STATISTICS IN MEDICINE, VOL. 15, 1713–1728 (1996)

THE POPULATION RISK AS AN EXPLANATORY VARIABLE IN RESEARCH SYNTHESIS OF CLINICAL TRIALS

MARTIN W. McINTOSH

Department of Statistics, Harvard University, 1 Oxford Street, Cambridge MA 02138, U.S.A.

SUMMARY

The population risk, for example the control group mortality rate, is an aggregate measurement of many important attributes of a clinical trial, such as the general health of the patients treated and the experience of the staff performing the trial. Plotting measurements of the population risk against the treatment effect estimates for a group of clinical trials may reveal an apparent association, suggesting that differences in the population risk might explain heterogeneity in the results of clinical trials. In this paper we consider using estimates of population risk to explain treatment effect heterogeneity, and show that using these estimates as fixed covariates will result in bias. This bias depends on the treatment effect and population risk definitions chosen, and the magnitude of measurement errors. To account for the effect of measurement error, we represent clinical trials in a bivariate two-level hierarchical model, and show how to estimate the parameters of the model by both maximum likelihood and Bayes procedures. We use two examples to demonstrate the method.

1. INTRODUCTION

Table I records the results of nine clinical trials that evaluated magnesium for treating acute myocardial infarction (AMI). Table II records the results of 14 clinical trials that evaluated β-mimetics for treatment of preterm labour. Both sets of data tabulate independent evaluations of a common treatment whose effects we may wish to summarize in a research synthesis, possibly by meta-analysis. One method that allows the representation of heterogeneity, and recommended in a National Research Council report,[1] summarizes the results in a random effects model, and may include attributes of the trials as covariates (see DerSimonian and Laird,[2] DuMouchel and Harris,[3] Hedges and Olkin,[4] Morris and Normand,[5] and more recently Berkey *et al.*[6]). For example, one may use the dose level of the drug or indicators of controlled factors in each study (for example did the study control for sex?) to explain some of the heterogeneity in treatment effect estimates. Figure 1 plots the treatment effect estimates versus the control group event rates for the data in Tables I and II. We estimate the regression lines by representing the treatment effects in a random effects model with the observed control group event rates as covariates. Both regressions reveal a negative association, suggesting that differences in the control group event rates may explain some heterogeneity of the treatment effect, and that populations with higher control group rates benefit more from the treatment than populations with low control group rates. Such association between the treatment effect and the control group event rates has been shown to occur commonly (see Schmid *et al.*[7]).

CCC 0277–6715/96/161713–16

Received March 1995
Revised October 1995

Table I. Data from magnesium trials. Treatment and control group outcomes $\hat{\alpha}_t$ and $\hat{\alpha}_c$ are mortality rates. treatment effect estimate defined as the log odds ratio is given by Y_t, and its variance by var(Y_t). The variance of the control group risk is given by var(Y_r), and $\beta_y = \sigma_{tr}/\sigma_r^2$ is the regression slope for estimating Y_t from Y_r

Source	Magnesium $\hat{\alpha}_t$	Control $\hat{\alpha}_c = Y_r$	Log (OR) Y_t	Var(Y_t) σ_t^2	Var(Y_r) σ_r^2	Slope β_y
Morton	1/40	2/36	− 0·8303	1·5550	0·00146	− 19·057
Abraham	1/48	1/46	− 0·0434	2·0434	0·00046	− 47·020
Feldsted	10/150	8/148	0·2231	0·2392	0·00035	− 19·557
Rasmussen	9/135	23/135	− 1·0561	0·1714	0·00105	− 7·074
Ceremuzynski	1/25	3/23	− 1·2809	1·4250	0·00493	− 8·816
Schechter	1/59	9/56	− 2·4075	1·1496	0·00241	− 7·413
LIMIT2	90/1150	118/1150	− 0·2976	0·0214	0·00008	− 10·859
ISIS 4	1997/27413	1897/27411	0·0552	0·0011	2·35E − 6	− 15·524
Schechter (1995)	4/92	17/98	− 1·5298	0·3325	0·00146	− 6·974

Table II. Data from β-mimetics trials. Data taken from Keirse, Grant and King with Mariona trial removed. Treatment and placebo group outcomes $\hat{\alpha}_t$ and $\hat{\alpha}_c$ are preterm labour rates, and the treatment effect estimate defined as the log odds ratio is given by Y_t, and its variance is given by var(Y_t). The variance of the control group risk is given by var(Y_r), and the regression slope for estimating Y_t from Y_r is $\beta_y = \sigma_{tr}/\sigma_r^2$. The variance of the treatment effect were derived by considering the 95 per cent confidence intervals given in Brand and Kragt as having width $2 \times 1{\cdot}96 \times \sigma$, where σ is the standard error of the odds-ratio. The variability of log(OR) was estimated by a Taylor series method as $\sigma_t^2 = \sigma^2/\mathrm{OR}^2$

Source	β-mimetics $\hat{\alpha}_t$	Placebo $\hat{\alpha}_c = Y_r$	Log (OR) Y_t	Var(Y_t) σ_t^2	Var(Y_r) σ_r^2	Slope β_y
Christiansen *et al.* (1980)	0/14	6/16	− 2·303	2·189	0·015	− 4·266
Spellacy *et al.* (1979)	6/14	11/15	− 1·204	1·041	0·013	− 5·113
Barden (unpublished)	0/12	8/13	− 2·659	1·446	0·018	− 4·225
Hobel (unpublished)	2/16	3/15	− 0·545	2·735	0·011	− 6·25
Cotton *et al.* (1984)	6/19	11/19	− 1·050	0·703	0·013	− 4·102
Howard *et al.* (1982)	2/15	2/18	0·199	3·971	0·005	− 10·125
Ingemarsson (1976)	0/15	10/15	− 2·813	0·956	0·015	− 4·500
Larsen *et al.* (1986)	5/19	16/50	− 1·309	0·332	0·004	− 4·595
Calder and Patel (1985)	4/37	9/39	− 0·844	0·568	0·005	− 5·633
Scommegna (unpublished)	1/15	5/17	− 1·427	1·969	0·012	− 4·816
Wesselius and DeCasparis (1971)	6/33	15/30	− 1·427	0·407	0·008	− 4·000
Laveno *et al.* (1986)	15/54	25/52	− 0·868	0·202	0·005	− 4·006
Larsen *et al.* (1980)	11/131	6/45	− 0·562	0·519	0·003	− 8·853

The population risk is a value that summarizes the aggregate health of a population treated or as an indicator of other 'clinical notions like "experience" or "seriousness of condition"' (Brand and Kragt[8]). Differences in population risk may be important in explaining heterogeneity in the effect of a treatment. Although we would ideally incorporate risk factors directly by, for example, controlling for age, sex or medical history of the patients, in many studies this information may not have been collected or published, and thus is unavailable for a meta-analysis. The population risk that measures these traits in aggregate may be useful in such circumstances.

The population treatment effect (TE) and population risk (PR) for a trial are those values that we would observe if one conducted the trial with an infinite sample size. In practice we do not

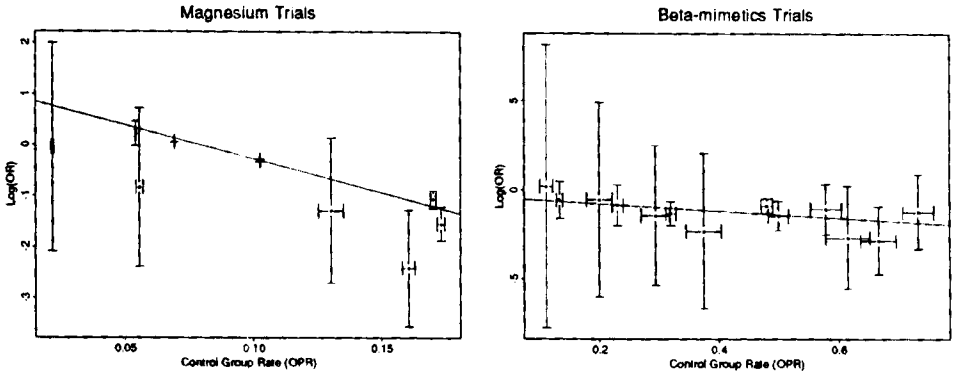

Figure 1. Relationship between the observed treatment effect (defined as log-odds ratio: log(OR)) and the observed control group risk (OPR). Left graph shows the relationship for the nine magnesium trials. The population risk is defined as the mortality rate in the control group. Right graph shows the relationship for thirteen of the β-mimetics trials (the Marion trial was dropped), and the population risk is defined as the preterm labour rates in the control group. The dotted lines are the estimated regression lines when using the observed control group risk as a covariate in a random effects regression. The horizontal error bars give a two standard-error interval for the true control group risk in each trial. The vertical error bars give a two standard-error interval for the true treatment effect in each trial

observe the TE and PR directly but we instead estimate them from a finite sample by the observed treatment effect (OTE) and the observed population risk (OPR). As we demonstrate below, the relationship between the OTE and OPR does not properly represent the relationship between the TE and PR, and a regression estimate of the OTE on the OPR is biased. Because the values plotted in Figure 1 are the OTE and OPR, the apparent negative association shown does not necessarily lead to the conclusion that populations with high PR benefit more from the treatment. That association, or some portion of it, may simply arise as a consequence of measurement error.

The attenuation of measurement errors that result in biased regression parameter estimates is a well known consequence of error-in-variables models, and extensive literature suggests corrections for such effects. (See Fuller,[9] Rosner and Willett,[10] and the body of work by Carrol and Stefanski. See Carrol and Stefanski[11] for application of measurement error models to meta-analysis in particular.) However, the usual assumption of independence between the error in the covariate estimate (here, OPR) and the estimate in the outcome (here, OTE) is violated. Senn[12] showed that correlation in measurement error could explain the association found by Brand and Kragt[8] when analysing the data in Table II, and gave specific results for treatment effects defined as odds ratios. In this paper we address the problem of estimating the association of the TE and the PR from the OTE and OPR, allowing for arbitrary definition of TE and PR, and correcting for the correlated measurement error peculiar to this application.

In Section 2 we construct a hierarchical model that represents both the observables and the structural model we seek to estimate. Section 2.3 demonstrates that, for general definitions of TE and PR, simple inclusion of the OPR as a covariate results in a biased estimate of the relationship we seek to estimate, discusses the conditions leading to severe biases, and indicates when we may ignore biases. Section 3 discusses both Bayes and maximum likelihood estimation procedures and Section 4 demonstrates the methods on the data sets shown in Tables I and II.

2. GENERAL FRAMEWORK

A clinical trial estimates a treatment effect, but this estimate is actually a function of two estimates: the treatment and control group outcomes. For example, the odds ratio is a treatment

effect estimate formed by dividing the observed odds of success (or failure) in the treatment group by the odds of success (or failure) in the control group. This section models the process of TE and PR definition and estimation, and then shows that estimating the association between the TE and PR from a linear regression of the OTE and OPR is biased.

2.1. Population Parameters and the Structural Model

Let θ_{ti} represent the true treatment effect (TE) for trial i, and let θ_{ri} represent true risk (PR) of the treated population in trial i. Both of these values are functions of treatment and control group parameters, α_{ti} and α_{ci}

$$\theta_{ti} \equiv f(\alpha_{ti}, \alpha_{ci})$$

$$\theta_{ri} \equiv f(\alpha_{ti}, \alpha_{ci}).$$

For example, the clinical trials in Table I estimate the true mortality rates α_{ti} and α_{ci}, has the TE defined as the log-odds ratio, $\theta_{ti} \equiv f(\alpha_{ti}, \alpha_{ci}) \equiv \log(\alpha_{ti}/1 - \alpha_{ti}) - \log(\alpha_{ci}/1 - \alpha_{ci})$, and has the PR defined as the mortality rate of the control group, $\theta_{ri} = h(\alpha_{ti}, \alpha_{ci}) \equiv \alpha_{ci}$. Alternatively, we could have chosen the TE as the risk difference, $f(\alpha_{ti}, \alpha_{ci}) \equiv \alpha_{ti} - \alpha_{ci}$, the risk ratio, $f(\alpha_{ti}, \alpha_{ci}) \equiv \alpha_{ti}/\alpha_{ci}$, or the PR as the total risk $h(\alpha_{ti}, \alpha_{ci}) \equiv \alpha_{ti} + \alpha_{ci}$. We assume that the functions $f(\cdot)$ and $h(\cdot)$ have been chosen so that a linear relationship holds between the TE and PR:

Structural model

$$\theta_{ti} \mid \theta_{ri} \sim \mathrm{N}(\mu_t + \beta_\theta(\theta_{ri} - \mu_r), \tau_t^2) \tag{1}$$

$$\theta_{ri} \sim \mathrm{N}(\mu_r, \tau_r^2). \tag{2}$$

We refer to distributions (1) and (2) as the 'structural model' and β_θ as the structural regression coefficient, or structural slope. For the TE and PR definition used in Figure 1, a structural model with $\beta_\theta < 0$ represents the case where a treatment's effectiveness increases with the risk of the control group, and $\beta_\theta = 0$ represents the case where the effect of a treatment is not associated with the risk of the control group. The parameter μ_t is the mean effect of all clinical trials and is the usual estimand in meta-analyses without use of covariates. We chose the 'deviation from means' representation in (1) to make this parameter an estimand. The residual variance τ_t^2 provides a measure of the heterogeneity in the treatment effect with populations matched on control group risk, and τ_r^2 provides a measure of population heterogeneity that is independent of the treatment.

2.2. Observed Quantities

If we could observe α_{ti} and α_{ci} from each clinical trial our analysis is quite straightforward because we simply compute θ_{ti} and θ_{ri} and estimate the structural model from a simple linear regression. Instead, in practice we only have estimates of α_{ti} and α_{ci} from a finite sample with the quantities $\hat{\alpha}_{ti}$ and $\hat{\alpha}_{ci}$. From these observed quantities we calculate the OTE and OPR for each trial as follows:

$$Y_{ti} \equiv f(\hat{\alpha}_{ti}, \hat{\alpha}_{ci})$$

$$Y_{ri} \equiv h(\hat{\alpha}_{ti}, \hat{\alpha}_{ci}).$$

Note that although the measurements $\hat{\alpha}_{ti}$ and $\hat{\alpha}_{ci}$ are independent, Y_{ti} and Y_{ri} may be dependent. Denote the covariance between Y_{ti} and Y_{ri} as σ_{tri} (see Appendix for approximating σ_{tri}). For convenience, we assume that, conditional on θ_{ti} and θ_{ri}, these values are normally

distributed with known variance and covariance according to the following bivariate *measurement error model*:

$$\begin{pmatrix} Y_{ti} \\ Y_{ri} \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} \theta_{ti} \\ \theta_{ri} \end{pmatrix}, \begin{pmatrix} \sigma_{ti}^2 & \sigma_{tri} \\ \sigma_{tri} & \sigma_{ri}^2 \end{pmatrix}\right). \tag{3}$$

We may also write the measurement error model as $Y_{ti} = \theta_{ti} + \beta_{yi}(Y_{ri} - \theta_{ri}) + e_i$ and $Y_{ri} = \theta_{ri} + d_i$, where e_i and d_i are independent normal deviates and $\beta_{yi} = \sigma_{tri}/\sigma_{ri}^2$. We refer to β_{yi} as the within-trial regression coefficient, or within-trial regression slope.

For notational convenience, let $\theta_i = (\theta_{ti}, \theta_{ri})'$ represent the population TE and PR for the ith trial, and $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$. Let $Y_i = (Y_{ti}, Y_{ri})'$ be the vector of estimates of θ_i assumed to follow (3). Also, let $\gamma = (\mu_t, \mu_r, \beta_\theta, \tau_t, \tau_r)'$ represent the parameters of the structural model.

We can conveniently represent the structural model (1) and (2) along with the measurement error model (3) as a two-stage bivariate hierarchical model, with first stage representing the measurement error and the second stage representing the structural model. Table III summarizes this hierarchical model in the form of a 'Morris diagram' (Morris[13]). The left column of Table III contains the measurement error model and structural model. The upper right corner describes the distribution of the OTE and OPR we observe when the TE and PR are not observed. The distribution in the lower left corner is important for the estimation procedures and represents the distribution of the TE and PR when the OTE and OPR are observed.

2.3. Bias

For ease of demonstration we assume equal variances, where all Σ_i are equal to Σ, and thus share a common within trial regression coefficient β_y. From the *between trial* distribution (refer to Table III) we can simply write down what a regression estimates:

$$E(Y_{ti} \mid Y_{ri}) = \mu_t + (\beta_y B + \beta_\theta(1 - B))(Y_{ri} - \mu_r)$$

where $B = \sigma_r^2/\tau_r^2 + \sigma_r^2$, the 'shrinkage factor' in the usual Bayes or empirical Bayes methods (see Morris[13]). That is, an ordinary least squares regression estimate of the OTE on the OPR, $\hat{\beta}_{\mathrm{ols}}$, estimates the quantity $E\hat{\beta}_{\mathrm{ols}} = \beta_y B + \beta_\theta(1 - B)$. Because $0 \leqslant B \leqslant 1$, that regression actually estimates a value somewhere between the slope of the structural model (β_θ) and the regression slope of the measurement error model (β_y), and thus $\hat{\beta}_{\mathrm{ols}}$ may either underestimate or overestimate the magnitude of the structural slope, depending on the relative values of β_y and β_θ. The magnitude of this bias depends on the relative size of τ_r^2 and σ_r^2. The expected bias in $\hat{\beta}_{\mathrm{ols}}$ is $B(\beta_\theta - \beta_y)$, which is $100 \times B$ per cent of the difference between β_θ and β_y. To have small bias we must have B small, and thus we must have σ_r^2 small relative to τ_r^2 (that is large sample sizes for a well measured covariate). A bias of $100 \times B$ per cent or less requires $\sigma_r^2/\tau_r^2 < B/1 - B$.

The data in Tables I and II do not have equal within-trial variances and so we cannot directly quantify the extent of the bias with a single quantity B. However, we can estimate a value of B for each trial by $B_i = \sigma_{ri}^2/(\tau_r^2 + \sigma_{ri}^2)$. If all the B_i are small then we can expect little bias. Because the B_i are shrinkage factors, we may estimate them independent of the structural model by applying Morris's,[13] empirical Bayes procedure* on the control group risks Y_r. Doing so for the two data sets here gives $\min \hat{B}_i = 0{\cdot}009$ to $\max \hat{B}_i = 0{\cdot}523$, with average $\bar{B}$ equal to 0·226, for the

* An Splus function 'norm.hm' which implements Morris[13] is available free, via anonymous ftp to hustat.harvard.edu in directory 'pub/HM'

Table III. Hierarchical diagram representation of the errors-in-variables model. The lower left corner describes the structural model between the population TE and PR that we seek to estimate. The upper left corner describes the error in measurement (note that Σ_i is not diagonal). The upper right hand corner describes the relationship between the OTE and the OPR we observe in our meta-analysis. The lower right hand corner gives the distribution of the unknown population TE and PR, via Bayes rule, conditional on the OTE and OPR

Within-trial (measurement error)	Between-trial (observed relationship)
$\mathbf{Y}_i \mid \theta_i \sim \mathrm{N}_2(\theta_i, \Sigma_i)$	$\mathbf{Y}_i \sim \mathrm{N}_2(\mu, \Sigma_i + \Lambda_0)$
$\theta_i = (\theta_{ti}, \theta_{ri})'$	$\mu = (\mu_t, \mu_r)'$
$\Sigma_i = \begin{pmatrix} \sigma_{ti}^2 & \sigma_{tri} \\ \sigma_{tri} & \sigma_{ri}^2 \end{pmatrix}$	$\Sigma_i + \Lambda_0 = \begin{pmatrix} \sigma_{ti}^2 + \tau_t^2 + \beta_\theta^2 \tau_r^2 & \sigma_{tri} + \beta_\theta \tau_r^2 \\ \sigma_{tri} + \beta_\theta \tau_r^2 & \sigma_{ri}^2 + \tau_r^2 \end{pmatrix}$
Random effect (structural model)	**Posterior**
$\theta_i \sim \mathrm{N}_2(\mu, \Lambda_0)$	$\theta_i \mid \mathbf{Y}_i = y_i \sim \mathrm{N}_2(\theta_i^*, \Lambda_i^*)$
$\mu = (\mu_t, \mu_r)'$	$\theta_i^* = \Lambda_i^* (\Lambda_0^{-1} \mu + \Sigma_i^{-1} y_i)$
$\Lambda_0 = \begin{pmatrix} \tau_t^2 + \beta_\theta^2 \tau_r^2 & \beta_\theta \tau_r^2 \\ \beta_\theta \tau_r^2 & \tau_r^2 \end{pmatrix}$	$\Lambda_i^* = (\Lambda_0^{-1} + \Sigma_i^{-1})^{-1}$

magnesium data, and min $\hat{B}_i = 0{\cdot}0592$ to max $\hat{B}_i = 0{\cdot}2931$, with average $\bar{B}_i$ equal to $0{\cdot}179$ for the β-mimetics data. The average values of the $\hat{B}_i$ suggest potential for substantial bias.

3. ESTIMATION

For both the maximum likelihood and Bayes procedures our estimand is the structural model parameter $\gamma = (\mu_t, \mu_r, \beta_\theta, \tau_t, \tau_r)'$ which we must estimate from k independent observations from the *between-trial* distribution, $Y = (Y_1, Y_2, \ldots, Y_k)'$. These data give us the following *observed likelihood*:

$$L(\gamma \mid \mathbf{Y} = \mathbf{y}) \propto \prod_{i=1}^{k} |\Sigma_i + \Lambda_0|^{-1/2} \exp\{ -\tfrac{1}{2}(y_i - \mu)'(\Sigma_i + \Lambda_0)^{-1}(y_i - \mu)\}. \tag{4}$$

For maximum likelihood estimation, we must maximize (4) and for Bayes estimation (4) multplied by some prior $\pi(\gamma)$ is proportional to the posterior distribution. Fuller[9] (pp. 105, 217–221) gives an iterative procedure that one can use to find the maximum likelihood estimate of γ, but here we adopt a missing data approach and estimate γ from the joint distribution of $(\mathbf{Y}, \theta)$, and treat θ as missing data. We show that the joint distribution is quite simple, and one can apply tools for handling missing data in a straightforward manner to exploit its simplicity. Specifically we make use of the EM algorithm (Dempster *et al.*[14]) for maximum likelihood estimates and the data augmentation algorithm (Tanner and Wong[15]) for Bayes estimates.

Together we refer to Y_i and θ_i as a complete observation and denoted as $Z_i = (Y_i', \theta_i')'$. The distribution of a complete observation is the product of the measurement error distribution and the structural distribution, summarized in the left hand column of Table III, and k independent

samples from this joint distribution, denoted $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_k)'$, give the *complete data likelihood*:

$$L(\gamma \mid \mathbf{Z} = \mathbf{z}) = \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} (y_i - \theta_i)' \Sigma_i^{-1} (y_i - \theta_i) \right\} \tag{5}$$

$$\times \tau_t^{-k} \exp\left\{ -\frac{1}{2\tau_t^2} \sum_{i=1}^{k} (\theta_{ti} - \mu_t - \beta_\theta(\theta_{ri} - \mu_r))^2 \right\} \tag{6}$$

$$\times \tau_r^{-k} \exp\left\{ -\frac{1}{2\tau_t^2} \sum_{i=1}^{k} (\theta_{ri} - \mu_r)^2 \right\}. \tag{7}$$

Note that the complete data likelihood has the following simple form:

Likelihood = (constant not depending on unknown parameters)

× (likelihood from k observations from the simple linear regression (1))

× (likelihood from k samples from the normal population (2))

and depends only on the complete data through the *complete data sufficient statistics*, $T(\theta) = (\bar{\theta}, S_\theta)$, the sample mean and covariance matrix of the unobserved θ_i, where

$$\bar{\theta} = \frac{1}{k} \sum_{i=1}^{k} \theta_i = \begin{pmatrix} \bar{\theta}_t \\ \bar{\theta}_r \end{pmatrix} \tag{8}$$

$$S_\theta = \frac{1}{k} \sum_{i=1}^{k} (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})' = \begin{pmatrix} s_t^2 & s_{tr} \\ s_{tr} & s_r^2 \end{pmatrix} \tag{9}$$

(To simplify later notation we compute S_θ dividing by k, not $k - 1$). That is, *if* we could observe the complete data we would have (5)–(7) to work with rather than (4).

3.1. Maximum Likelihood Estimation by the EM algorithm

If we had actually observed the sufficient statistics, $T(\theta)$, we could compute the maximum likelihood estimate $\hat{\gamma}_{ml} = (\hat{\mu}_t, \hat{\mu}_r, \hat{\beta}_\theta, \hat{\tau}_t^2, \hat{\tau}_r^2)'$ as (for example, see Weisberg,[16] pp. 10–12):

M-step

$$\hat{\beta}_\theta = \frac{s_{tr}}{s_r^2}, \quad \hat{\mu}_t = \bar{\theta}_t, \quad \hat{\mu}_r = \bar{\theta}_r, \quad \hat{\tau}_r^2 = s_r^2, \quad \hat{\tau}_t^2 = s_t^2 - \hat{\beta}_\theta^2 s_r^2.$$

Conversely, if we knew γ, we could estmiate $T(\theta)$ by:

E-step

$$E(\bar{\theta} \mid \mathbf{Y} = \mathbf{y}, \gamma) = \bar{\theta}^* = \frac{1}{k} \sum_{i=1}^{k} \theta_i^*$$

$$E(S_\theta \mid \mathbf{Y} = \mathbf{y}, \gamma) = \frac{1}{k} \sum_{i=1}^{k} \Lambda_i^* + \frac{1}{k} \sum_{i=1}^{k} (\theta_i^* - \bar{\theta}^*)(\theta_i^* - \bar{\theta}^*)'$$

where θ_i^* and Λ_i^* are the posterior means and variances given in the lower right hand corner of Table III. With γ known, these expectations are simple to compute. Dempster *et al.*[14] showed that iterating between the two steps above, which use only the complete data likelihood (7), yields an estimate of γ that is a mode of the observed likelihood (4). First, with a current estimate $\gamma = \gamma^{(n)}$

('n' denoting the nth iteration), estimate the complete data sufficient statistics (E-step above) by setting γ equal to $\gamma^{(n)}$. This yields a current estimate of the sufficient statistic $T(\theta)^{(n)}$. Next, with the sufficient statistics set at the current estimate, $T(\theta) = T(\theta)^{(n)}$, compute the complete data maximum likelihood estimate $\hat{\gamma}_{\text{ml}}$ from (7) (M-step), and set the next value equal to this estimate, $\gamma^{(n+1)} = \hat{\gamma}_{\text{ml}}$. Repeat these steps until convergence.

For information on assessing convergence see Dempster *et al.*[14] or Gelman *et al.*[17]. One can compute asymptotic standard errors by the SEM algorithm,[18] by numerical approximation using the log of the observed likelihood (4) (the method used in Section 4, see Gelman *et al.*,[17] p. 273) or by the methods given in Fuller[9] (pp. 220, 221).

3.2. Bayes Estimation

For Bayes estimation we wish to investigate the entire posterior distribution proportional to the observed likelihood (4) multiplied by some prior, $\pi(\gamma)$. Because this does not have a convenient closed form, we estimate it by simulation, using the data augmentation algorithm (Tanner and Wong[15]). The data augmentation algorithm is a Markov Chain Monte Carlo (MCMC) algorithm and a special form of a Gibbs sampler. It is analogous to the EM algorithm, but instead of computation of expectations of missing data (E-step) and formation of maximum likelihood estimates of parameters (M-step), we simulate the missing data (Augmentation step), and simulate parameters (Parameter step). Notice that if we knew the parameter γ, we could simulate the missing data from the posterior distribution shown in Table III by:

Augmentation step:

$$\theta_i \sim \mathrm{N}(\theta_i^*, \Lambda_i^*), \quad \text{for } i = 1, \ldots, k$$

where θ_i^* and Λ_i^* are given in Table III. Conversely, if we had observed the missing data θ, we could simply sample a value of γ from the complete data posterior, as shown below.

We can rewrite the complete data likelihood (7) in the following form, which makes the parameters we wish to estimate more explicit (see Box and Tiao,[19] p. 114):

$$L(\gamma \mid \mathbf{Y} = \mathbf{y}, \theta) \tag{10}$$

$$\propto \tau_r^{-(k-1)} \exp\left\{-\frac{k s_r^2}{2\tau_r^2}\right\} \tag{11}$$

$$\times \tau_r^{-1} \exp\left\{-\frac{k}{2\tau_r^2}(\mu_r - \bar{\theta}_r)^2\right\} \tag{12}$$

$$\times \tau_t^{-(k-2)} \exp\left\{-\frac{k\hat{\tau}_t^2}{2\tau_t^2}\right\} \tag{13}$$

$$\times \tau_t^{-2} \exp\left\{-\frac{1}{2\tau_t^2}((\mu_t, \beta_\theta) - (\hat{\mu}_t, \hat{\beta}_\theta))' \mathbf{X}'\mathbf{X}((\mu_t, \beta_\theta)) - (\hat{\mu}_t, \hat{\beta}_\theta))\right\} \tag{14}$$

where $\bar{\theta}_r$ and s_r^2 are defined in (8) and (9), $\hat{\mu}_t$ and $\hat{\beta}_\theta$ are the usual slope and intercept estimates from an ordinary least squares regression of θ_{ti} on $d_i = \theta_{ri} - \mu_r$, $\hat{\tau}_t^2 = \frac{1}{k}\sum_{i=1}^{k}(\theta_{ti} - \hat{\mu}_t - \hat{\beta}_\theta d_i)^2$, which is the maximum likelihood estimate of the residual variation of the regression just mentioned, and $\mathbf{X}$ is the matrix of regressors $\mathbf{X} = ((1, d_1)', (1, d_2)', \ldots, (1, d_k)')'$.

The complete data posterior is proportional to this likelihood times a prior $\pi(\gamma)$. Here we use priors uniform on the location and slope parameters μ_r, μ_t and β_θ, but for the variance

components τ_t^2 and τ_r^2 we use the inverse-gamma priors. We may simulate γ by performing the following steps in order (see Box and Tiao[19], p. 114):

Parameter step:

$$\tau_r^2 \sim (\lambda_r + ks_r^2)/\chi^2_{q_r+k-1} \tag{15}$$

$$\mu_r \sim \mathrm{N}(\bar{\theta}_r, \tau_r^2/k) \tag{16}$$

$$\tau_t^2 \sim (\lambda_t + ks_t^2)/\chi^2_{q_t+k-2} \tag{17}$$

$$(\mu_t, \beta_\theta)' \sim \mathrm{N}_2((\bar{\mu}_t, \hat{\beta}_\theta)', X'X^{-1}\tau_t^2) \tag{18}$$

where q_t, q_r, λ_t and λ_r are the parameters of the inverse gamma distribution:

$$g(\tau^2)\mathrm{d}\tau^2 \propto \tau^{-(q+2)}\exp(-\lambda/2\tau^2)\mathrm{d}\tau^2. \tag{19}$$

The data augmentation algorithm proceeds by iterating between the augmentation and imputation steps. First, with a current parameter value $\gamma^{(n)}$ ('n' meaning the nth iteration), we impute the missing data as described in the Augmentation step. We denote the missing values imputed with parameter $\gamma^{(n)}$ as $\theta^{(n)}$. Second, we form $\bar{\theta}_t, \bar{\theta}_r, s_r^2, s_t^2, \hat{\mu}_t$ and $\hat{\beta}_\theta$ from $\theta^{(n)}$, and draw the next iteration $\gamma^{(n+1)}$ from the Parameter step (15)–(18). The sequence $\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, \ldots$, converges to the posterior distribution (see Gelfand and Smith[20]). We can use the resulting sequence to approximate the posterior moments and the posterior distribution of the parameters.

The choice of a starting value and the assessment of convergence for MCMC methods are very important, but are not discussed here is detail. In Section 4 we use the method* proposed by Gelman and Rubin[21] (see the companion piece by Geyer[22] for an alternative view). This method uses the normal approximation resulting from maximum likelihood estimation to select the starting values for several parallel MCMC simulations (thus maximum likelihood estimation is the first step for Bayes estimation). Convergence is determined when the between-simulation variability is dominated by the within-simulation variability (see Gelman *et al.*[17] for many examples).

3.3. Discussion

The estimate of $1/\tau_r^2$ determines how far the estimate of β_θ is from the observed slope (see Section 2.3). The uncertainty of β_θ also increases with large estimates of $1/\tau_r^2$. The intuition for this is that β_θ, being the coefficient of a simple linear regression model, has uncertainty determined by the ratio of the residual variation and the variation of the covariate (see Weisberg,[16] p. 23); here τ_t^2/τ_r^2. Because both of these effects (bias adjustment and variance) depend on $1/\tau_r^2$, the inference for this parameter is very important for estimating β_θ. In particular, inferences for β_θ are particularly sensitive to small values of τ_r^2.

With large k all reasonably vague priors result in similar inferences, but with small k inferences for β_θ are sensitive to the choice of λ_r and q_r. With $\lambda = 0$, prior distribution (19) provides two candidates that may be considered as vague prior distributions:† prior 1, uniform on the standard

* An Splus program that implements this method is available free. See instructions on p. 461 of Gelman and Rubin.[21]

† Jeffreys' prior, which is uniform on $\log(\tau^2)$, and corresponds to (19) with $q = 0$ and $\lambda = 0$ does not lead to a proper posterior. See DuMouchel and Waternaux[23] for a discussion. Recall that τ_t^2 and τ_r^2 are the heterogeneity of θ_t and θ_r, not of Y_t and Y_r.

deviation ($q = -1$), and prior 2, uniform on the variance ($q = -2$). Prior 2 places weight on larger values of τ^2 than does prior 1. If prior 1 or 2 is chosen for τ_r^2, the Bayes procedure typically estimates τ_r^2 larger, and results in smaller bias adjustment, than the maximum likelihood procedure. Prior 2, because it places weight on larger values of τ^2, typically results in less bias adjustment than prior 1.

Prior parameters λ and q may also be chosen to correspond to informative prior opinion. If opinion is expressed in the form of a mean and variance, or two prior quantiles (for example the 1st and 99th), λ and q can be chosen so that (19) matches this opinion (see Gelman, *et al.*[17], pp. 139–140, for a discussion).

The maximum likelihood estimate of τ_r^2 may be zero. When this happens the maximum likelihood estimates are invalid (if $\tau_r^2 = 0$, the structural model has a vertical line). When τ_r^2 approaches zero the covariance matrix Λ in Table III approaches singularity, and the maximum likelihood estimate of β_θ approaches $\pm\infty$ (see Fuller,[9] pp. 105–106, or this can be shown by following the steps of the EM algorithm as τ_r^2 approaches zero), but even if the maximum likelihood estimate of τ_r^2 is zero, the likelihood will support larger values of τ_r^2 and the Bayes estimates are valid. When this occurs, we may either conclude that there is no heterogeneity in the control group risk, or we must use the Bayes estimates. When k is small, unless there is good reason to think $\tau_r^2 = 0$, the Bayes estimates should be preferred.

4. EXAMPLES

In this section we apply the methods discussed above to estimate the structural model for the data in Tables I and II. First, we must estimate the measurement covariances Σ_i from the definitions of $f(\cdot)$, $h(\cdot)$, and α_{ti} and α_{ci}, possibly by a Taylor series method (see Appendix). As we have done throughout this paper we use the log-odds ratio of event rates for the TE definition and the control group event rates as the PR definition. For both data sets we use the usual binomial variance approximation for the variability of the $\hat{\alpha}_c$; $\hat{\sigma}_{ri}^2 = \hat{\alpha}_{ci}(1 - \hat{\alpha}_{ci})/n_{ci}$, where n_{ci} is the size of the control group sample for trial i. The Appendix gives the covariance approximation $\hat{\sigma}_{tri} = -1/n_{ci}$. For the magnesium data we use the common large sample approximation for $\hat{\sigma}_{ti}^2 = 1/n_{ti}\hat{\alpha}_{ti}(1 - \hat{\alpha}_{ti}) + 1/n_{ci}\hat{\alpha}_{ci}(1 - \hat{\alpha}_{ci})$, but for the β-mimetics data we approximate σ_{ti}^2 from the original data given in Keirse *et al.*,[24] which provides 95 per cent confidence intervals for treatment effects defined as odds-ratio (OR). We approximate the variance of the OR by assuming the 95 per cent confidence interval has width $2 \times 1{\cdot}96 \times \sigma_i$, where σ_i is the standard error estimate of the OR. From this we approximate σ_{ti}^2 with Taylor series method by $\hat{\sigma}_{ti}^2 \approx \hat{\sigma}_i^2/\mathrm{OR}^2$. These are the values shown in Tables I and II.

4.1. β-mimetics Trials

Brand and Kragt[8] summarized the data in Table II with a weighted least squares regression of the log-odds ratio on the control group event rates, and were able to explain a statistically significant amount of treatment effect heterogeneity. However, because their model assumes the log-odds ratio is a fixed covariate, their result may simply have been a consequence of measurement error. This was pointed out by Senn,[12] and in response Brand and Kragt[12] agreed, and then suggested correction for the correlated error by regressing the treatment effect defined as risk difference, $f(\alpha_t, \alpha_c) = \alpha_t - \alpha_c$, on the population risk defined as the total risk, $h(\alpha_t, \alpha_c) = \alpha_t + \alpha_c$. Although this redefinition does reduce the correlation, doing so only makes β_{yi} nearer to zero, and consequently may only change the direction of the bias. Here we estimate the association originally intended in Brand and Kragt.[8]

Table IV. Results from analysis of β-mimetics trials. Structural model maximum likelihood and Bayes estimates are from the methods discussed in this paper. The Bayes estimates are posterior means and the values in parentheses are standard errors for the maximum likelihood estimates and posterior standard deviations for the Bayes estimates. The Bayes estimates are labelled by the prior distributions used for the variance components τ_t^2 and τ_r^2. For example, the estimate labelled $q_t = -1$, $q_r = -2$, is from a Bayesian analysis with priors uniform on τ_t^2 and τ_r^2. The row labelled 'random effects' gives the estimate of the observed relationship ignoring the measurement error, and computed as recommended by Morris.[13] The final column gives the one-sided p-value (for maximum likelihood and random effects estimation) of the one sided test H_0: $\beta_\mu = 0$, and the posterior probability (for Bayes estimates) of a non- negative structural slope; $P(\beta_\mu > 0 | \mathbf{Y})$

Method	μ_t	μ_r	β_θ	τ_t	τ_r	p-value/ prob
β-mimetics data parameter estimates						
ML	−1·099 (0·235)	0·384 (0·054)	−1·425 (1·411)	0·037 (0·297)	0·172 (0·044)	0·156
Bayes:						
$q_t = -1, q_r = -2$	−1·158 (0·280)	0·389 (0·031)	−1·613 (1·812)	0·350 (0·260)	0·220 (0·067)	0·160
$q_t = -2, q_r = -2$	−1·160 (0·316)	0·388 (0·031)	−1·783 (1·824)	0·556 (0·355)	0·220 (0·065)	0·148
$q_t = -2, q_r = -1$	−1·155 (0·303)	0·385 (0·030)	−1·696 (1·896)	0·514 (0·310)	0·202 (0·060)	0·175
Random effects	−1·163 (0·226)		−1·954 (1·341)	0		0·072
Magnesium data parameter estimates						
ML	−0·290 (0·289)	0·094 (0·020)	−12·948 (2·991)	0 (NA)	0·048 (0·019)	0·001
Bayes:						
$q_t = -1, q_r = -2$	−0·395 (0·269)	0·097 (0·013)	−13·438 (5·31)	0·277 (0·260)	0·071 (0·031)	0·015
$q_t = -2, q_r = -2$	−0·482 (0·341)	0·097 (0·013)	−13·401 (7·180)	0·508 (0·382)	0·068 (0·030)	0·033
$q_t = -2, q_r = -1$	−0·479 (0·351)	0·096 (0·012)	−12·55 (8·228)	0·523 (0·429)	0·059 (0·024)	0·051
Random effects	−10370 (0·087)		−13·287 (2·709)	0		0·001

The top half of Table IV tabulates four estimates of the β-mimetics structural model; one maximum likelihood and three Bayes estimates. For comparison we also show the results of a random effects regression that ignores the measurement error. Each of the structural model slope estimates given in Table IV are closer to zero than the estimates that ignore the measurement error (the random effects estimates), and so we can expect that the line in Figure 1 exaggerates the true strength of the structural relationship. Figure 2 plots the estimated structural models. Treating the control group event rate as a fixed covariate gives reasonable evidence for a negative association (the one-sided p-value for testing $\beta < 0$ is $p < 0{\cdot}072$), but the structural model estimates tell another story. The maximum likelihood procedure gives an asymptotic one-sided p-value of $p = 0{\cdot}156$, and the Bayes procedures give the probability that the structural relationship is positive as greater than 0·148. Figure 3 shows a posterior distribution of β_θ. Thus for the β-mimetics data we cannot conclude with any reasonable certainty that there exists a true association between the effect of the treatment and the risk of the population.

Overall we can be confident that β-mimetics is beneficial. The posterior distribution of μ_t, the mean treatment effect, gives almost complete support to values that are less than zero. Figure 3 shows a posterior distribution of μ_t.

4.2. Magnesium Trials

The lower half of Table IV summarizes the four structural model estimates for the magnesium data. Each of the structural slope estimates is similar to the observed association (compares to random effects), suggesting that the magnitude of the association in Figure 1 is not highly biased. Figure 4 plots the estimated structural regression lines.

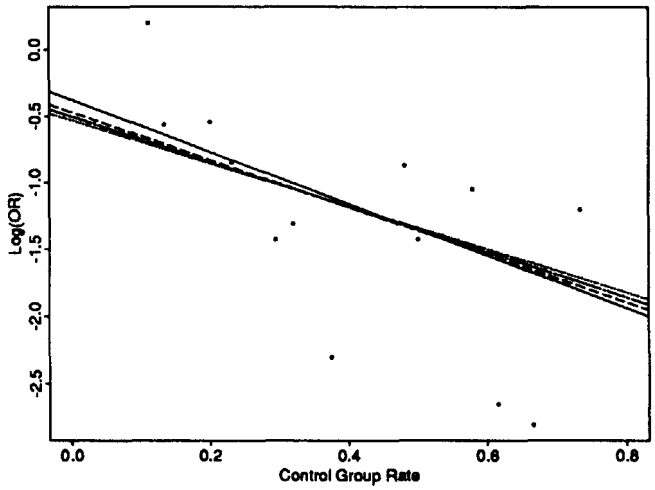

Figure 2. Estimated structural relationships for the β-mimetics data: solid line (——) observed regression line estimated by random effects regression; dotted line (.....), Bayes estimate with $q_t = -1$, $q_r = -2$; dashed line (---) Bayes estimated with $q_t = q_c = -2$; mixed (-·-·-) Bayes estimate with $q_t = -2$, $q_r = -1$

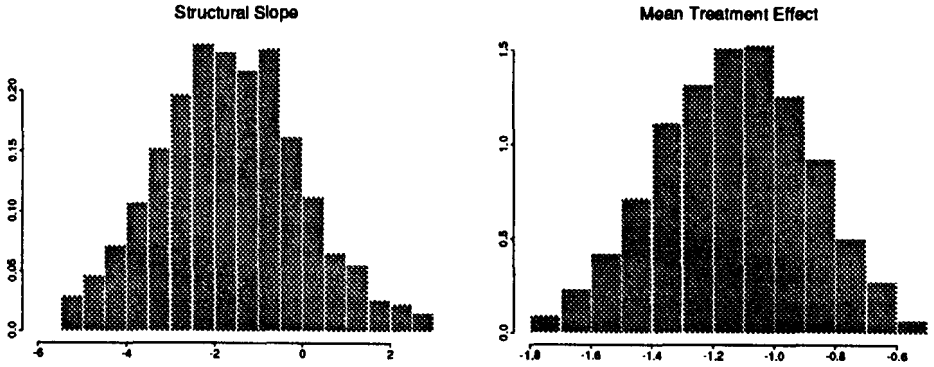

Figure 3. Estimated marginal posterior distributions for the structural slope β_θ and mean treatment effect μ_t for the β-mimetics data. This is from the Bayes analysis with variance component priors $q_t = -1$, $q_r = -2$

The conclusion of a negative structural slope may have important policy implications. The largest single clinical trial, ISIS 4 with over 50,000 patients, suggests no benefit and possible harm from magnesium therapy, and taken alone this trial may lead to the conclusion that magnesium therapy is not beneficial, but a structural model with a non-zero slope may have a critical control group mortality rate that estimates a null treatment effect; $\zeta = -\mu_t/\beta_\theta + \mu_r$. Populations with control group rates on one side of ζ can expect benefit from magnesium, and population on the other side can expect harm from magnesium. Because the ISIS 4 trial has a mortality rate of 6·92 per cent, Figure 1 suggests no benefit, possibly because the population treated was too 'healthy' but magnesium therapy may still have benefit for less healthy populations. The structural models shown in Figure 4 provide even greater evidence for the benefit of magnesium than does the observed line. Figure 4 shows each of the structural regression lines beneath the random effect estimate, suggesting that magnesium has greater benefits, and benefits healthier population, than suggested by the observed line.

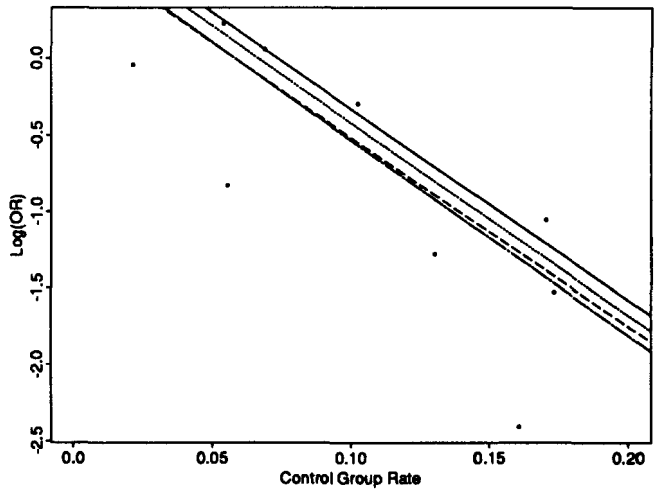

Figure 4. Estimated structural relationships for the magnesium data: solid line (——) observed regression line estimated by random effects regression; dotted line (.....), Bayes estimate with $q_t = -1$, $q_r = -2$; dashed line (---) Bayes estimated with $q_t = q_c = -2$; mixed (-·-·-) Bayes estimate with $q_t = 2$, $q_r = -1$

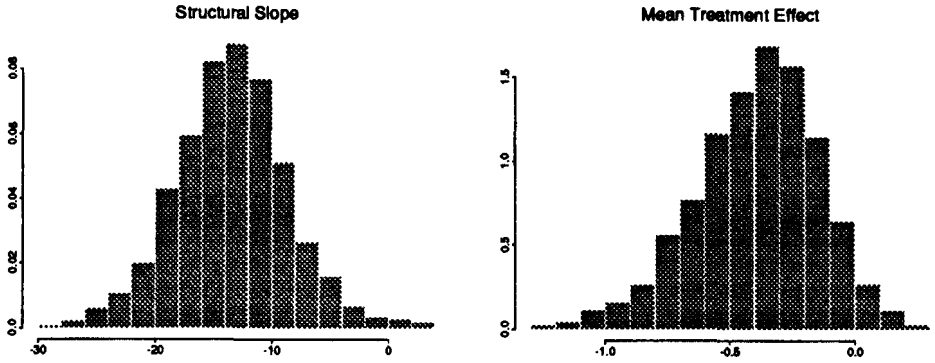

Figure 5. Posterior distribution of the control group mortality rate which will yield a null effect of magnesium, computed from the marginal posterior distributions with $q_t = -1$, $q_r = -2$. The mean of this distribution is 0·062 and the 5th, 50th and 95th percentiles are 0·016, 0·073 and 0·089, respectively

We may investigate the critical value ζ by computing the posterior distribution of ζ (defined above) from the posterior distribution of γ. We simply compute $\zeta^{(i)}$ from each $\gamma^{(i)}$ sampled in the data augmentation scheme and produce a histogram. Figure 5 shows a marginal posterior distribution for ζ. The 95th percentile of this posterior distribution is 0·089, meaning that with probability 0·95 populations with control group mortality rates above 8·9 per cent will have an expected benefit from magnesium.

We can evaluate the overall effect of magnesium with the parameter μ_t. Figure 6 shows a posterior distribution for this parameter. This distribution is skewed to the left, with only 4·6 per cent of the distribution greater than zero, suggesting that the mean effect of magnesium is beneficial (the probabilities computed from the other posterior distributions are all less than 0·05).

There is large disagreement between the uncertainties given by the Bayes estimates and the maximum likelihood estimate; the latter gives an asymptotic p-value of $p < 0{\cdot}001$. Theory

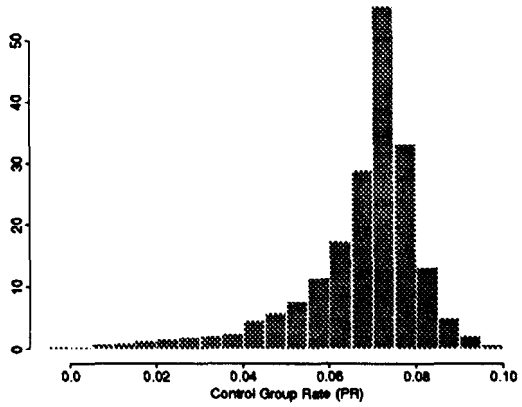

Figure 6. Estimated marginal posterior distributions for the structural slope β_θ and mean treatment effect μ_t for the magnesium data. This is from the Bayesian analysis with variance component priors $q_t = -1$, $q_r = -2$

suggests that if there are sufficient data, the point estimates and standard errors from the Bayes and maximum likelihood procedures will agree, and the posterior distribution of the parameters should look as a normal distribution. The disagreement between the Bayes and maximum likelihood estimates, and the skewness in the posterior distributions (see Figure 6), suggest that there is an insufficient number of clinical trials to treat the maximum likelihood estimates as valid for these data.

5. SUMMARY AND CONCLUSIONS

An observed relationship between the estimated effect of a treatment and the estimated risk of the population may indicate that important differences in the populations relate to the effect of a treatment, or it may result from measurement error. Applying the framework of measurement error models to this problem can correct for the biasing affects.

Maximum likelihood and Bayes estimation is straightforward with procedures of missing data. The Bayes estimates have many advantages over the maximum likelihood estimates. With small samples the asymptotic standard errors of the maximum likelihood estimates may be too small, and because the likelihood may be significantly skewed) as one can see through the marginal posterior distributions), symmetric confidence intervals formed may be meaningless. Also, one can use samples from the marginal posterior distributions to make further inferences about functions of the parameters, as was done to investigate the point of null treatment effect, ζ, for the magnesium data.

To make the methods discussed here most flexible to different choices of $h(\cdot)$ and $f(\cdot)$, Section 2 assumed unbiased normal measurement errors with known variance. Although the bias results demonstrated in Section 2.3 do not depend on normality of the errors, the estimation procedures do, and so one must assess the validity of these assumptions for each application. Often one can justify these by the central limit theorem when large within-trial samples exist, but, because of the low event rates in Tables I and II, these assumptions do not hold for the data given here. In this case we recommended evaluation of the procedure presented here by simulation. When applying the methods given here to data of the type in Tables I and II (binomial measurements, small number of trials, low event rates per trial, log-odds ratio treatment effect definition), simulations show a considerable reduction in bias and mean-squared-error compared

to methods that ignore the measurement error, whether weighted least squares or random effects regression. The standard errors and posterior standard deviations are also wider than with methods that ignores the measurement error, giving better coverage properties.

Although here we treat controlling for population risk, the model and estimation procedures described are appropriate for meta-analysis whenever a covariate of interest is an aggregate measurement of the treated population. For example, we may wish to explain heterogeneity of a treatment effect by the mean age of the population treated; $\mathrm{TE}_i = \beta_0 + \beta_1 \times \text{mean age}_i$. Here, β_0 and β_1 are structural regression parameters. If for each clinical trial we estimate the mean age by the average age of the individuals treated ($\overline{\text{age}}$), the regression $\hat{\mathrm{T}}\mathrm{E}_i = \beta_0 + \beta_1 \times \overline{\text{age}}$ is biased towards the within-trial regression slope ($\hat{\mathrm{T}}\mathrm{E}_i$ is the estimated mean TE for trial i). One can use the model and estimation methods described here to correct for this measurement error as follows. In the notation used throughout this paper, setting $Y_i = (\hat{\mathrm{T}}\mathrm{E}_i, \overline{\text{age}})'$ and $\Sigma_i = S_y^2/n_i$, where S_y^2 is the sampling variance matrix for Y_i, and applying the estimation procedures given here one will estimate the structural model that corrects for the measurement error.

APPENDIX: COVARIANCE APPROXIMATIONS

Let $Y_t \equiv f(\hat{\alpha}_t, \hat{\alpha}_c)$ and $Y_r \equiv h(\hat{\alpha}_t, \hat{\alpha}_c)$, where $\hat{\alpha}_t \sim \mathrm{N}(\alpha_t, \sigma_t^2)$ and $\hat{\alpha}_c \sim \mathrm{N}(\alpha_c, \sigma_r^2)$ are independent. We abbreviate $\alpha = (\alpha_t, \alpha_c)'$. A Taylor series expansion of $f(\cdot)$ and $h(\cdot)$ around α gives the following covariance approximation:

$$\begin{aligned}\operatorname{cov}(Y_t, Y_r) &\approx \operatorname{cov}\{(\hat{\alpha} - \alpha)'\nabla f(\alpha), (\hat{\alpha} - \alpha)\nabla h(\alpha)\} \\ &= \frac{\partial f(\alpha)}{\partial \alpha_t}\frac{\partial h(\alpha)}{\partial \alpha_t}\operatorname{var}(\hat{\alpha}_t) + \frac{\partial f(\alpha)}{\partial \alpha_c}\frac{\partial h(\alpha)}{\partial \alpha_c}\operatorname{var}(\hat{\alpha}_c) + \left(\frac{\partial f(\alpha)}{\partial \alpha_t}\frac{\partial h(\alpha)}{\partial \alpha_c} + \frac{\partial f(\alpha)}{\partial \alpha_c}\frac{\partial h(\alpha)}{\partial \alpha_t}\right)\operatorname{cov}(\hat{\alpha}_t, \hat{\alpha}_c) \qquad (20) \\ &= \frac{\partial f(\alpha)}{\partial \alpha_t}\frac{\partial h(\alpha)}{\partial \alpha_t}\sigma_t^2 + \frac{\partial f(\alpha)}{\partial \alpha_c}\frac{\partial h(\alpha)}{\partial \alpha_c}\sigma_c^2 \qquad (21)\end{aligned}$$

Expression (20) follows from the linear terms of Taylor series expansion and (21) follows from the independence of $\hat{\alpha}_t$ and $\hat{\alpha}_c$. Thus an estimate for the covariance between Y_t and Y_r is

$$\hat{\sigma}_{tr} = \frac{\partial f(\hat{\alpha})}{\partial \hat{\alpha}_t}\frac{\partial h(\hat{\alpha})}{\partial \hat{\alpha}_t}\sigma_t^2 + \frac{\partial f(\hat{\alpha})}{\partial \hat{\alpha}_c}\frac{\partial h(\hat{\alpha})}{\partial \hat{\alpha}_c}\sigma_c^2 \qquad (22)$$

In this paper we used $f(\alpha_t, \alpha_c) = \log(\alpha_t/1 - \alpha_c) - \log(\alpha_c/1 - \alpha_c)$ and $h(\alpha_c) = \alpha_c$. With α_c defined as a mortality rate we have $\sigma_c^2 = \alpha_c(1 - \alpha_c)/n_c$, where n_c is the sample size of the control group, $\partial f(\alpha)/\partial \alpha_c = -1/\alpha_c(1 - \alpha_c)$, $\partial h(\alpha)/\partial \alpha_c = 1$, and $\partial h(\alpha)/\partial \alpha_t = 0$. Thus our approximation (22) becomes $\hat{\sigma}_{tr} = -1/n_c$.

ACKNOWLEDGEMENTS

Supported by AHCPR Proposal No. 1R01HS07118-01, and grant no. PORT-HS06341. The author would like to thank Elliott Antman, Catherine Berkey, Frederick Mosteller, Carl Morris and Christopher Schmid for many valuable discussions, and Frederick Mosteller for comments on numerous drafts of this manuscript.

REFERENCES

1. National Research Council. *Combining Information: Statistical Issues and Opportunities for Research*, National Academy Press, Washington, D.C., 1992.

2. DerSimonian, R. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
3. DuMouchel, W. H. and Harris, J. E. 'Bayes methods for combining the results of cancer studies in humans and other species (with comment)', *Journal of the American Statistical Association*, **78**, 293–308 (1983).
4. Hedges, L. V. and Olkin, I. *Statistical Methods for Meta-analysis*, Academic Press, 1985.
5. Morris, C. N. and Normand, S. L. 'Hierarchical models for combining information and for meta-analyses (with discussion)', *in* Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics*, 4, Oxford University Press, 1992, pp. 321–335.
6. Berkey, C. S., Hoaglin, D. C., Mosteller, F. and Colditz, G. A. 'A random-effects regression model for meta-analysis', *Statistics in Medicine*, **14**, 395–411 (1995).
7. Schmid, H.C., McIntosh, M. W., Cappelleri, J. C., Lau, J. and Chalmers, T. C. 'Measuring the impact of the control rate in meta-analysis of clinical trials', abstract in *Controlled Clinical Trials*, **16**, 665 (1995).
8. Brand, R. and Kragt, H. 'Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials', *Statistics in Medicine*, **11**, 2077–2082 (1992).
9. Fuller, W. A. *Measurement Error Models*, Wiley, 1987.
10. Rosner, B. and Willett, W. C. 'Interval estimates for correlation coefficients corrected for within-person variation: Implications for study design and hypothesis testing (C/R: V131/1990 p. 573–575)', *American Journal of Epidemiology*, **127**, 377–386 (1988).
11. Carrol, R. J. and Stefanski, L. A. 'Measurement error, instrumental variables and corrections for attenuation with applications to meta-analysis', *Statistics in Medicine*, **13**, 1265–1282 (1994).
12. Senn, S. 'Letters to the Editor', *Statistics in Medicine*, **13**, 293–296 (1994).
13. Morris, C. N. 'Parametric empirical Bayes inference: theory and applications (with discussion)', *Journal of the American Statistical Association*, **78**, 515–529 (1983).
14. Dempster, A. P., Laird, N. M. and Rubin, D. B. 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society, Series B*, **39**, 1–22 (1977).
15. Tanner, M. A. and Wong, W. H. 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association*, **82**, 528–540 (1987).
16. Weisberg, S. *Applied Linear Regression*, Wiley, 1985.
17. Gelman, A., Carlin, C. B., Stern, H. S. and Rubin, D. B. *Bayesian Data Analysis*, Chapman & Hall, 1995.
18. Meng, X. and Rubin, D. B. 'Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm', *Journal of the American Statistical Association*, **86**, 899–909 (1991).
19. Box, G. E. P. and Tiao, G. C. *Bayesian Inference in Statistical Analysis*, Wiley, 1992.
20. Gelfand, A. E. and Smith A. F. M. 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1992).
21. Gelman, A. and Rubin, D. B. 'Inference from iterative simulation using multiple sequences', *Statistical Science*, **7**, 457–472 (1992).
22. Geyer, C. S. 'Practical Markov chain Monte Carlo', *Statistical Science*, **7**, 473–483 (1992).
23. DuMouchel, W. and Waternaux, C. Comment on 'Hierarchical models for combining information and for meta-analyses', *in* Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), Oxford University Press, 1992, pp. 338–339.
24. Keirse, M. J. N. C., Grant, A. and King, J. F., in Chalmers, I., Enkin, M. and Keirse, M. J. N. C. (eds) *Labour, Effective Care in Pregnancy and Childbirth*, Oxford University Press, 1989, pp. 694–745.