.04193670	0.4007874	0.7754968	1.00	1.00	17
[2,]	0.2136090	0.02817267	0.02286181	0.4043562	0.8111425	0.99	0.80	18
[3,]	0.1902500	0.15105143	-0.06944911	0.4499491	0.9149076	0.80	0.50	26
[4,]	0.1825510	0.24529215	-0.12539770	0.4904998	0.9462107	0.60	0.30	40
[5,]	0.1816679	0.26417895	-0.13722218	0.5005580	0.9590936	0.40	0.10	88
[6,]	0.1823535	0.25877165	-0.13413152	0.4988385	0.9662029	0.20	0.01	500

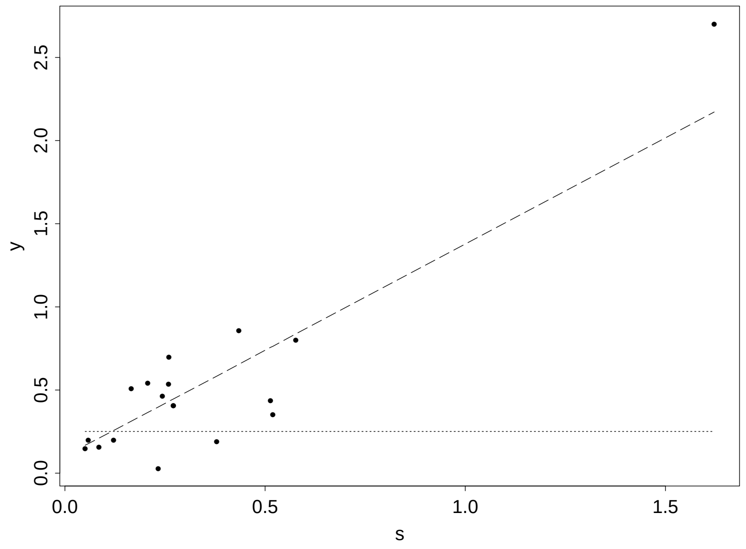

Figure 1: Passive smoking and coronary heart disease data: Funnel plot with fitted values — output of $scode1$. The dotted line is the usual random effects model, the dashed line is the selection model.

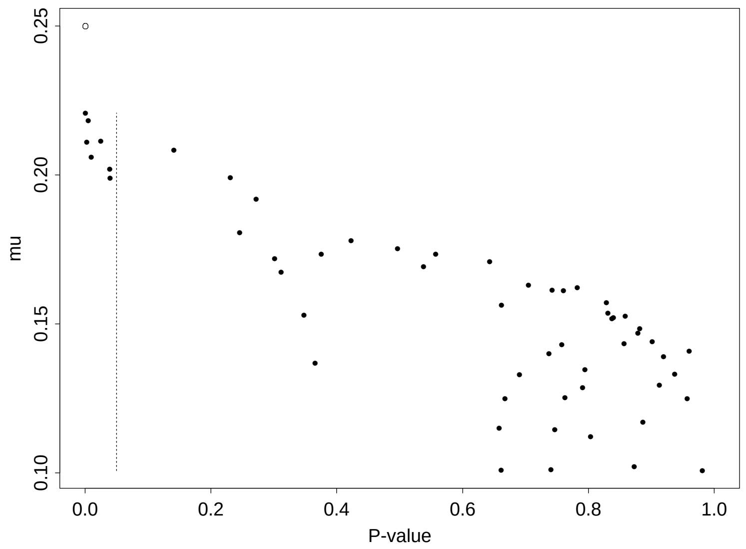

Figure 2: Passive smoking and coronary heart disease data: $\hat{\mu}$ against the P-value of $H_0 : \beta = 0$ — output of $scode2$. The dotted line is P=0.05.

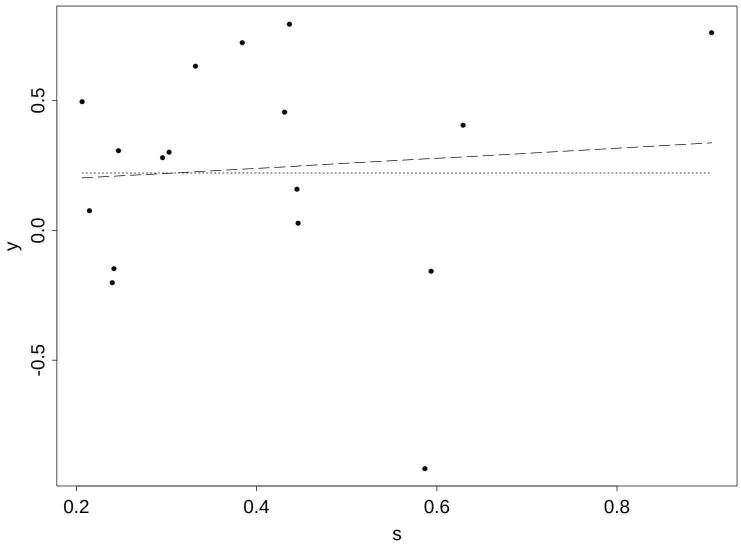

Figure 3: Antibiotic data — reduction in mortality for CA1: Funnel plot with fitted values — output of $scode1$. The dotted line is the usual random effects model, the dashed line is the selection model.

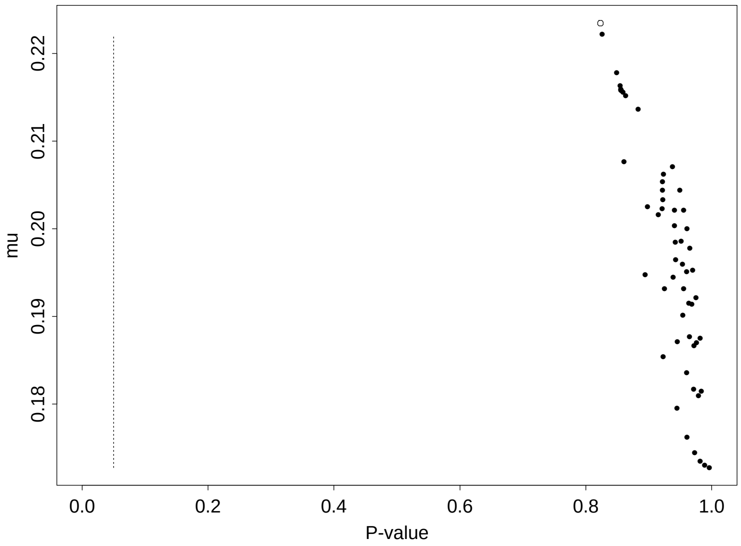

Figure 4: Antibiotic data — reduction in mortality for CA1: $\hat{\mu}$ against the P-value of $H_0 : \beta = 0$ — output of $scode2$. The dotted line is P=0.05.