Meta-Analysis for Trend Estimation

Jian Qing Shi[1*] and J. B. Copas[2]

[1] *School of Mathematics and Statistics, University of Newcastle, Newcastle NE1 7RU, UK.*

[2] *Department of Statistics, University of Warwick, Coventry CV4 7AL, UK*

SUMMARY

Grouped dose measures, heterogeneity and publication bias are three major problems for meta-analysis in trend estimation. In this paper, we propose a model which allows for arbitrarily aggregated dose levels, and show that the resulting estimates and standard errors can be quite different from those given by the usual assigned value method. Based on fitting a model to the funnel plot, we discuss a method for random effects sensitivity analysis which deals with the problems of heterogeneity and publication bias. A meta-analysis of epidemiological studies on the effect of alcohol on the risk of breast cancer is used to illustrate the method. Our analysis suggests that the rate of increase in risk with alcohol consumption is substantially less than has been previously suggested.

Keywards: Grouped dose levels; Heterogeneity; Meta-analysis; Publication bias; Trend estimation.

* *Address for correspondence*: School of Mathematics and Statistics, University of Newcastle, Newcastle NE1 7RU, UK. Email: j.q.shi@ncl.ac.uk

1 Introduction

In epidemiological studies of the association between disease and exposure to some agent or hazard, it is often of interest to estimate how much risk increases as exposure increases. For example, in smoking and lung cancer, we would like to know how much the risk increases as we increase the number of cigarettes smoked per day. Alcohol use and breast cancer is another example. It would be meaningless to try and quantify the association between the incidence of breast cancer and whether of not alcohol has *ever* been consumed: the risk clearly depends on the actual amount of alcohol. Studies which measure the risk at different levels of exposure are usually analysed by trend estimation. Meta-analysis is used to combine together the evidence from several such studies in order to get a clearer assessment of the nature of the association.

The usual approach to trend estimation is to estimate a log-odds ratio (or adjusted log-odds ratio) at each of several dose levels, and to fit a weighted linear regression through the origin. In practice the dose levels are often not recorded exactly but are grouped into class intervals — the usual method is to assign a single representative dose value for each category. Given several such studies, the resulting regression slopes and their standard errors are combined together into a single trend estimate using the standard methods of meta analysis.

There are (at least) three important statistical problems which arise in meta analysis for trend estimation. Firstly, as already mentioned, is the problem of *grouped dose levels*. If the class intervals cover a wide range of levels of exposure, the assigned value method is clearly unsatisfactory — different choices of assigned values will give different estimates. Secondly, as in meta analysis more generally, there is the problem of *heterogeneity*, recognising that the design quality and sample selection can be very variable between the different studies (see, e.g., Thompson and Sharp [1] and Copas [2]). The third problem is *publication bias*, recognising that the studies used in meta-analysis are themselves selected. It is almost certainly the case that studies with significant results and/or large sample sizes are more likely to be published, or more likely to be published in journals accessible to the reviewer, than studies with inconclusive outcomes or small sample sizes [3]. The aim of our paper is to develop a method for the meta analysis of trend estimates which address these three

problems.

Section 2 introduces our motivating example, a meta analysis of thirteen case-control studies on the association between alcohol consumption and the risk of breast cancer. A conventional fixed-effect meta analysis of these data gives a trend of around 0.01 as the regression slope of cancer log-odds on the consumption of alcohol in grams per day. In practical terms this means that, very roughly, one extra drink per day increases risk by about 16%. Our aim is to review this figure using more appropriate statistical methods.

The methodology leading to our method is set out in Section 3. Section 3.1 discusses a basic model which allows for heterogeneity between studies and which takes into account the correlation between log-odds ratios caused by the fact that each study uses a single estimate of base level risk. Section 3.2 adapts the model to allow for grouped exposure levels by assuming that the actual dose levels follow a distribution across the different subjects in the studies. In Section 3.3 we show how to carry out a sensitivity analysis for publication bias following our previous work in [4] and [5].

In Section 4 we revisit the example of alcohol and breast cancer. Allowing for heterogeneity and grouped dose levels reduces the overall trend estimate very considerably, from about 0.01 to about 0.005. The funnel plot for these data gives evidence of publication bias. Our sensitivity analysis suggests that only a few 'unobserved studies' are needed for the estimated trend to reduce still further. The example demonstrates that the problems we discuss in this paper are not just minor refinements to theory, but can have a substantial practical impact.

Some final comments are given in Section 5.

2 A motivating example

Table 1, taken from Rohan and McMichael [6], reports data from a case-control study investigating the association between the consumption of alcohol and the risk of breast cancer.

Table 1. Case-control data on alcohol use and breast cancer

Alcohol (g/day)	Assigned dose x	No. of cases	No. of controls	Crude OR	Adjusted OR (CI)
0	0	165	172	1.0	1.0
< 2.5	2	74	93	0.83	0.80 (0.51-1.27)
2.5-9.3	6	90	96	0.98	1.16 (0.73-1.85)
> 9.3	11	122	90	1.41	1.57 (0.99-2.51)

Here, for each of 451 cases and an equal number of controls, we have an estimate of alcohol consumption grouped into one of four exposure bands, including the base line group with zero dose. Each row gives an estimate of the odds of being a case versus being a control, and hence the estimates of the odds ratios reported in the fifth column. For example, $1.41 = (122/90)/(165/172)$. In practice, as in this study, we usually adjust these estimates to allow for other covariates known to be related to risk, such as age. The sixth column in Table 1 gives the adjusted odds ratios and their confidence limits. The standard errors of the corresponding adjusted log odds ratios can be deduced from the limits of the reported confidence intervals in the obvious way.

If the jth dose class can be assumed to have a single dose level x_j, and if the corresponding log odds ratio (or adjusted log-odds ratio) y_j has standard error s_j, the usual approach of trend estimation is to fit a linear regression through the origin,

$$y_j = \beta x_j + s_j \epsilon_j, \quad j = 1, \cdots, n, \tag{1}$$

where n is the number of non-zero exposure levels, equal to 3 for Table 1, and the ϵ_js are assumed to be standard normal residuals. A crude application of least squares gives the estimated dose-response regression slope $\hat{\beta}$, and its variance, as

$$\hat{\beta} = \frac{\sum_j x_j y_j / s_j^2}{\sum_j x_j^2 / s_j^2}; \quad s_\beta^2 = \frac{1}{\sum_j x_j^2 / s_j^2}. \tag{2}$$

Taking the x_js to be the assigned dose levels listed in the second column of Table 1, formulae (2) gives $\hat{\beta} = 0.033$ with standard error 0.017, suggesting a marginally significant association between risk and exposure to alcohol.

Greenland and Longnecker [7], and other papers by these authors (see the references at the end of this paper), discuss the meta analysis of several different studies similar to that in Table 1. If there are m such studies, formulae (2) give individual study estimates $\hat{\beta}_i$ and s_{β_i} for $i = 1, 2, \cdots, m$. A fixed effects meta analysis, assuming that the true value of β is

constant across the studies, gives the pooled estimate of the regression slope and its variance to be

$$\hat{\beta} = \frac{\sum_i \hat{\beta}_i / s_{\beta i}^2}{\sum_i 1/s_{\beta i}^2} = \frac{\sum_{ij} x_{ij} y_{ij} / s_{ij}^2}{\sum_{ij} x_{ij}^2 / s_{ij}^2} \quad \text{and} \quad var(\hat{\beta}) = \frac{1}{\sum_i 1/s_{\beta i}^2} = \frac{1}{\sum_{ij} x_{ij}^2 / s_{ij}^2}. \tag{3}$$

Here, in the obvious notation, x_{ij}, y_{ij} and s_{ij} are the values of x_j, y_j and s_j specific to the ith study. This fixed effects estimate is often used as the overall estimate of trend in meta-analysis. See Section 4 below for details of $m = 13$ case-control studies in the cancer/alcohol application.

There are a number of statistical problems with this analysis. First, and most obviously, is the implied assumption of independence in the variance formula in (2). Greenland and Longnecker [7] suggested the use of generalized least squares to allow for the correlation between log-odds ratios — this correlation results from the fact that the y_js in each study use the same base-line estimate for the unexposed group. Their approach is followed in Section 3.1 below.

The second problem is the grouped dose levels and the need for assigned values x_{ij} in calculating the least squares estimates. Table 1 uses assigned values 2, 6, 11 for the categories $0 < x < 2.5$, $2.5 \leq x \leq 9.3$ and $x > 9.3$ respectively, leading to the trend $\hat{\beta}_i = 0.033$ for this particular study. But if we take the assigned values instead as 1.25, 5.24 and 25.01, which are the mean dose levels in each category using the model suggested in Section 3.2, $\hat{\beta}_i$ becomes 0.018. Clearly this difference cannot be ignored. The fewer categories there are the more sensitive the trend estimate is to the choice of assigned values. Accurate estimation of the variances of the $\hat{\beta}_i$s is also important, because they are used as weights in meta-analysis.

The third problem is the assumption implicit in (3) that each study is estimating the same underlying trend. Between study heterogeneity is usually resolved by assuming that the true values follow a random effects model — this is the approach we follow in Section 3.

The fourth problem is the presence of publication bias, immediately evident from the funnel plot in Figure 1(i). This plots the $m = 13$ values of $\hat{\beta}_i$ calculated by the method of Section 3.2 against their corresponding standard errors. There is clear evidence that the small studies (those with large $s_{\beta i}$) tend to give more positive outcomes than the large studies (those with small $s_{\beta i}$), just as one would expect of publication bias. If we assume that small studies are less likely to be published than large studies, then the plot suggests there may

be several unpublished small studies which give lower trend estimates than those observed, and so ignoring them will lead to an overestimate of β. This is examined in Section 3.3.

3 Methodology for meta-analysis in trend estimation

3.1 Dose analysis model with heterogeneity

As above, suppose there are m studies in the meta-analysis, with the log-odds ratios (or their adjusted values) for the individual studies satisfying model (1). The assumptions that the slopes are equal, and that the log odds ratios are uncorrelated within each study, both need to be relaxed. To model heterogeneity between studies, assume that

$$\beta_i \sim N(\beta, \tau^2). \tag{4}$$

To allow for within-study correlation between log-odds ratios, we follow the discussion in Greenland and Longnecker [7] and suppose that the correlation between y_{ij} and y_{ik}, the jth and kth log odds ratio in the ith study, is

$$r_{i,jk} = v_{i0}/(v_{ij}v_{ik})^{1/2}$$

for $j \neq k$, where

$$v_{i0} = 1/C_{i0} + 1/D_{i0}, \quad v_{ij} = 1/C_{i0} + 1/D_{i0} + 1/C_{ij} + 1/D_{ij}$$

and (C_{i0}, D_{i0}) and (C_{ij}, D_{ij}) are the numbers of cases and controls respectively for the unexposed group ($x = 0$) and the exposed group with assigned dose x_{ij}. Defining $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{in_i})'$, let

$$\boldsymbol{\Omega}_i = var(\boldsymbol{y}_i|\beta_i), \tag{5}$$

a matrix with jth diagonal element equal to s_{ij}^2 and (j,k)th off-diagonal element equal to $r_{i,jk}s_{ij}s_{ik}$. The log-likelihood for the unknown parameters (β, τ) is therefore

$$L(\beta, \tau) = \sum_{i=1}^{m} \left\{ -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\boldsymbol{y}_i - \beta\boldsymbol{x}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \beta\boldsymbol{x}_i) \right\} \tag{6}$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \tau^2\boldsymbol{x}_i\boldsymbol{x}_i'$ is the marginal covariance matrix of $\boldsymbol{y}_i$, with

$$\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{\Omega}_i^{-1} - \boldsymbol{\Omega}_i^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\Omega}_i^{-1}/(1 + \tau^2\boldsymbol{x}_i'\boldsymbol{\Omega}_i^{-1}\boldsymbol{x}_i). \tag{7}$$

The maximum likelihood estimate (MLE) of (β, τ) can be calculated by maximising $L(\beta, \tau)$, and homogeneity ($H_0 : \tau = 0$) can be tested by the corresponding likelihood ratio test.

The above method is approximately equivalent to the following two step procedure. Firstly, we calculate the estimate of β_i for the ith study as $\hat{\beta}_i = s^2_{\beta i} \boldsymbol{x}_i{}' \boldsymbol{\Omega}_i^{-1} \boldsymbol{y}_i$ and $s^2_{\beta i} = (\boldsymbol{x}_i{}' \boldsymbol{\Omega}_i^{-1} \boldsymbol{x}_i)^{-1}$. Then we assume that, conditionally on the true value β_i, $\hat{\beta}_i \sim N(\beta_i, s^2_{\beta i})$ so that, marginally,

$$\hat{\beta}_i \sim N(\beta, s^2_{\beta i} + \tau^2). \tag{8}$$

The overall MLE of (β, τ) is calculated by maximising

$$\sum_{i=1}^{m} \left\{ -\frac{1}{2} \log(\tau^2 + s^2_{\beta i}) - \frac{1}{2} (\hat{\beta}_i - \beta)^2 / (\tau^2 + s^2_{\beta i}) \right\}.$$

The standard error s_β is calculated by $s^2_\beta = 1/\sum(\tau^2 + s^2_{\beta i})^{-1}$, where τ is evaluated at its MLE. This method is simple and easy to use in practice, and can be extended to allow for grouped vales of x_{ij} as discussed in the next subsection.

3.2 Grouped exposure levels

As in Table 1, the exposure levels for each subject are often not recorded exactly but grouped into class intervals. We suppose that exposure (or dose) is an underlying continuous covariate belonging to observable intervals J_{ij}, so that each subject in the jth group in the ith study has $x \in J_{ij}$.

The customary method of analysis is to assign a single value x_{ij} as representative of J_{ij}. Longnecker [8] used historical data to obtain the population distribution of exposure levels, and then calculated the assigned values as the medians or means within each category. There are two obvious disadvantages with this method. Firstly the distribution obtained from historical data may not be that of the population from which the particular subjects in a study are selected. For example, Longnecker [8] used historical data for US women, but about half of the 39 studies being considered were outside the US (Longnecker recognised the problem by stating (p.79) *"US women, in general, do not drink much"*). We suggest that historical data can be used as prior information, but that the observed frequencies of J_{ij} in each study should inform us about the distribution of x appropriate to that study. The second disadvantage, as mentioned earlier, is that using a single assigned value for each

J_{ij} leads to inaccurate estimates and underestimation of variance.

Suppose the exposure levels of all individuals in a particular study are sampled from some distribution with probability density function $f(x)$. Then if the probability of being a case given dose x is p_x, the probability that an individual in class interval J is a case is

$$\frac{\int_J p_x f(x)dx}{\int_J f(x)dx}. \tag{9}$$

The observed number of individuals in this class interval who are cases therefore follows a binomial distribution with this as the probability of success. Using the derivation given in Appendix I, the likelihood for β_i is

$$L_i = \log l_i = -\frac{1}{2}\log|\boldsymbol{\Omega}_i| - \frac{1}{2}\left\{\boldsymbol{y}_i - \boldsymbol{A}_i(\beta_i)\right\}'\boldsymbol{\Omega}_i^{-1}\left\{\boldsymbol{y}_i - \boldsymbol{A}_i(\beta_i)\right\}), \tag{10}$$

where $\boldsymbol{A}_i(\beta_i) = (A_{ij}(\beta_i))$ is given by (20), and $\boldsymbol{\Omega}_i$ is defined by (5). It is fairly easy to calculate the integrals involved in L_i because they are all one-dimensional. For instance Monte Carlo and Gaussian quadrature [9] are all efficient approaches to calculating these integrals. Once the MLE of β_i and its standard error are obtained, model (8) is used to calculate the overall pooled estimate in the same way as before.

To implement this procedure we need to know the dose distribution $f(x)$. We suggest using a parametric model fitted to the observed frequencies of the dose intervals. Let n_{ij} be the number of subjects in J_{ij}, and suppose that the interval J_{ij} is (J_{ij*}, J_{ij}^*) (taking the end points as 0 or ∞ when needed). One possible model is to assume that the dose levels (or transformations of them) have a normal distribution $x \sim N(\xi, \lambda^2)$. Under this model, the frequencies n_{ij} give the log likelihood

$$\sum_{i,j} n_{ij}\log P(x \in J_{ij}) = \sum_{i,j} n_{ij}\log\left\{\Phi(\frac{J_{ij}^* - \xi}{\lambda}) - \Phi(\frac{J_{ij*} - \xi}{\lambda})\right\}, \tag{11}$$

where Φ is the standard normal distribution function. The parameters ξ and λ can be estimated by maximising this log likelihood. In principle any parametric or nonparametric method could be used at this stage to estimate the density of x. Prior knowledge from historical data can also be incorporated, perhaps estimating the posterior distribution by a Markov Chain Monte Carlo algorithm (see, e.g., Carlin and Louis [9]; Shi and Lee [10], and the references therein).

It is shown in Appendix II that the above approach is also approximately valid for

adjusted log odds ratios provided the covariate adjustments are not too big. In practice we estimate β_i by using the values of the adjusted rather than the crude log odds ratios $\boldsymbol{y}_i$ in (10).

3.3 Publication bias and sensitivity analysis

As noted earlier, the funnel plot for the alcohol and breast cancer meta-analysis in Figure 1(i) shows clear signs of publication bias. This graph is a scatter plot of the estimates of $\hat{\beta}_i$ and their standard errors estimated by the ML approach discussed in the last subsection. We follow Copas and Shi [4] and [5] by developing a sensitivity analysis which monitors the effect on the overall slope estimate of different degrees of study selection. The idea is to formulate a family of selection models which explains the trend we have seen in the funnel plot, and then to study the possible values of $\hat{\beta}$ given by the models within that family. The important question in the example is to assess whether some or even all of the dose response is an artifact of publication bias rather than a real effect of alcohol on breast cancer.

Our selection model is based on the idea that there is a population of comparable studies which been done in the area of interest, and that the studies available for the meta-analysis are a selection from this population. We imagine that the studies which have not been selected were either never written up, not accepted for publication, or were published in obscure journals beyond the reach of the literature search. Publication bias arises if the selection is biased, i.e. if some studies in the population are more likely to be selected than others. Let $\mathcal{S}_i$ be the event that the ith study in the population is selected. To allow for the possibility that selection is biased in favour of the studies with more positive outcomes and/or with larger sample sizes, Copas and Shi [4] and [5] supposed that a study reporting estimate $\hat{\beta}_i$ and standard error $s_{\beta i}$ is selected with probability

$$q(\boldsymbol{y}_i|\beta_i) = P(\mathcal{S}_i|\hat{\beta}_i, s_{\beta i}, \beta_i) = \Phi\left\{\frac{u_i + \rho\frac{\hat{\beta}_i - \beta_i}{s_{\beta i}}}{(1-\rho^2)^{1/2}}\right\} \tag{12}$$

where $u_i = a + b/s_{\beta i}$. This probability is conditional on the true value β_i. With $\beta_i \sim N(\beta, \tau^2)$, the overall selection probability for a study with data $(\hat{\beta}_i, s_{\beta i})$ is

$$q(\boldsymbol{y}_i) = P(\mathcal{S}_i|\hat{\beta}_i, s_{\beta i}) = \int q(\boldsymbol{y}_i|\beta_i)p(\beta_i)d\beta_i = \Phi(v_i), \tag{13}$$

where

$$v_i = \frac{u_i + \tilde{\rho}\frac{\hat{\beta}_i - \beta}{(\tau^2 + s_{\beta i}^2)^{1/2}}}{(1-\tilde{\rho}^2)^{1/2}}, \quad \tilde{\rho} = \frac{s_{\beta i}}{(\tau^2 + s_{\beta i}^2)^{1/2}}\rho.$$

See Copas and Shi [4] for a discussion of the form of these equations and the motivation behind them.

The marginal selection probability for a study with standard error $s_{\beta i}$ is

$$P(\mathcal{S}_i | s_{\beta i}) = \Phi(u_i) = \Phi\left(a + \frac{b}{s_{\beta i}}\right). \tag{14}$$

This gives the interpretation of the parameters a and b. Parameter a controls the overall proportion of studies selected. Parameter b, expected to be positive, controls how the chance of selection depends on study size. To interpret the parameter ρ, (12) shows how ρ controls the dependence of selection on study outcome. If $\rho = 0$ there is no publication bias, and the model is equivalent to the model we discussed in the last section. If $\rho > 0$, which we suspect for the meta-analysis of alcohol and breast cancer, the selected studies will be biased in favour of those with more significant results (larger $\hat{\beta}_i$), hence inducing a trend in the funnel plot as we have seen in Figure 1(i).

Assume now that the studies have resulted from the above selection model. The likelihood is now formed from the conditional distribution of the $\boldsymbol{y}_i$s given that the events $\mathcal{S}_i$ have occurred, so that the log likelihood for the unknown parameters (β, τ, ρ) is

$$L(\beta, \tau, \rho | a, b) = \sum_{i=1}^{m} \log p(\boldsymbol{y}_i | \mathcal{S}_i) = \sum_{i=1}^{m} \{\log p(\boldsymbol{y}_i) + \log q(\boldsymbol{y}_i) - \log \Phi(u_i)\}, \tag{15}$$

where $p(\boldsymbol{y}_i)$ is the distribution of $\boldsymbol{y}_i$ studied in Section 3.2, $q(\boldsymbol{y}_i)$ is given by (13), and $\Phi(u_i)$ is given by (14). With grouped exposure levels, the marginal density function $p(\boldsymbol{y}_i)$ is the integral

$$p(\boldsymbol{y}_i | \boldsymbol{J}_i, \boldsymbol{\theta}) = \tau^{-1} \int p(\boldsymbol{y}_i | \boldsymbol{J}_i, \beta_i) \phi\left(\frac{\beta_i - \beta}{\tau}\right) d\beta_i$$

where $p(\boldsymbol{y}_i | \boldsymbol{J}_i, \beta_i)$ is the density function of $N(\boldsymbol{A}_i(\beta_i), \boldsymbol{\Omega}_i)$, and ϕ is the standard normal density function.

Since we have no information about the number of unpublished studies, i.e. the studies in the population which have not been selected, it is intuitively clear that the likelihood can give us no useful information about the selection parameters a and b. We follow the same

procedure as argued in Copas and Shi [4]: we estimate (β, τ, ρ) for a range of different values of (a, b), and then monitor how sensitively the results depend on the particular choice of these selection parameters.

For any given pair (a, b), the unknown parameters can be estimated by maximising (15). The goodness-of-fit of the model can be assessed by calculating fitted values and plotting the corresponding residuals from the funnel plot. Following Copas and Shi [4], the fitted values of β_i are

$$E(\hat{\beta}_i|\mathcal{S}_i) = \beta + \rho s_{\beta i}\frac{\phi(u_i)}{\Phi(u_i)}, \tag{16}$$

giving the approximate residuals

$$r_i = \{\hat{\beta}_i - E(\hat{\beta}_i|\mathcal{S}_i)\}/s_{\beta_i}. \tag{17}$$

A plot of r_i against $s_{\beta i}$ can be used as an informal test of goodness-of-fit. Copas and Shi [4] suggest testing the hypothesis $H_0 : \gamma = 0$ in the extended model

$$\beta_i \sim N(\beta + \gamma s_{\beta i}, \tau^2), \tag{18}$$

fitted under the restriction that $\rho \geq 0$. This is testing whether the trend in the funnel plot is real or whether it is simply an artifact of the selection process. Note that this analysis is conditional on the selection events $\mathcal{S}_i$, and so all these quantities are functions of a and b. Accepting this null hypothesis for any particular pair (a, b) suggests that the model gives a satisfactory fit to the funnel plot.

An exact analysis is difficult because of the integrals in the likelihood (15), but we propose the following approximate method. Assuming that, for each study in the population, $\hat{\beta}_i$ follows (8), we can write the model generating the $\hat{\beta}_i$s as

$$\hat{\beta}_i = \beta_i + s_{\beta i}\epsilon_i, \quad \epsilon_i \sim N(0, 1) \quad \beta_i \sim N(\beta, \tau^2).$$

Now the selection model defined in (12) can equivalently be written in the form of a selection equation

$$z_i = u_i + \delta_i, \quad \delta_i \sim N(0, 1),$$

where z_i is now a latent variable: the ith study is selected if $z_i \geq 0$ and not selected if $z_i < 0$. The parameter ρ is now the correlation between the residuals of these two equations,

$\rho = corr(\epsilon_i, \delta_i)$. This is now very similar to the formulation discussed in Copas and Shi [5], and so the algorithms developed there can now be applied to fit the model.

4 Example: association between breast cancer and alcohol

4.1 Meta-analysis for alcohol use and breast cancer

The association between alcohol consumption and risk of breast cancer was studied in the systematic review reported by Longnecker et *al.* [11]. A reanalysis of the same data, allowing for the correlation between estimated log-odds ratios, was subsequently published by Greenland and Longnecker [7]. These authors included 16 studies in their meta-analysis, of which we use $m = 13$ here. We have omitted two of them (their studies 12 and 13) because of difficulties in extracting the raw figures, and a third study (number 11) because of the poor quality of data. This particular study has only two non-zero dose categories and only 2 cases at the higher level (actually if this study is included it makes no appreciable difference to the analysis because of the very high variance of its $\hat{\beta}_i$). Because of the reduced number of studies our analysis is not directly comparable to the results published earlier. The data in Table 1 is study number 10 in our meta analysis.

With the fixed effects model (Section 3.1 with $\tau = 0$), and using the same assigned values for x_{ij} as used by these authors, our overall estimate of the trend parameter β is 0.0088 with standard error 0.0014, comparing quite closely with the figures of 0.0082 and 0.0013 for the 16 studies of Greenland and Longnecker [7]. Details are given in the columns headed 'GL approach' in Table 2. Replacing fixed effects by the random effects model (8), gives $\hat{\beta} = .0112$ with standard error 0.0026, and $\hat{\tau} = .0065$ (penultimate two rows). The heterogeneity is significant ($P = 0.02$, the last row of the table). The corresponding P-value for $H_0 : \beta = 0$ is less than 0.001, suggesting that the association between breast cancer and the consumption of alcohol is highly significant.

These results are based on the assigned values of alcohol consumption chosen by these previous authors. As discussed earlier, the estimates $\hat{\beta}_i$ can be very sensitive to these choices. Following Section 3.2, suppose that alcohol consumption has a lognormal distribution over

all the subjects in these studies. Taking x to be the log of alcohol consumption, we maximise (11) to give the estimates $\hat{\xi} = 1.533$ and $\hat{\lambda} = 1.380$. We next take $f(x)$ to be this fitted log normal distribution and maximise L_i in (10), followed by calculating the standard errors from (21). We have used Monte Carlo to calculate the integrals in (10), with a sample size of 2000. The results are reported in the columns headed 'ML approach' in Table 2.

For a further comparison of methods, we use this fitted dose distribution to suggest another method of calculating assigned doses, leading to the columns headed 'Mean approach'. Here we have defined x_{ij} by finding the expected log dose calculated from the appropriate truncated section of the lognormal distribution.

Table 2 shows that for many of the individual studies there are considerable differences between the methods of analysis. In a few cases, the estimates and standard errors for both assignment methods are far away from those calculated using the ML approach, especially for studies with a small number of categories. As expected, the mean approach estimates are somewhat closer to ML, but the differences can still be large relative to the underlying sampling variability.

When the studies are pooled together using the fixed effects model, the Mean and GL approaches are suggesting a much stronger dose trend than the ML approach. With the random effects model, the estimates from the assigned values approaches are even higher. The value of $\hat{\tau}$ for the ML approach is much smaller than for the assigned values approaches, in fact the fixed and random effects models are almost indistinguishable when ML is used. It seems that the extra uncertainty through the variability of dose levels within classes has almost entirely explained the heterogeneity. The consequence of having a large value of τ is that more weight is given to the smaller studies. Because of the evident publication bias, the smaller studies tend to be those with larger regression slopes, and hence the pooled estimates by the assigned value methods will be over-estimated.

Table 2. Meta analysis estimates for alcohol and breast cancer studies

	Mean approach*		ML approach*		GL approach*		Number of non-zero
Study number	β	SE	β	SE	β	SE	categories
1	0.0054	0.0028	0.0027	0.0016	0.0043	0.0025	3
2	0.0111	0.0044	0.0070	0.0029	0.0109	0.0041	4
3	0.0149	0.0030	0.0067	0.0010	0.0284	0.0057	4
4	0.0328	0.0164	0.0157	0.0069	0.1180	0.0476	3
5	0.0107	0.0033	0.0087	0.0030	0.0121	0.0043	4
6	0.0256	0.0080	0.0204	0.0076	0.0870	0.0232	2
7	0.0030	0.0033	0.0024	0.0027	0.0031	0.0037	6
8	0.0000	0.0072	0.0000	0.0043	0.0000	0.0094	2
9	0.0026	0.0030	0.0020	0.0024	0.0060	0.0066	4
10	0.0211	0.0089	0.0190	0.0094	0.0479	0.0205	3
11†	0.0081	0.0041	0.0053	0.0035	0.0111	0.0048	4
12†	0.0112	0.0054	0.0080	0.0045	0.0148	0.0064	2
13†	0.0092	0.0067	0.0066	0.0054	-0.0008	0.0087	2
Pooled estimate by fixed effect model							
	0.0084	0.0012	0.0055	0.0007	0.0088	0.0014	
Pooled estimate by random effect model							
	0.0090	0.0016	0.0054	0.0009	0.0112	0.0026	
$\hat{\tau}$	0.0034		0.0013		0.0065		
p-value of $\tau = 0$	0.1164		0.3814		0.0201		

*ML approach is based on grouped data; the others are based on assigned values using the mean of each category or using the values in Greenland and Longnecker [7];

†Since original studies 11, 12 and 13 are excluded, these three studies are numbers 14, 15 and 16 in Greenland and Longnecker [7].

With the GL approach the random effects trend estimate is 0.0112, implying that one extra drink daily (about 13g of alcohol, according to Longnecker [8]) increases risk by about 16%. By the ML approach the risk increase is less than half that figure, about 7% (from an estimated trend of 0.0054) .

4.2 Sensitivity analysis for publication bias

Figure 1(ii) plots the residuals $(\hat{\beta}_i - \hat{\beta})/s_{\beta i}$ against s_{β_i} for the random effects model without publication bias, i.e. the plot of r_i in (17) when $a = b = \infty$ or when $\rho = 0$. The clear upwards trend in this plot is informal confirmation of the presence of publication bias. More formally, a likelihood ratio test of the hypothesis $H_0 : \gamma = 0$ in (18) derived from the likelihood function without selection gives a P-value of $P_{\gamma=0} = 0.05$.

To implement the approximate method discussed in Section 3.3, we use the *Splus* algo-

rithms *scode1* and *scode2* listed in Copas and Shi [5]. Taking the estimates of the individual regression slopes and their standard errors (the figures in the columns headed 'ML approach' in Table 2) as input, we first run algorithm *scode1*. This gives the eight quantities presented in Table 3, for each of six pairs of values of a and b representing increasing severity of selection bias. Copas and Shi [5] describes how these representative values of a and b are chosen. The eight quantities calculated by *scode1* are the columns of the table:

1. $\hat{\beta}$, the pooled estimate β
2. $P_{\beta=0}$, the P-value for testing the null effect hypothesis $H_0 : \beta = 0$
3. $\hat{\beta}_L$, lower 95% confidence limit for β
4. $\hat{\beta}_U$, upper 95% confidence limit for β
5. $P_{\gamma=0}$, the P-value for fit to the funnel plot (testing $\gamma = 0$ in (18))
6. Max $P(\mathcal{S}_i)$, the marginal selection probability for the largest observed study (the one with the smallest standard error)
7. Min $P(\mathcal{S}_i)$, the marginal selection probability for the smallest observed study (the one with the largest standard error)
8. $\mathcal{N}$, the estimated total number of studies given by

$$\mathcal{N} = \sum_i \{P(\mathcal{S}_i)\}^{-1}.$$

This last quantity is an estimate of the size of the hypothetical population of studies from which those in the meta analysis are assumed to have been sampled, and can be used as an aid to interpreting the severity of selection being invoked.

The first row of Table 3 takes $a = b = \infty$ and so is assuming no selection; these figures are just reproducing the analysis in the last subsection. The last entry in the first row is therefore just 13, the number of studies actually in the meta-analysis. The second row admits a small amount of selection, imagining the 13 studies to be selected out of 15 (2 unpublished studies). The estimate of β has decreased. The next row has the 13 studies selected from 22. There is now considerably more publication bias, and the P-value for $H_0 : \beta = 0$ becomes

rather less clear-cut ($P_{\beta=0} = 0.01$). The corresponding fitted values in the funnel plot are the dashed line in Figure 1(i), giving the residuals in Figure 1(iii). These are calculated from (16) and (17). Similarly, the fourth row gives the dotted line in Figure 1(i) and the residual plot in Figure 1(iv). Both third and fourth rows give a reasonable explanation of the trend in the funnel plot, confirmed by the fact that $P_{\gamma=0}$ in Table 3 is well in excess of 5% from the second row downwards.

Table 3 shows the drastic effect of selection bias. The fourth row (13 studies out of 35) no longer gives a significant trend, $P_{\beta=0} = 0.16$. The apparent association between alcohol and breast cancer in the conventional meta-analysis is being explained away as an artifact of the way in which the studies have been selected. Of course we have no evidence that the (a, b) pair underpinning this row in the table is 'correct' — all we can say is that this model fits the data well, and that the possibility that there are $35 - 13 = 22$ further studies which have been done somewhere in the world but have not turned up in the literature search, does not seem entirely unreasonable.

Table 3. Sensitivity analysis for publication bias

(a, b) selection	$\hat{\beta}$	$P_{\beta=0}$	$\hat{\beta}_L$	$\hat{\beta}_U$	$P_{\gamma=0}$	Max $P(\mathcal{S}_i)$	Min $P(\mathcal{S}_i)$	$\mathcal{N}$
1	0.0054	0.0000	0.0029	0.0080	0.0591	1.00	1.00	13
2	0.0046	0.0002	0.0022	0.0070	0.3184	0.95	0.80	15
3	0.0036	0.0108	0.0008	0.0063	0.5548	0.80	0.50	22
4	0.0029	0.1592	-0.0012	0.0071	0.8219	0.60	0.30	35
5	0.0029	0.1894	-0.0014	0.0071	0.9723	0.40	0.10	91
6	0.0029	0.1872	-0.0014	0.0071	0.6545	0.20	0.01	670

The rows of Table 3 select six 'representative' (a, b) pairs. A much wider selection of possible values of a and b (essentially all reasonable possibilities) is taken by the algorithm *scode2* in Copas and Shi [5]. This results in Figure 2, which plots values of $\hat{\beta}$ against the corresponding P-values for the fit to the funnel plot (value of $P_{\gamma=0}$). The general downwards drift of the points in this plot suggests that, on the whole, the better a selection model fits the data, the smaller is the estimate of β. The points in the graph are divided into those corresponding to $\mathcal{N} \leq 26$ (the dots), those with $26 < \mathcal{N} \leq 39$ (denoted +), and those with $\mathcal{N} > 39$ (the circles). Respectively, these represent marginal selection probabilities of more

than 1/2, between 1/3 and a 1/2, and less than 1/3. The average value of $\hat{\beta}$ for the dots in Figure 2, where the number of unpublished studies is less than the actual number of studies in the meta analysis, is around 0.0038, a reduction of a third from the value with no selection. A trend estimate of 0.0038 means that the risk increase attributed to one extra drink daily is about 5%, compared to the earlier figures of 16% (GL approach) and 7% (ML approach). Allowing for a modest amount of selection, the evidence for the association becomes much more marginal. The appearance of lines in Figure 2 is purely an artifact of the numerical method used by the algorithm, which takes (a, b) pairs over a uniform rectangular grid . Taking a finer grid (by changing one of the arguments of *scode2*) gives a denser cloud of points but the same overall shape.

5 Discussion

In this paper we have discussed three major problems for meta-analysis in trend estimation. The example shows that the estimates and standard errors given by the assigned value methods can be markedly different from those given by a model which allows for the variation in exposure level within each class interval. This can have a major effect on the pooled estimate of the underlying trend. The method we have used (Section 3.2) can only be regarded as a rather crude attempt to overcome deficiencies in the data, and more refined methods may be needed, for example to allow subsets of the studies to have different underlying exposure distributions. These difficulties suggest that, whenever possible, data for finer exposure intervals should be routinely reported in such studies.

The second difficulty is heterogeneity. We have followed the usual method of adding a normally distributed random effect, but this itself can be problematical. If publication bias exists then a random effects model makes things worse, as more weight will be given to the smaller studies. At a more fundamental level, a random effects model makes the objective of meta-analysis less clear — are we interested in a global 'average' effect, or the effect prevailing under particular circumstances? Under a random effects model these are not the same thing. What our example does suggest is that if we model the within-study variability correctly then the random effect will have a smaller variance or even be eliminated altogether.

The third difficulty, and the most troublesome, is publication bias. Study selection can have a drastic effect on the outcome of a meta-analysis, as seen in the second column of Table 3. Several methods for correcting for publication bias have been proposed in the literature [3]. These methods inevitably rest on assumptions which cannot be checked from the available data. Following Copas and Shi [4] and [5], we have argued for a sensitivity approach. The essential indeterminacy of the problem means that we cannot give a single 'best' estimate, all we can do is study a range of possibilities which conform to the evidence in the funnel plot. We have argued in the example that the range of plausible trend estimates does *not* include the conventional fixed effects or random effects results — the true figure is likely to be smaller and possibly substantially smaller. If we entertain the possibility that the systematic review has only found about 50% of all relevant studies, then the evidence that alcohol is a risk factor for breast cancer becomes unclear.

Our more cautious analysis of the studies in Longnecker *et al.* [11] is consistent with the later more extensive meta-analysis of 39 studies in Longnecker [8]. Using the mean approach for assigning exposure levels, their overall random effects estimate is 0.0073 (Longnecker [8], p77) compared with the figure of 0.0090 in our Table 2. Longnecker's comment ([8], p 73) that "the modest size of the association and variation in results across studies leaves the causal role of alcohol in question" seems to be confirmed by our sensitivity analysis. Unfortunately we have not been able to extract the raw data needed to reanalyse all 39 studies in Longnecker [8].

References

1. Thompson, S. G. and Sharp, S. J. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 1999; **18**: 2693-2708.

2. Copas, J. B. What works?: selectivity models and meta-analysis. *J. R. Statist. Soc. A*, 1999; **162**: 95-100.

3. Song, F., Eastwood, A. J., Gilbody, S., Duley, L. and Sutton, A. J. Publication and other selection biases in systematic reviews. *Health Technology Assessment*, 2000; **4**: 1-115.

4. Copas, J.B. and Shi, J.Q. Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, 2000; **1**: 247-262.

5. Copas, J.B. and Shi, J.Q. A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research*, 2001; **10**: 251-165.

6. Rohan, T. E. and Mcmichael, A. J. Alcohol consumption and risk of breast cancer. *Int. J. Cancer*, 1988; **41**: 695-699.

7. Greenland, S. and Longnecker, M.P. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am. J. Epidemiology*, 1992; **135**: 1301-1309.

8. Longnecker, M. P. Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review. *Cancer Causes and Control*, 1994; **5**: 73-82.

9. Carlin, B.P. and Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis.* (2nd edition). London: Chapman & Hall/CRC, 2000.

10. Shi, J. Q. and Lee, S. Y. Latent variable models with mixed continuous and polytomous data. *J. R. Statist. Soc. B*, 2000; **62**: 77-87.

11. Longnecker, M. P, Berlin, J. A., Orza, M. J. and Chalmers, T. C. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA*, 1988; **260**: 652-656.

Appendix I: Likelihood theory for Section 3.2

Following the dose analysis model (1), we assume that

$$\log \frac{p_x}{1-p_x} = \log \frac{p_0}{1-p_0} + \beta x,$$

where p_0 is the probability of being a case for the unexposed group. For the ith study, $\log(p_0/(1-p_0))$ is estimated by $\hat{\alpha}_i = \log(C_{i0}/D_{i0})$. Thus from (9), the probability that an individual in class interval J_{ij} in the ith study is a case is

$$p_{ij}(\beta_i) = \frac{1}{q_{ij}} \int_{J_{ij}} \frac{\exp(\hat{\alpha}_i + \beta_i x)}{1 + \exp(\hat{\alpha}_i + \beta_i x)} f(x) dx \tag{19}$$

where $q_{ij} = \int_{J_{ij}} f(x)dx$.

Define

$$A_{ij}(\beta_i) = \log \frac{p_{ij}(\beta_i)}{1 - p_{ij}(\beta_i)} - \hat{\alpha}_i. \tag{20}$$

Then, following the approximations in Section 3.1, the vector of log odds ratios $\boldsymbol{y}_i = (y_{ij})$ is approximately normally distributed with mean $\boldsymbol{A}_i(\beta_i) = \{A_{ij}(\beta_i)\}$ and covariance matrix $\boldsymbol{\Omega}_i$. This gives (10) as the likelihood for β_i. The MLE $\hat{\beta}_i$ is then given by the non-linear least squares equation

$$\frac{\partial L_i}{\partial \beta_i} = \{\boldsymbol{y}_i - \boldsymbol{A}_i(\beta_i)\}' \boldsymbol{\Omega}_i^{-1} (\frac{\partial \boldsymbol{A}_i}{\partial \beta_i}) = 0.$$

The variance of $\hat{\beta}_i$ is equal to the inverse of the negative second derivative of L_i, namely

$$\frac{\partial^2 L_i}{\partial \beta_i^2} = -(\frac{\partial \boldsymbol{A}_i}{\partial \beta_i})' \boldsymbol{\Omega}_i^{-1} (\frac{\partial \boldsymbol{A}_i}{\partial \beta_i}) + \{\boldsymbol{y}_i - \boldsymbol{A}_i(\beta_i)\}' \boldsymbol{\Omega}_i^{-1} (\frac{\partial^2 \boldsymbol{A}_i}{\partial \beta_i^2}). \tag{21}$$

The derivatives appearing in these expressions are vectors formed in the obvious way from

$$\frac{\partial A}{\partial \beta} = \frac{\dot{p}}{p(1-p)}, \quad \frac{\partial^2 A}{\partial \beta^2} = \frac{\ddot{p}}{p(1-p)} - \frac{\dot{p}^2(1-2p)}{p^2(1-p)^2},$$

$$\dot{p} = \frac{\partial p}{\partial \beta} = \frac{1}{q} \int_J \frac{e^u}{(1+e^u)^2} x f(x) dx \quad \text{and} \quad \ddot{p} = \frac{\partial \dot{p}}{\partial \beta} = \frac{1}{q} \int_J \frac{e^u(1-e^u)}{(1+e^u)^3} x^2 f(x) dx,$$

with $u = \hat{\alpha} + \beta x$.

Appendix II: Use of adjusted log odds ratios in Section 3.2

In applying the formulae in Section 3.2 we assumed that we can use covariate adjusted log-odds ratios in place of the actual log-odds ratios y_{ij}. To see why this is at least approximately

valid, suppose there are just two exposure groups, an unexposed group $d = 0$ and an exposed group $d = 1$. The cases are labelled $c = 1$ and the controls $c = 0$. The exposed group have independent dose levels x, sampled from a known distribution, and all subjects have covariate vectors z which we wish to use to adjust the estimate of the log-odds ratio. Let the mean of z in each group be μ_1 for the exposed group and μ_0 for the unexposed group. We assume that in the exposed group x and z are independent.

If we knew all the values of x and z, we could fit the logistic models

$$\log \frac{P(c = 1|d = 0, z)}{P(c = 0|d = 0, z)} = \alpha + \delta' z, \tag{22}$$

and

$$\log \frac{P(c = 1|d = 1, x, z)}{P(c = 0|d = 1, x, z)} = \alpha + \beta x + \delta' z. \tag{23}$$

Then if the covariate regression effect δ is sufficiently small, we can use the lemma below to find

$$P(c = 1|d = 0) = E_z\{P(c = 1|d = 0, z)\} \simeq p_\alpha + p_\alpha(1 - p_\alpha)\delta'\mu_0, \tag{24}$$

where

$$p_\alpha = \frac{e^\alpha}{1 + e^\alpha}, \tag{25}$$

and

$$P(c = 1|d = 1) = E_{x,z}\{P(c = 1|d = 1, x, z)\} \simeq p_{\alpha\beta} + p_{\alpha\beta}(1 - p_{\alpha\beta})\delta'\mu_1, \tag{26}$$

where

$$p_{\alpha\beta} = E_x \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}.$$

Thus the log-odds ratio which compares the two marginal probabilities (26) and (24) is

$$A_{crude} = A_{adj} + \delta'(\mu_1 - \mu_0), \tag{27}$$

where

$$A_{adj} = \log \frac{p_{\alpha\beta}(1 - p_\alpha)}{p_\alpha(1 - p_{\alpha\beta})}.$$

If, however, the varying dose x is ignored but we still observed z, the model would be

$$\log \frac{P(c = 1|d, z)}{P(c = 0|d, z)} = \alpha + \gamma d + \delta' z.$$

Here, γ is the covariance adjusted log-odds ratio under this model. If δ were known, the equations for the MLEs of α and γ would be

$$\sum \left(c - \frac{e^{\alpha+\gamma d+\delta' z}}{1+e^{\alpha+\gamma d+\delta' z}} \right) \begin{pmatrix} 1 \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Again using the lemma, the solution to these equations reduce to

$$\hat{\alpha} \simeq \log \frac{C_0}{D_0} - \delta' \mu_0,$$

and

$$\hat{\alpha} + \hat{\gamma} \simeq \log \frac{C_1}{D_1} - \delta' \mu_1,$$

where C_0 and D_0 are the numbers of cases and controls for the unexposed group, and C_1 and D_1 are the same quantities for the exposed group. Thus

$$\hat{\gamma} \simeq y - \delta'(\mu_1 - \mu_0),$$

where

$$y = \log \frac{C_1 D_0}{C_0 D_1}$$

so that

$$\hat{\gamma} - A_{adj} \simeq y - A_{crude}. \tag{28}$$

The left hand side of (28) is the difference between the adjusted log-odds ratio which is generally reported in studies of this kind, and its expected value under the full model (22) and (23). The right hand side is the difference between the observed and expected values used in the non-linear least squares analysis of Section 3.2, namely the difference between the crude log-odds ratio calculated from the observed marginal frequences C_d and D_d, and its expected value under the full model. For small values of δ these are approximately equal. Further use of the lemma shows that, to the same order of approximation, A_{adj} is approximately the same whether α is estimated by the adjusted estimate $\hat{\alpha}$ or by the unadjusted estimate $\log(C_0/D_0) = \hat{\alpha} + \delta' \mu_0$.

A more careful analysis with more than two groups follows similar lines. The approximations depend on the linearisation of the logit function implicit in the lemma, and so are good approximations when δ and β are small.

Lemma: If ϵ is small then

$$\frac{e^{\alpha+\epsilon}}{1+e^{\alpha+\epsilon}} = p_\alpha + p_\alpha(1-p_\alpha)\epsilon + O(\epsilon^2).$$

where p_α is defined in (25).

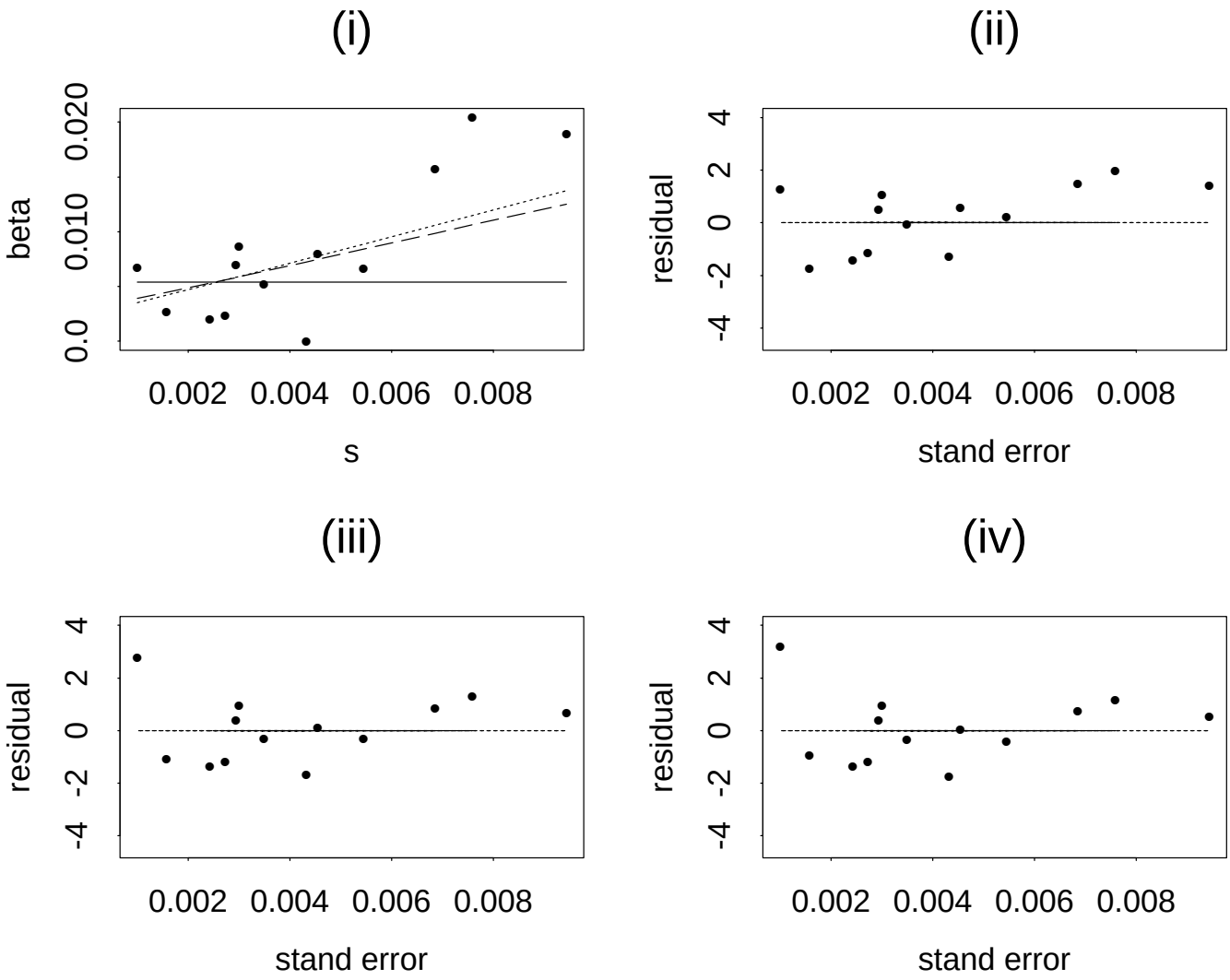

Figure 1: Alcohol use and breast cancer data. (i) Funnel plot of $\hat{\beta}_i$ against $s_{\beta i}$. The solid line represents the conventional estimate $\hat{\beta} = .0054$, the dashed and dotted lines represent the fitted values for the 3rd and 4th rows of Table 3 ($\hat{\beta} = .0036$ and 0.0029). (ii)–(iv) Residual plots corresponding to the solid, dashed and dotted lines respectively in (i).

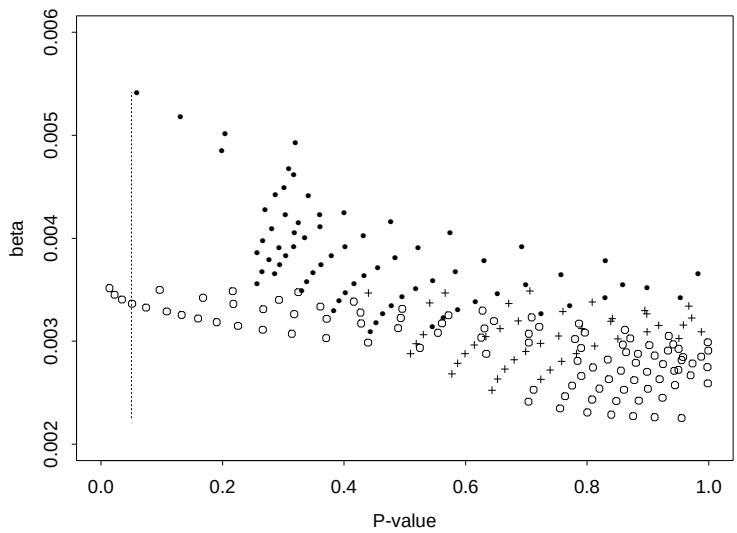

Figure 2: Sensitivity analysis plot: $\hat{\beta}$ against $P_{\gamma=0}$ for selection models with $\mathcal{N} \leq 26$ $(\cdot)$, $26 < \mathcal{N} \leq 39$ $(+)$ and $\mathcal{N} > 39$ $(\circ)$.