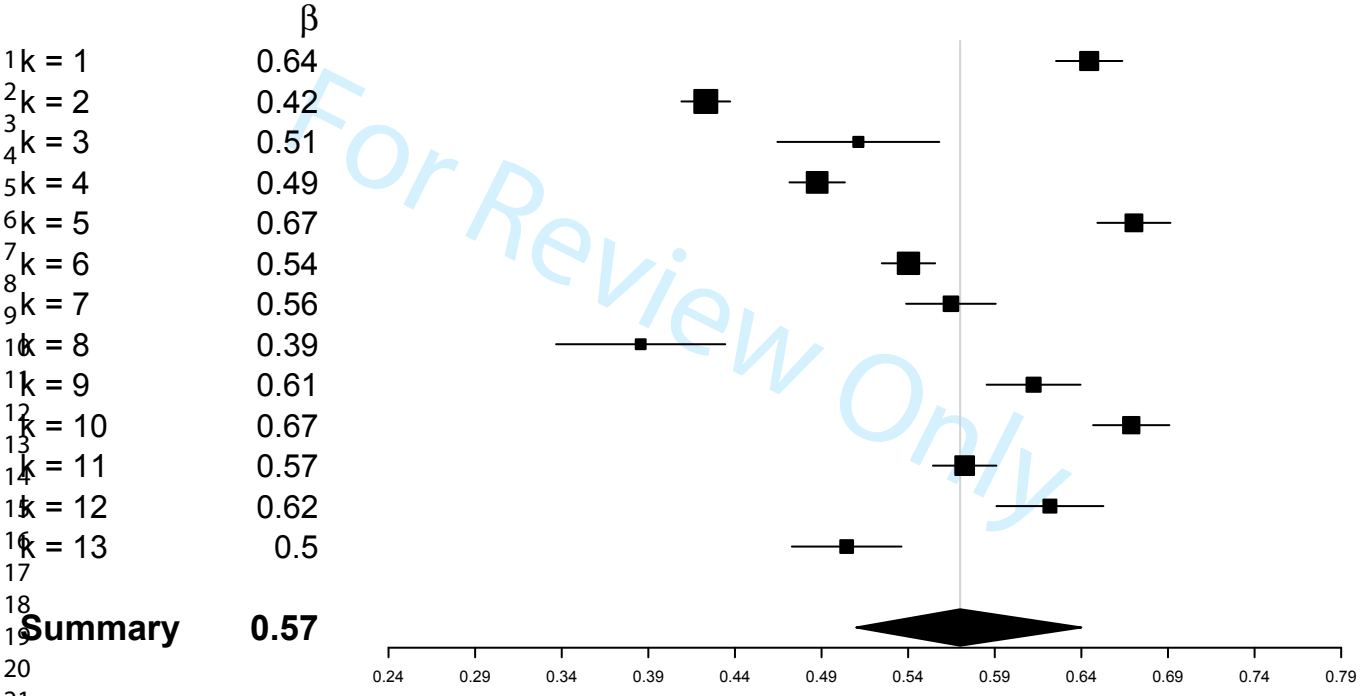
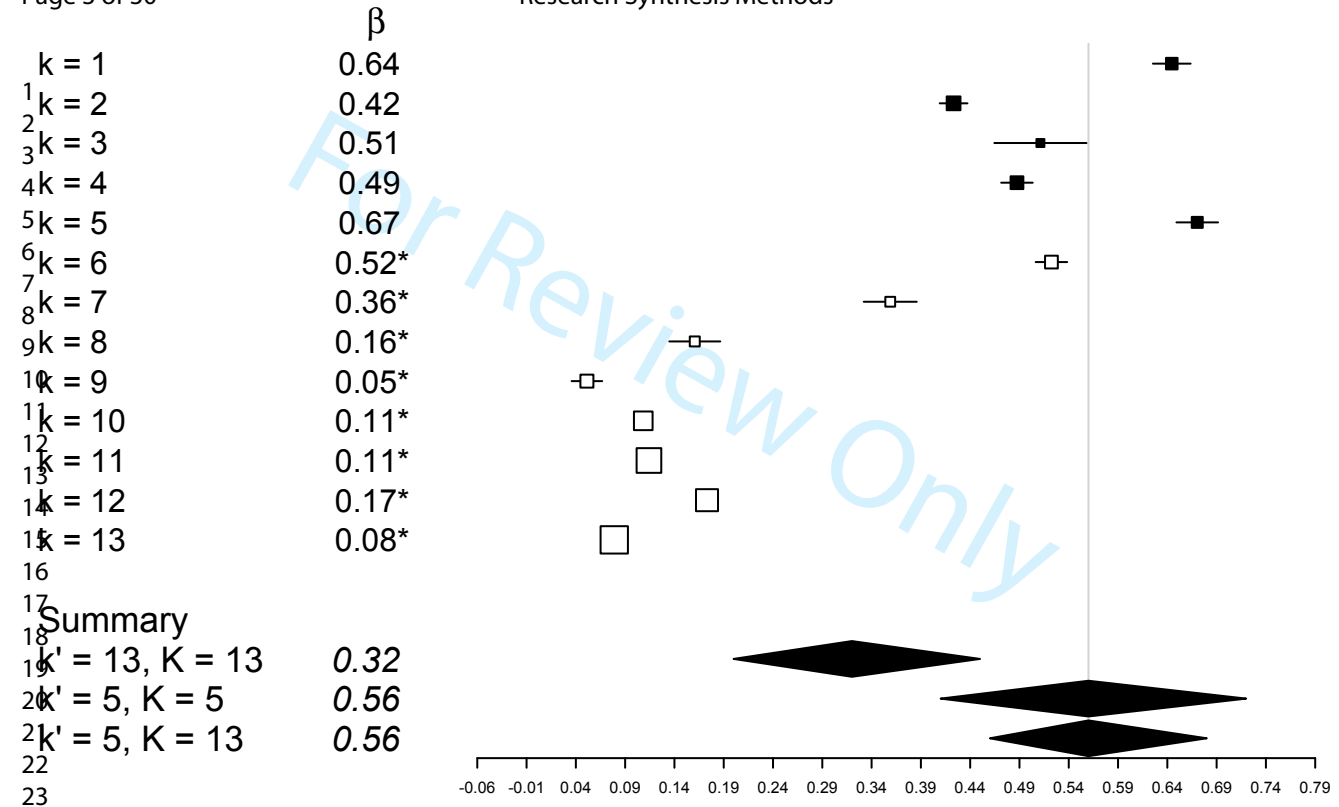
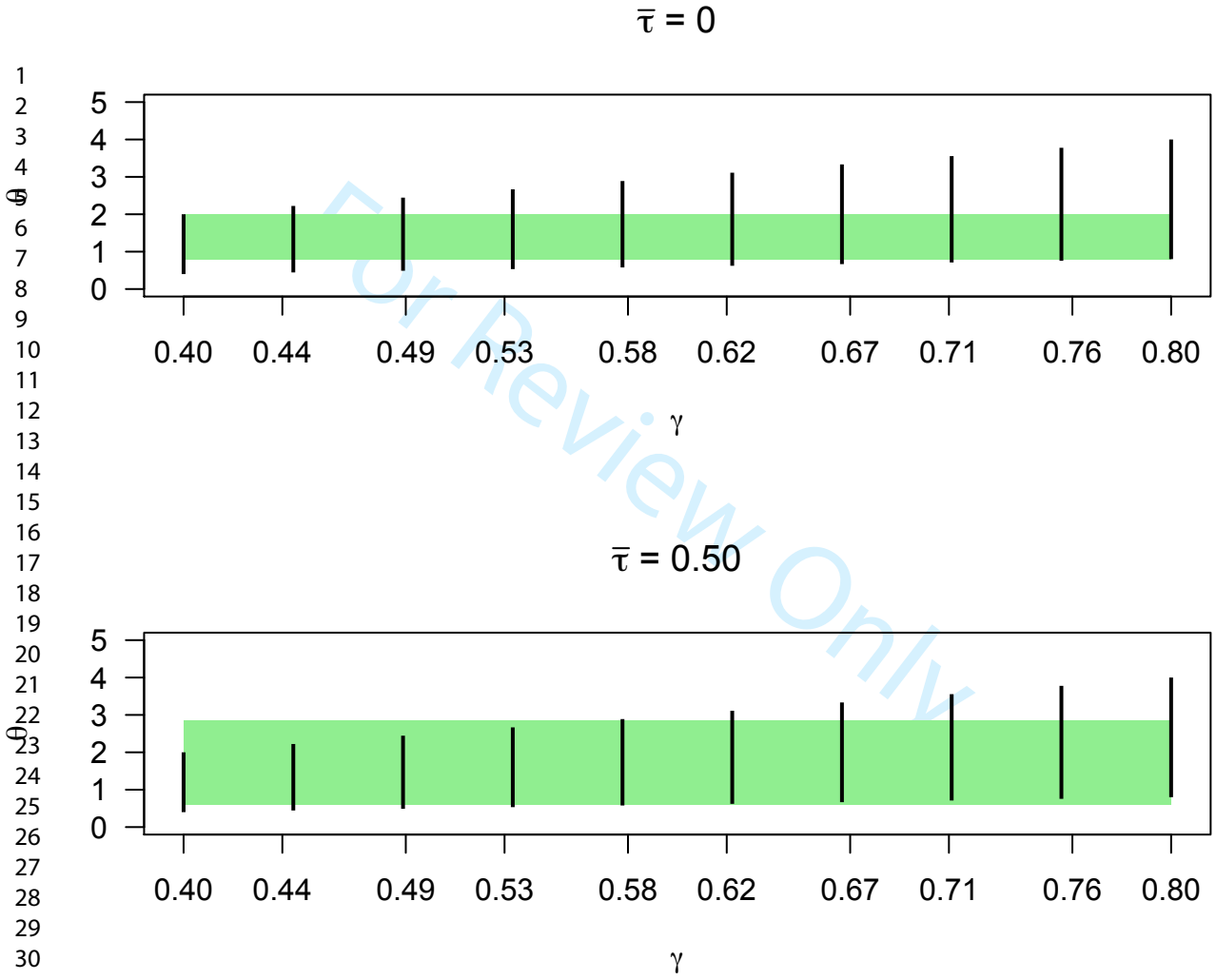
Measurement Error in Meta-Analysis (MEMA) - \\ a Bayesian framework for continuous outcome data

Journal:	*Research Synthesis Methods*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Campbell, Harlan; The University of British Columbia, Statistics de Jong, Valentijn; Utrecht University, Julius Center for Health Sciences and Primary Care Maxwell, Lauren; University Hospital Heidelberg, Heidelberg Institute for Global Health Debray, Thomas; University Medical Center Utrecht, Julius Center for Health Sciences and Primary Care; University Medical Center Utrecht, The Dutch Cochrane Centre, Julius Center for Health Sciences and Primary Care Jaenisch, Thomas; University Hospital Heidelberg, Heidelberg Institute for Global Health; Colorado School of Public Health, Department of Epidemiology Gustafson, Paul; University of British Columbia, Statistics
Manuscript Keywords:	Measurement error, Meta-analysis, Bayesian evidence synthesis, partial identification
Keywords to match reviewers:	Stats: Meta-analytical models < Methods Areas, Measurement < Methods Areas, Reporting Bias < General Methods < Methods Areas
Abstract:	Ideally, a meta-analysis will summarize data from several unbiased studies. Here we consider the less than ideal situation in which contributing studies may be compromised by measurement error. Measurement error affects every study design, from randomized controlled trials to retrospective observational studies. We outline a flexible Bayesian framework for continuous outcome data which allows one to obtain appropriate point and interval estimates with varying degrees of prior knowledge about the magnitude of the measurement error. We also demonstrate how, if individual-participant data (IPD) are available, the Bayesian meta-analysis model can adjust for multiple participant-level covariates, measured with or without measurement error.







Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

Measurement Error in Meta-Analysis (MEMA) - a Bayesian framework for continuous outcome data

Harlan Campbell*[1] | Valentijn M.T. de Jong[2] | Lauren Maxwell[3] | Thomas P.A. Debray[2,4] | Thomas Jaenisch[3,5] | Paul Gustafson[1]

[1]Department of Statistics, University of British Columbia, BC, Canada
[2]Julius Center for Health Sciences and Primary Care, Utrecht University, Utrecht, The Netherlands
[3]Heidelberg Institute for Global Health, Heidelberg University Hospital, Heidelberg, Germany
[4]Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[5]Department of Epidemiology, Colorado School of Public Health, CO, USA

Correspondence
*Harlan Campbell, Email: harlan.campbell@stat.ubc.ca

Summary

Ideally, a meta-analysis will summarize data from several unbiased studies. Here we consider the less than ideal situation in which contributing studies may be compromised by measurement error. Measurement error affects every study design, from randomized controlled trials to retrospective observational studies. We outline a flexible Bayesian framework for continuous outcome data which allows one to obtain appropriate point and interval estimates with varying degrees of prior knowledge about the magnitude of the measurement error. We also demonstrate how, if individual-participant data (IPD) are available, the Bayesian meta-analysis model can adjust for multiple participant-level covariates, measured with or without measurement error.

KEYWORDS:
meta-analysis, measurement error, misclassification, partial identification, Bayesian evidence synthesis

1 | INTRODUCTION

Increasingly often, traditional meta-analysis methods are used to synthesize results from multiple observational studies such as epidemiological surveys, cohort studies, and diagnostic test accuracy studies [1,2,3]. Observational studies are, by definition, non-randomized and are notoriously prone to a wide range of biases [4]. For instance, since exposure variables in an observational study are typically measured using imperfect tools (e.g., questionnaires, surveys, public health records), results are susceptible to "bias caused by measurement error" [5].

If the measurement error affecting a particular study is of known magnitude, adjustment for bias can be achieved by modifying the study's effect size estimate and uncertainty interval prior to its inclusion in a meta-analysis. Typically, however, the magnitude of measurement error in any particular study is unknown, and appropriate adjustments are rarely done [6,7]. To be clear,

issues of measurement error are not restricted to observational studies. Indeed, measurement error is potentially problematic for a wide range of research studies regardless of study design. However, the assumption of no measurement error (or that measurement error does not affect the results) becomes more difficult to defend when the exposure of interest is not randomized or when variables of interest are difficult to quantify (e.g. no gold standard measurement tool exists, social constructs, stigmatized behaviors).

In a meta-analysis of observational studies, failure to acknowledge and appropriately adjust for the possibility of measurement error amongst contributing studies will no doubt weaken or even invalidate the overall results[8]. And yet measurement error has received relatively little attention in the meta-analysis literature.

Hunter and Schmidt (2004) discuss various pragmatic statistical approaches to correct for known impacts of measurement error[9]; and see more recently Wiernik et al. (2020)[10]. Other work includes Carroll et al. (1991)[11] who consider the merits of various attenuation factors to correct for measurement error in a meta-analysis; and more recently: Lian et al. (2019)[12] who introduce Bayesian meta-analysis models accounting for exposure misclassification, and Zeisser (2014)[13] who discuss the bias due to measurement error in a meta-analysis estimating the relationship between alcohol consumption and breast cancer.

In this paper, we consider a meta-analysis of observational studies with continuous outcome and exposure variables in which (a subset of) contributing studies may be compromised by non-differential measurement error, i.e., error in the exposure variable(s) that is conditionally independent of the outcome variable[14]. We then develop a Bayesian hierarchical model to adjust for the measurement error when either aggregate study-level data or individual participant-level data (IPD) are available.

Bayesian methods for handling measurement error are well established for single studies and offer "a number of statistical advantages"[15]. Bayesian methods also offer a "number of specific advantages" for meta-analysis[16]. In Section 2, we outline a proposed Bayesian framework for the case of meta-analysis with measurement error with a single exposure variable and in Section 3, we generalize this framework for the case of multiple explanatory variables. We conclude with a summary of findings in Section 4.

2 | META-ANALYSIS OF SIMPLE LINEAR REGRESSION AGGREGATE DATA

2.1 | A traditional random-effects meta-analysis

Suppose we have data from K independent observational studies for a meta-analysis. For the k-th study in $1, \dots, K$, there are $n^{[k]}$ independent observations, and the total sample size is therefore $N = \sum_{k=1}^{K} n^{[k]}$. For each study, both the exposure and the outcome are continuous variables. Let $\left(X_j^{[k]}, Y_j^{[k]}\right)$ be the exposure and outcome for the j-th observation in the k-th study.

The sufficient statistics from each study can be summarized using a simple linear regression model: $Y_j^{[k]} | X_j^{[k]} \sim \mathcal{N}\left(\alpha^{[k]} + \beta^{[k]} X_j^{[k]}, \sigma^{[k]2}\right)$; for k in $1, \dots, K$ and j in $1, \dots, n^{[k]}$. We also assume that each study has its own exposure distribution governed by a univariate Normal distribution: $X_j^{[k]} \sim \mathcal{N}\left(\mu^{[k]}, \lambda^{[k]2}\right)$.

For each of the K studies, the observable aggregate data are the sufficient statistics: $\hat{\alpha}^{[k]}$, $\hat{\beta}^{[k]}$, $\sigma^{[k]2}$, and $\lambda^{k]2}$. The estimated regression coefficients, $\hat{\alpha}^{[k]}$ and $\hat{\beta}^{[k]}$, follow a bi-variate Normal distribution such that:

$$\begin{matrix}\hat{\alpha}^{[k]} \\ \hat{\beta}^{[k]}\end{matrix}\Bigg| \begin{matrix}\alpha^{[k]} \\ \beta^{[k]}\end{matrix} \sim \mathcal{N}\left(\begin{pmatrix}\alpha^{[k]} \\ \beta^{[k]}\end{pmatrix}, \begin{pmatrix}\Sigma_{11}^{[k]} & \Sigma_{12}^{[k]} \\ \Sigma_{12}^{[k]} & \Sigma_{22}^{[k]}\end{pmatrix}\right), \tag{1}$$

$$\text{where:} \quad \Sigma_{11}^{[k]} = \frac{(\lambda^{[k]2} + \mu^{[k]2}) \times \sigma^{[k]2}}{\lambda^{[k]2} \times n^{[k]}}; \tag{2}$$

$$\Sigma_{22}^{[k]} = \frac{\sigma^{[k]2}}{\lambda^{[k]2} \times n^{[k]}}; \text{ and} \tag{3}$$

$$\Sigma_{12}^{[k]} = -\mu^{[k]} \times \frac{\sigma^{[k]2}}{\lambda^{[k]2} \times n^{[k]}}. \tag{4}$$

We consider a Normal random-effects meta-analysis model which assumes that the regression coefficient parameters, $\alpha^{[k]}$ and $\beta^{[k]}$, vary between studies with:

$$\alpha^{[k]} \sim \mathcal{N}(\xi, \omega^2), \tag{5}$$

and

$$\beta^{[k]} \sim \mathcal{N}(\theta, \tau^2). \tag{6}$$

for k in $1, \ldots, K$, where ξ is the overall mean intercept parameter, ω^2 represents the variance in intercepts across studies, θ is the overall mean slope parameter, and τ^2 is the variance of slopes across studies. Note that θ is the primary parameter of interest and that τ^2 is only one of four different sources of variance. The meta-analysis model also includes: $\sigma^{[k]2}$, $\lambda^{[k]2}$ and ω^2. The $\sigma^{[k]2}$ parameter represents the residual variance from the k-th study, and the $\lambda^{[k]2}$ parameter represents the variance in the distribution of exposures in the k-th study. The ω^2 parameter represents the across-study variance in baseline levels.

While we assume that the regression coefficient parameters, $\alpha^{[k]}$ and $\beta^{[k]}$, are independent of each other, for all $k = 1, \ldots, K$, their corresponding parameter estimates, $\hat{\alpha}^{[k]}$ and $\hat{\beta}^{[k]}$, are correlated. The unconditional distribution of $(\hat{\alpha}^{[k]}, \hat{\beta}^{[k]})$ is bivariate Normal such that:

$$\begin{pmatrix}\hat{\alpha}^{[k]} \\ \hat{\beta}^{[k]}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\xi \\ \theta\end{pmatrix}, \begin{pmatrix}\Sigma_{11}^{[k]} + \omega^2 & \Sigma_{12}^{[k]} \\ \Sigma_{12}^{[k]} & \Sigma_{22}^{[k]} + \tau^2\end{pmatrix}\right), \tag{7}$$

for k in $1, \ldots, K$, where $\Sigma_{11}^{[k]}$, $\Sigma_{12}^{[k]}$, and $\Sigma_{22}^{[k]}$ are given by equations (2), (3), and (4) above.

Note that, for $k = 1, \ldots, K$, the $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$ parameters are assumed to be known, and equal to their observed estimates ($\hat{\mu}^{[k]}$, $\hat{\sigma}^{[k]}$, and $\hat{\lambda}^{[k]}$) obtained from the individual studies. So long as the number of observations in each study is sufficiently large, this simplifying assumption should make little practical difference[16].

Finally, note that aggregate data typically reported in a study may not include $\hat{\mu}^{[k]}$ and $\hat{\lambda}^{[k]}$ and instead include standard errors for the regression coefficients, se($\hat{\alpha}^{[k]}$) and se($\hat{\beta}^{[k]}$). This is only a minor inconvenience since values for $\hat{\mu}^{[k]}$ and $\hat{\lambda}^{[k]}$ can be

easily obtained from the given values of $\text{se}(\hat{\alpha}^{[k]})$, $\text{se}(\hat{\beta}^{[k]})$, and $\hat{\sigma}^{[k]}$ as follows:

$$\hat{\lambda}^{[k]} = \frac{\hat{\sigma}^{[k]}}{\sqrt{(n^{[k]} - 1) \times \text{se}(\hat{\beta}^{[k]})}}, \tag{8}$$

$$\hat{\mu}^{[k]} = \sqrt{\left(\frac{\text{se}(\hat{\alpha}^{[k]})}{\text{se}(\hat{\beta}^{[k]})}\right)^2 - \frac{\hat{\sigma}^{[k]2}}{n^{[k]} \times (\text{se}(\hat{\beta}^{[k]}))^2}}. \tag{9}$$

A Bayesian model for combining the evidence from several linear regressions was first proposed by Novick et al. (1972)[17] and we consider a similar framework. There are many reasons why a Bayesian model may be advantageous over a frequentist model. For instance, frequentist meta-analysis models are known to have difficulty estimating variance parameters if these parameters are near-zero, particularly when sample sizes are small[18,19]. Bayesian models also offer substantial flexibility for handling complicated data structures that may arise with multiple covariates and the possibility of measurement error[16,15]. See Kim and Peiris (2019)[20] for a review the many different frequentist meta-analytic methods for the synthesis of linear regressions.

For our unknown parameters of interest (θ, τ, ξ, ω, α, and β), and data from K independent studies (we require $\hat{\alpha}^{[k]}$, $\hat{\beta}^{[k]}$, and values for $\hat{\mu}^{[k]}$, $\hat{\sigma}^{[k]}$, and $\hat{\lambda}^{[k]}$, for k in $1, \ldots, K$), Bayes' Law takes the form:

$$\begin{aligned} p((\theta, \tau, \xi, \omega, \alpha, \beta)|\text{data}) &\propto p(\text{data}|(\theta, \tau, \xi, \omega, \alpha, \beta))p(\theta, \tau, \xi, \omega, \alpha, \beta) \\ &= \left(\prod_{k=1}^{K} p(\hat{\alpha}^{[k]}, \hat{\beta}^{[k]}|\alpha^{[k]}, \beta^{[k]})p(\alpha^{[k]}|\xi, \omega)p(\beta^{[k]}|\theta, \tau)\right) \\ &\quad \times p(\theta)p(\xi)p(\tau)p(\omega). \end{aligned} \tag{10}$$

We have that $p(\hat{\alpha}^{[k]}, \hat{\beta}^{[k]}|\alpha^{[k]}, \beta^{[k]})$ is defined according to a multivariate Normal as stated in equation (1) and that $p(\alpha^{[k]}|\xi, \omega)$ and $p(\beta^{[k]}|\theta, \tau)$ are defined according to (5) and (6) respectively. We are left to define prior distributions for the unknown parameters: θ, τ, ξ, and ω.

Our strategy will be to adopt wide Normal distributions (with variance of 100) for the mean parameters and weakly-informative half-Cauchy priors (with scale parameter of 2) for the τ and ω parameters; following the recommendations of Polson et al. (2012)[21] and the simulation results of Williams et al. (2018)[22]. Consider the following priors:

$\theta \sim \mathcal{N}(0, 100)$; (mean of 0, variance of 100)

$\xi \sim \mathcal{N}(0, 100)$;

$\tau \sim \text{half-Cauchy}(0, 2), \quad \tau > 0;$ (location of 0, scale of 2) and:

$\omega \sim \text{half-Cauchy}(0, 2), \quad \omega > 0.$

We emphasize that the performance of any Bayesian estimator will depend on the choice of priors. Particularly when few data are available, the choice of priors can substantially influence the posterior[23,24,25].

Going forward, the above Bayesian model will be referred to as *BayesMA*.

2.1.1 | Example: the NELS88 dataset

Let us now demonstrate how the *BayesMA* model can be used for data analysis with a simple example analysis of the NELS88 datset. The NELS88 dataset has been used previously as an example dataset by[26] and is from a survey of U.S. grade 10 high-school students in 1988 from over 1,000 schools.

Becker et al. (2007)[26] include for analysis only the 13 schools with samples of a minimum of 45 students ($n^{[k]} \geq 45; k = 1, \ldots, 13$) and consider each school as an independent study for meta-analysis. We will use the same subset of schools for our example analysis. The outcome of interest, $Y_j^{[k]}$, will be the science achievement test score, and the exposure of interest, $X_j^{[k]}$, will be the reading test score, for the j-th student in the k-th school. The total sample size is $N = 664$ students from $K = 13$ different schools.

Table 1 displays the aggregate data from the NELS88 dataset to be used for meta-analysis. We fit the data with the *BayesMA* model described in Section 2.1 and report posterior medians and equal-tailed 95% credible intervals. All models in this paper are fit using the probabilistic programming language JAGS which employs the Gibbs sampling Markov chain Monte Carlo (MCMC) algorithm and is compatible with the R statistical programming language[27]. Each model is fit based on 100,000 monte carlo draws from each of three chains (thinning of 10).

We obtain the following posterior medians: $\hat{\theta} = 0.57$ with 95% equal-tailed credible interval of $CI(\theta)_{95\%} = [0.50, 0.64]$; $\hat{\xi} = 5.40$ with 95% equal-tailed credible interval of $CI(\xi)_{95\%} = [4.06, 6.81]$; $\hat{\tau} = 0.04$; and $\hat{\omega} = 1.82$.

2.2 | Adjusting for non-differential measurement error

Now suppose that each study is hampered by a distinct extent of classical and non-differential measurement error. In other words, suppose that the vector of imprecise surrogate exposures X^* is recorded, instead of X itself. The assumption of *non-differential* measurement error refers to the fact that the distribution of the surrogate exposures depends only on the actual exposure variables and not on the response variable or other variables in the model. In other words, we have that the conditional distribution of $(X^*|X, Y)$ is identical to the conditional distribution of $(X^*|X)$.

We consider the relationship between the outcome, Y, and the exposure, X, having only data on Y and X^* and assume the vector of independent surrogates, X^*, arises from a classical additive measurement error model. Consider:

$$X_j^{[k]*}|X_j^{[k]} \sim X_j^{[k]} + \mathcal{N}\left(0, \phi^{[k]2}\right), \text{ for } j \text{ in } 1, \ldots, n^{[k]}, \tag{11}$$

for k in $1, \ldots, K$, where the $\phi^{2[k]}$ parameter corresponds to the variance in measurement error for the k-th study. To be clear, for each study, the average error is zero but individual measurements can be biased, such that:

$$X_j^{[k]*} \sim \mathcal{N}\left(\mu^{[k]}, \lambda^{[k]*2}\right) \tag{12}$$

where $\lambda^{[k]*2} = \lambda^{[k]2} + \phi^{[k]2}$; for j in $1, \ldots, n^{[k]}$ and for k in $1, \ldots, K$.

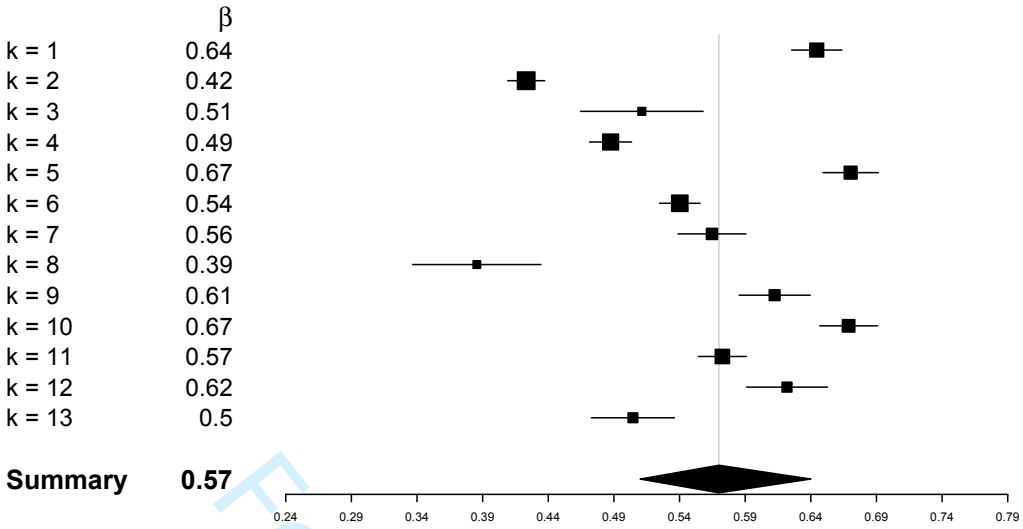


FIGURE 1 Forest plot for the meta-analysis of the NELS88 aggregate data (Table 1). The $K = 13$ black squares correspond to the $\hat{\beta}^{[k]}$ values, for k in $1, \ldots, K$; with horizontal lines corresponding to symmetrical $1.96 \times \text{se}(\hat{\beta}^{[k]})$ confidence intervals. The *BayesMA* posterior estimate of θ is plotted as a diamond, the lateral points of which indicate the equal-tailed 95% credible interval for this estimate.

Let $\gamma^{[k]} = \lambda^{[k]2}/(\lambda^{[k]2} + \phi^{[k]2}) = \left(1 + \phi^{[k]2}/\lambda^{[k]2}\right)^{-1} < 1$ be the "attenuation factor" for the k-th study, for k in $1, \ldots, K$. The range of values for $\phi^{[k]}$ is therefore restricted to: $0 \leq \phi^{[k]} \leq \lambda^{[k]*}$. The regression coefficients estimated in the individual studies will be biased and governed by:

$$\left.\begin{matrix}\hat{\alpha}^{[k]*} \\ \hat{\beta}^{[k]*}\end{matrix}\right| \begin{matrix}\alpha^{[k]} \\ \beta^{[k]}\end{matrix} \sim \mathcal{MVN}\left(\begin{pmatrix}\alpha^{[k]} + (1-\gamma^{[k]})\beta^{[k]}\mu^{[k]} \\ \gamma^{[k]}\beta^{[k]}\end{pmatrix}, \begin{pmatrix}\Sigma_{11}^{[k]*} & \Sigma_{12}^{[k]*} \\ \Sigma_{12}^{[k]*} & \Sigma_{22}^{[k]*}\end{pmatrix}\right), \tag{13}$$

$$\text{where:} \quad \Sigma_{11}^{[k]*} = \frac{(\lambda^{[k]*2} + \mu^{[k]*2}) \times \sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}};$$

$$\Sigma_{22}^{[k]*} = \frac{\sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}}; \quad \text{and}$$

$$\Sigma_{12}^{[k]*} = -\mu^{[k]*} \times \frac{\sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}};$$

where $\mu^{[k]*} = \mu^{[k]}$; $\sigma^{[k]*2} = \sigma^{[k]2} + (1-\gamma^{[k]})(\beta^{[k]2})(\lambda^{[k]2})$; and $\lambda^{[k]*2} = \lambda^{[k]2} + \phi^{[k]2}$, for k in $1, \ldots, K$.

If we ignore measurement error, then, for the parameter of interest θ, we will mistakenly target:

$$\theta^* = \text{E}\left(\gamma^{[k]}\beta^{[k]}\right) = \text{E}\left\{\left(1 + \frac{\phi^{[k]2}}{\lambda^{[k]2}}\right)^{-1} \times \beta^{[k]}\right\}, \tag{14}$$

in place of $\theta = \mathrm{E}\left(\beta^{[k]}\right)$, for all k in $1, \ldots, K$. If we presume independence of $\beta^{[k]}$ and $\left(\phi^{[k]}, \lambda^{[k]}\right)$, then we have:

$$\theta^{*} = \mathrm{E}\left(\gamma^{[k]}\right)\theta. \tag{15}$$

This is intuitive: the attenuation factor induced by measurement error in estimating the typical exposure-outcome association is the expectation of the study-specific attenuation factors. This suggests that an unbiased estimate of the overall effect, θ, can be derived analytically if one knows the cross-study average degree of measurement error. In other words, it is not necessary to know each individual value of $\phi^{[k]}$, for k in $1, \ldots, K$. One need only know where the distribution of the $\phi^{[k]}$s is centered in order to adequately adjust the meta-analytic point estimate of θ in the presence of measurement error.

The bias in estimating τ^2 is perhaps less intuitive. If we ignore the presence of measurement error, then our estimation procedures will mistakenly target: $\tau^{*2} = \mathrm{Var}\left(\beta^{[k]*}\right) = \mathrm{Var}\left(\gamma^{[k]}\beta^{[k]}\right)$, for all k in $1, \ldots, K$. Invoking independence of $\gamma^{[k]}$ and $\beta^{[k]}$, we have:

$$\tau^{*2} = \left\{\mathrm{E}\left(\gamma^{[k]}\right)\right\}^2 \tau^2 + \mathrm{Var}\left(\gamma^{[k]}\right)\left(\tau^2 + \theta^2\right), \tag{16}$$

or alternatively: $\tau^{*2} = \mathrm{E}\left(\gamma^{[k]2}\right)\tau^2 + \mathrm{Var}\left(\gamma^{[k]}\right)\theta^2$. The first term alone speaks to underestimating cross-study heterogeneity if we ignore measurement error, since we know $0 < \mathrm{E}\left(\gamma^{[k]2}\right) < 1$, for all k in $1, \ldots, K$. However, the second term could counteract this. More precisely, if the magnitude of the cross-study variation in measurement error (i.e., if $\mathrm{Var}(\gamma^{[k]})$ is large) and/or the average effect size (i.e., θ is large) are large enough, we could end up overestimating cross-study heterogeneity instead.

It follows that, without any cross-study heterogeneity, i.e., when $\tau^2 = 0$, equation (16) reduces to $\tau^{*2} = \mathrm{Var}\left(\gamma^{[k]}\right)\theta^2$. Therefore, if the cross-study variability in measurement error is sufficiently large, one could erroneously select a random-effects model ($\tau^2 > 0$) instead of the correct fixed-effects model ($\tau^2 = 0$).

The bias in estimating ξ and ω^2 must also be considered. If we ignore measurement error, then our estimation procedures will mistakenly target: $\xi^{*} = \mathrm{E}\{\alpha^{[k]} + (1-\gamma^{[k]})\beta^{[k]}\mu^{[k]}\} = \xi + \mathrm{E}\{(1-\gamma^{[k]})\beta^{[k]}\mu^{[k]}\}$, for k in $1, \ldots, K$. For the ω parameter, ignoring measurement error will lead one to mistakenly target: $\omega^{*2} = \mathrm{Var}\left(\alpha^{[k]*}\right) = \mathrm{Var}\left(\alpha^{[k]} + (1-\gamma^{[k]})\beta^{[k]}\mu^{[k]}\right)$, for k in $1, \ldots, K$.

The correlation between different elements of the data will also be impacted by measurement error. For example, in the absence of measurement error (i.e., when $\gamma^{[k]} = 1$), the estimators, $\hat{\beta}^{*[k]}$ and $\hat{\lambda}^{*[k]}$ will be entirely independent. However, in the presence of heterogeneous measurement error, this is no longer the case. A large amount of measurement error will lead to a larger value of $\hat{\lambda}^{*[k]}$ and simultaneously to a smaller value of $\hat{\beta}^{*[k]}$. As such, if a study has a relatively (as compared to the other studies) small value of $\hat{\beta}^{*[k]}$ and a relatively large value of $\hat{\lambda}^{*[k]}$, this suggests that (relative to other studies) it may be compromised by a substantial degree of measurement error.

2.3 | Issues of identifiability

Note that, without any specific knowledge regarding the degree of measurement error for each of the contributing studies, the meta-analysis parameters of interest may not be identifiable. With this in mind, let us briefly consider an asymptotic argument for so-called "partial identifiability" by considering the degree to which θ can be estimated in the presence of unspecified measurement error.

Presume that *a priori* defensible information about the amount of bias caused by measurement error in the k-th study is expressed in the form: $\gamma^{[k]} \in [\underline{\gamma}^{[k]}, 1]$, where $\underline{\gamma}^{[k]}$ is an investigator-specified lower bound for the k-th attenuation factor. Then the set of possible values for $\beta^{[k]}$, given $\beta^{[k]*}$, is restricted to:

$$I_k(\underline{\gamma}^{[k]}) = \left[\frac{\beta^{*[k]}}{\underline{\gamma}^{[k]}}, \beta^{*[k]}\right]. \tag{17}$$

To be clear, this represents the study-specific identification interval for $\beta^{[k]}$. As $n^{[k]} \to \infty$, all values inside the interval remain plausible, while all values outside are ruled out[28]. This is the essence of the *partial identification* inherent to this problem.

Thinking now about the meta-analytic task of combining information, the $\gamma^{[k]}$ could exhibit considerable variation across studies while τ (i.e., the variation in $\beta^{[k]}$) could be small. Suppose that τ does not exceed an investigator-specified upper bound of $\bar{\tau}$, i.e., $\tau \leq \bar{\tau}$. Then an identification region for θ can be specified as:

$$I(\underline{\gamma}, \bar{\tau}) = \left\{\theta : \tau \leq \bar{\tau}, \beta^{[k]} \in I_k(\underline{\gamma}^{[k]}), \forall k \in \{1, \ldots, K\}\right\}. \tag{18}$$

Again, the interpretation is direct: in the asymptotic limit, all values of θ inside this interval are compatible with the observed data, and all values outside are not. The primary question of interest is whether this interval is narrow or wide under realistic scenarios, since this governs the extent to which we can learn about θ from the data.

In general, evaluating (18) for given inputs is an exercise in quadratic programming nested within a grid search, hence can be handled with standard numerical optimisation. However, the special "fixed-effects" case of $\bar{\tau} = 0$ is noteworthy. Mathematically, the case is much simpler, with (18) reducing to $I(\underline{\gamma}, 0) = \cap_k I_k(\underline{\gamma}^{[k]})$. As intuition must have it, without heterogeneity, a putative value for θ is compatible with the observed data if and only if it is compatible with the data from *every* individual study.

To illustrate, consider a scenario with $K = 10$ studies, with $\theta = 1.0$, and $\tau = 0$ i.e., $\beta^{[k]} = 1.0$, for k in $1, \ldots, 10$. Suppose the observed $\hat{\beta}^{[k]*}$ values for these studies lie equally spaced between 0.40 and 0.80, (since the unknown $\gamma^{[k]}k$ values range between 0.40 and 0.80). Furthermore, say the investigator pre-specifies $\underline{\gamma}^{[k]} = 0.2$ for all k. The resulting study-specific identification intervals, I_k, are depicted in the upper panel of Figure 2. Also depicted by the green rectangle is the global identification interval, i.e., the intersection of the individual intervals. The global identification interval is indeed narrow, ranging from 0.8 to 2.0.

Evaluating (18) for $\overline{\tau} > 0$ can be done via quadratic programming; see Appendix for details. Figure 2 (lower panel) shows that, when $\overline{\tau} = 0.5$, the global identification interval is wider: 0.6 to 2.9. In summary, depending on the the upper limit in τ and the range in $\gamma^{[k]}$ values, i.e., the "heterogeneity of bias," it appears that data can indeed contribute substantial information about θ.

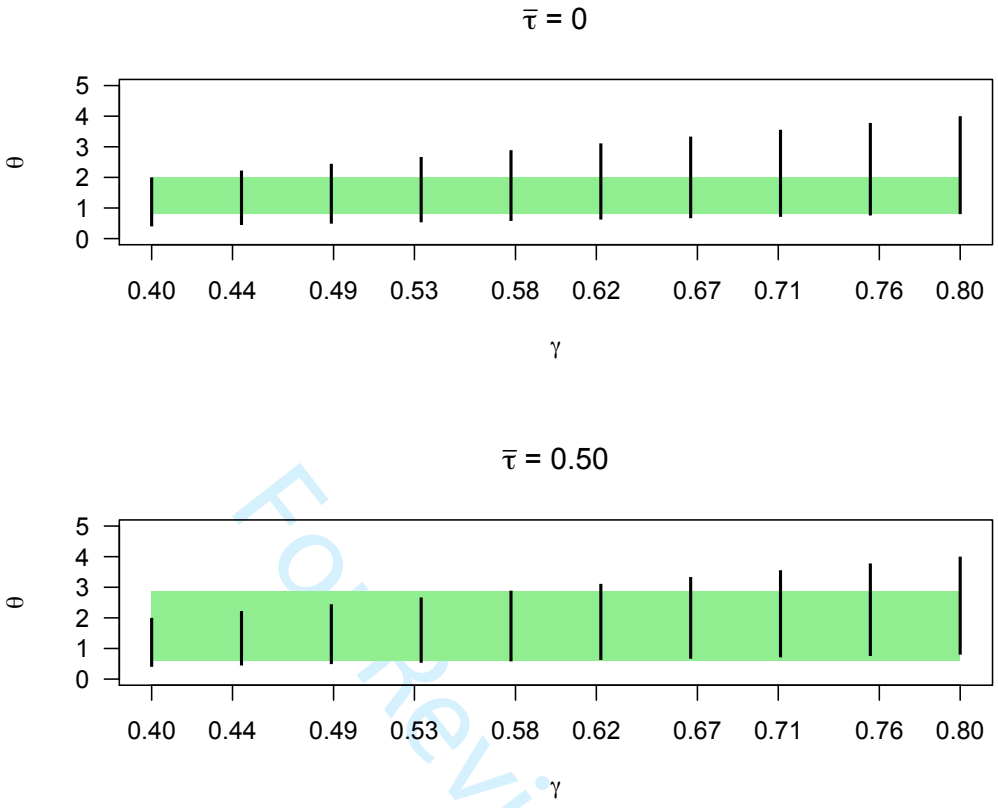


FIGURE 2 Black vertical lines correspond to jurisdiction-specific identification intervals and the green rectangle corresponds to the global identification interval. Upper panel corresponds to assumption of $\bar{\tau} = 0$ such that the global identification interval is simply the intersection of the individual intervals. Lower panel corresponds to $\bar{\tau} = 0.50$.

2.4 | The BMEMA model

The traditional random-effects model (the *BayesMA* model) described in Section 2.1, can be easily adapted to account for non-differential measurement error. Let the *BMEMA* model be defined by a joint structure such that, for j in $1, \ldots, n^{[k]}$, and for k in $1, \ldots, K$:

$$X_j^{[k]} \sim \mathcal{N}\left(\mu^{[k]}, \lambda^{[k]2}\right);$$
$$X_j^{[k]*} | X_j^{[k]} \sim \mathcal{N}\left(X_j^{[k]}, \phi^{[k]2}\right); \text{ and}$$
$$Y_j^{[k]} | X_j^{[k]} \sim \mathcal{N}\left(\alpha^{[k]} + \beta^{[k]} X_j^{[k]}, \sigma^{[k]2}\right).$$

For the unknown parameters of interest (θ, τ, ξ, ω, α, and β), and measurement error-tainted aggregate data ($data^*$) from the K studies (we require: $\hat{\alpha}^{[k]*}$, $\hat{\beta}^{[k]*}$, for k in $1, \ldots, K$ and values for $\mu^{[k]*}$, $\sigma^{[k]*}$, and $\lambda^{[k]*}$), Bayes theorem states that:

$$\begin{aligned} p((\theta, \tau, \xi, \omega, \alpha, \beta)|data^*) &\propto p(data^*|(\theta, \tau, \xi, \omega, \alpha, \beta))p(\theta, \tau, \xi, \omega, \alpha, \beta) \\ &= \Big(\prod_{k=1}^{K} p(\hat{\alpha}^{[k]*}, \hat{\beta}^{[k]*}|\alpha^{[k]}, \beta^{[k]}, \phi^{[k]})p(\alpha^{[k]}|\xi, \omega)p(\beta^{[k]}|\theta, \tau) \Big) p(\theta)p(\xi)p(\tau)p(\omega)p(\phi), \end{aligned} \tag{19}$$

where $p(\hat{\alpha}^{[k]*}, \hat{\beta}^{[k]*}|\alpha^{[k]}, \beta^{[k]}, \phi^{[k]})$ is defined according to a multivariate Normal as stated in equation (13); and $p(\alpha^{[k]}|\xi, \omega)$ and $p(\beta^{[k]}|\theta, \tau)$ are defined according to equations (5) and (6) respectively. For parameters θ, τ, ξ, and ω, we define prior distributions as in Section 2.1.

Our strategy for modelling the degree of measurement error is to place an inverse-gamma prior on the study-specific $\phi^{[k]2}$ parameters, such that:

$$\phi^{[k]2} \sim \text{Inv-Gamma}(\zeta_1, \zeta_2); \qquad \text{and} \quad \zeta_1 \sim \text{Exp}(\rho), \quad \zeta_2 \sim \text{Exp}(\rho), \tag{20}$$

for k in $1, \ldots, K$; and for $\zeta_1 > 0$ and $\zeta_2 > 0$. Note that the mean and variance of the inverse-gamma distribution have that, *a priori*, $\text{E}(\phi^{[k]2}) = \zeta_2/(\zeta_1 - 1)$ and $\text{Var}(\phi^{[k]2}) = \zeta_2^2/((\zeta_1 - 1)^2(\zeta_1 - 2))$. However, note that only values of $\phi^{[k]2} \leq \lambda^{[k]*}$ will be consistent with the data. In order to reflect very vague prior knowledge, we set $\rho = 0.1$.

Knowledge about the magnitude of measurement error in a study may come from a variety of sources, e.g. replicate measurements, validation data, or expert opinion. In some scenarios, one might have a subset of studies for which $\phi^{2[k]}$ is known and equal to 0 (i.e., have data from some studies known to be unaffected by measurement error). We will focus on how one might proceed under such a scenario. Without loss of generality, suppose a subset of "gold standard" studies is the first k' studies, such that for $k = 1, \ldots, k'$, we have $\phi^{[k]2} = 0$ and $\gamma^{[k]} = 1$. Thus, in a situation where all studies are known to be unbiased by measurement error, k' will equal K and the *BayesMA* and *BMEMA* models will be identical.

2.4.1 | Example: the NELS88* dataset

Returning to our example with the NELS88 dataset, we will illustrate the impact of measurement error by intentionally adding non-differential measurement error as described in equation (11) so as to corrupt the reading test scores for 8 out of the $K = 13$ schools. As such, the contaminated dataset, NELS88*, has $k' = 5$ schools for which the data is "clean." We set $\phi^{[k]} = 0$, for $k = 1, \ldots, 5$; and, for $k = 6, \ldots, 13$, increasing values from 1 to 12: $\phi^{[6]} = 1.00$, $\phi^{[7]} = 2.57, \ldots, \phi^{[12]} = 10.43$ and $\phi^{[13]} = 12.00$. Table 2 lists the aggregate data obtained after adding measurement error to the reading scores. We also list the values of $\phi^{[k]}$ and $\gamma^{[k]}$ for reference.

With the NELS88* dataset, we have that $\text{E}(\gamma^{[k]}) = 0.63$ and $\text{Var}(\gamma^{[k]}) = 0.15$. Based on equation (15), we have that $\theta^* = 0.66 \times 0.57 = 0.36$; and based on equation (16), we have that $\tau^{*2} = 0.54^2 \times 0.04^2 + 0.15 \times \left(0.04^2 + 0.57^2\right) = 0.05$, or $\tau^* = 0.22$. Indeed, if we ignore the possibility of any measurement error, we obtain, with the NELS88* dataset, estimates similar to the numbers given by equations (15) and (16) (see Table 3 line 1): when measurement error is added to the data, estimates for θ are

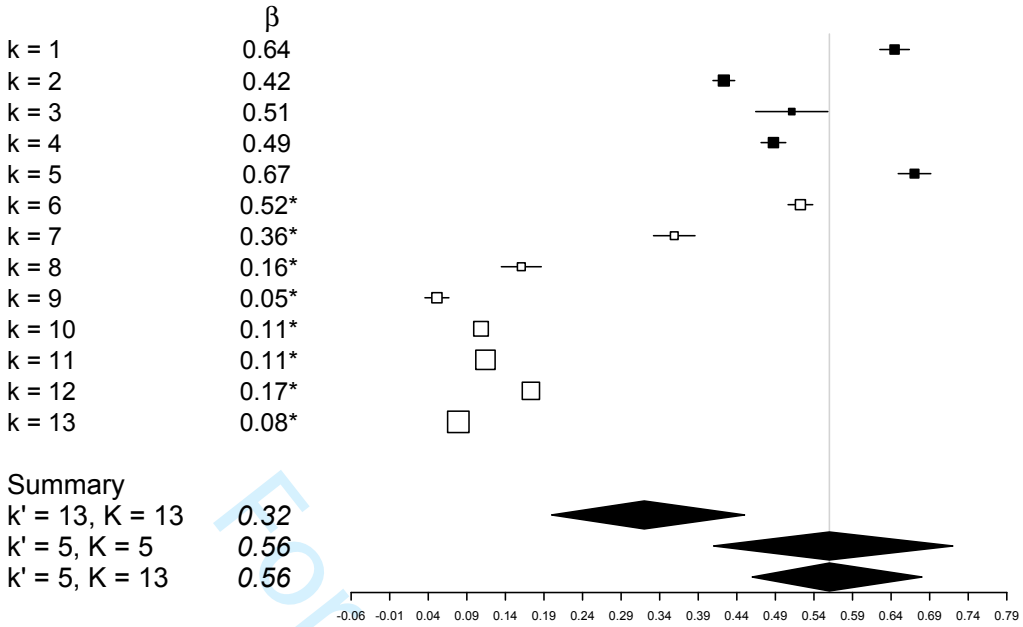


FIGURE 3 Forest plot for the meta-analysis of the NELS88* aggregate data, with three summaries corresponding to line 1 ($k' = 13$, $K = 13$), line 4 ($k' = 5$, $K = 5$), and line 3 ($k' = 5$, $K = 13$) of Table 3). The 5 black full squares correspond to the $\hat{\beta}^{[k]}$ values unaffected by measurement error; the 8 empty squares correspond to the $\hat{\beta}^{[k]}$ values potentially compromised by measurement error; with horizontal lines corresponding to symmetrical $1.96 \times \text{se}(\hat{\beta}^{[k]})$ confidence intervals. The *BMEMA* posterior estimates of θ are plotted as diamonds, the lateral points of which indicate the equal-tailed 95% credible intervals for these estimates.

biased downwards, while estimates for τ are biased upwards. Also, note that the estimates for both ξ and ω are biased upwards (compare Table 3 line 1 to line 2).

For the NELS88* data, $\text{cor}(\hat{\beta}^{*[k]}, \hat{\lambda}^{*[k]}) = -0.66$, whereas for the NELS88 data, $\text{cor}(\hat{\beta}^{[k]}, \hat{\lambda}^{[k]}) = -0.02$. The fact that the values of $\hat{\beta}^{*[k]}$ and $\hat{\lambda}^{*[k]}$ are negatively correlated in the presence of heterogeneous measurement error and independent otherwise suggests that by combining data from several studies, the presence of bias caused by measurement error can be better identified (and suggests the possibility of a simple diagnostic test for heterogeneous measurement error).

The estimate of $\hat{\theta} = 0.57$ obtained with the *BayesMA* model and the NELS88 dataset (see Table 3 line 2) serves as a target. For the NELS88* dataset, the *BMEMA* model, with $k' = 5$, obtains an estimate of $\hat{\theta} = 0.56$ (see Table 3 line 3). MCMC diagnostic plots are presented in the Appendix and show little prior-posterior overlap (PPO) which suggests that the prior is suitably overwhelmed by the signal provided in the data; see Figures 4 and 5.

We also fit the Bayesian model to data from only the first five schools that are known to be unaffected by measurement error (see Table 3 line 4) and obtain $\hat{\theta} = 0.56$ with a notably wider credible interval of $CI(\theta)_{95\%} = [0.41, 0.72]$. This suggests that

the data from the additional eight schools, while compromised by measurement error, provide "added value" and sharpen our inference. However, we also note that knowing that the $k = 1, \ldots, 5$ schools are untainted by measurement error is crucial to obtaining appropriate estimates: the *BMEMA* model, with $k' = 0$, obtains an estimate of θ much too high: $\hat{\theta} = 0.80$ (see Table 3 line 5).

Finally, note that if the *BMEMA* model is fit to the original NELS88 data, we will end up overestimating the θ parameter: $\hat{\theta} = 0.59$, with $k' = 5$, and $\hat{\theta} = 0.60$, with $k' = 0$ (see Table 3 lines 6 and 7).

3 | META-ANALYSIS OF MULTIVARIABLE LINEAR REGRESSION

3.1 | In the absence of measurement error

When studies are not measured with error, we can generalize the *BayesMA* model described in Section 2 (in which each study is summarized as a simple linear regression) to a general case where each study is summarized as a multivariable linear regression. We denote $X^{[k]}_{j\times}$ as the $(Q+1)$-length row vector with elements 1, and the Q covariates measured for the j-th observation in the kth observational study. Similarly, we denote $X^{[k]}_{\times q}$ as the $n^{[k]}$-length column vector of values of the q-th covariate for the k-th study. Finally the $n^{[k]} \times (Q+1)$ design matrix, $X^{[k]}$, consists of $n^{[k]}$ rows, one for each observation in the k-th study: $X^{[k]}_{j\times}$, for j in $1, \ldots, n^{[k]}$. Note that, in this multivariate setting, $X^{[k]}$ is a matrix with the first column consisting of 1s. We will use $X^{[k]}_{\times,-1}$ to denote the design matrix excluding the column of 1s.

Each study can be summarized as a multivariable linear regression model where, for k in $1, \ldots, K$, and j in $1, \ldots, n^{[k]}$, we have:

$$Y^{[k]}_j | X^{[k]}_{j\times} \sim \mathcal{N}\left(X^{[k]}_{j\times}\beta^{[k]}, \sigma^{[k]2}\right), \tag{21}$$

where $\beta^{[k]} = (\beta^{[k]}_0, \beta^{[k]}_1, \ldots, \beta^{[k]}_Q)'$ is the column-vector of regression coefficients. We assume that each study has its own exposure distribution governed by a multivariate Normal distribution:

$$X^{[k]}_{j,-1} \sim \mathcal{MVN}\left(\mu^{[k]}, \Lambda^{[k]}\right), \tag{22}$$

where $\mu^{[k]}$ is a Q-length vector, and $\Lambda^{[k]}$ is a $Q \times Q$ covariance matrix. The random-effects meta-analysis model can be summarized as:

$$\beta^{[k]} | \theta, \mathrm{T} \sim \mathcal{MVN}(\theta, \mathrm{T}), \tag{23}$$

for all k in $1, \ldots, K$; where θ is a $(Q+1)$-length vector, and T is a $(Q+1)\times(Q+1)$ diagonal covariance matrix. We first consider a situation in which individual participant data (IPD) are available. For our unknown parameters of interest (θ, T, β, σ, μ, Λ, and

T), and the data from K studies (we require: $Y_j^{[k]}$, and $X_{j\times}^{[k]}$ for j in $1, \ldots, n^{[k]}$, and for k in $1, \ldots, K$), Bayes' theorem states that:

$$\begin{aligned} p(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda|\text{data}) &\propto p(\text{data}|(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda))p(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda) \\ &= \prod_{k=1}^{K}\Big(\prod_{j=1}^{n^{[k]}}\Big\{p(Y_j^{[k]}|X_{j\times}^{[k]}, \beta^{[k]}, \sigma^{[k]})p(X_{j\times}^{[k]}|\mu^{[k]}, \Lambda^{[k]})\Big\} \\ &\quad \times p(\beta^{[k]}|\theta, \mathrm{T})p(\mu^{[k]})p(\Lambda^{[k]})p(\sigma^{[k]})\Big)p(\theta)p(\mathrm{T}). \end{aligned} \tag{24}$$

We have that $p(Y_j^{[k]}|X_{j\times}^{[k]}, \beta^{[k]}, \sigma^{[k]})$ is defined according to (21), that $p(X_{j\times}^{[k]}|\mu^{[k]}, \Lambda^{[k]})$ is defined according to (22), and that $p(\beta^{[k]}|\theta, \mathrm{T})$ is defined by (23). We are left to define prior distributions for the parameters $\mu^{[k]}$, $\Lambda^{[k]}$, $\sigma^{[k]}$, θ, and the $Q+1$ diagonal elements of the T covariance matrix. We specify multivariate normal priors for $\mu^{[k]}$ and θ:

$\mu^{[k]} \sim \mathcal{MVN}(0_{[Q]}, 100 \times I_{[Q]})$, for k in $1, \ldots, K$; and

$\theta \sim \mathcal{MVN}(0_{[Q+1]}, 100 \times I_{[Q+1]})$;

where $0_{[Q]}$ is a Q-length vector of zeros, and $I_{[Q]}$ is a $Q \times Q$ identity matrix. We specify a half-Cauchy priors for:

$\sqrt{\mathrm{T}_{qq}} \sim$ half-Cauchy$(0, 2)$, for q in $1, \ldots, Q+1$; and

$\sigma^{[k]} \sim$ half-Cauchy$(0, 2)$, for k in $1, \ldots, K$,

such that, when $Q = 1$, these priors are equivalent to those described for the univariate model in Section 2. Finally, for $\Lambda^{[k]}$ we specify an inverse-Wishart prior to reflect vague prior knowledge such that:

$\Lambda^{[k]} \sim$ Inv-Wishart$(I_{[Q]}, Q)$, for k in $1, \ldots, K$.

Note that, if IPD are not available and only aggregate data are available, an alternative model can be defined. In this situation, the $\hat{\beta}^{[k]}$ regression coefficients are governed by:

$$\hat{\beta}^{[k]}|\beta^{[k]} \sim \mathcal{MVN}(\beta^{[k]}, COV^{[k]}), \text{ where } COV^{[k]} = (X^{[k]T}X^{[k]})^{-1}\sigma^{[k]2}, \tag{25}$$

for k in $1, \ldots, K$. Assuming that $COV^{[k]}$ is available and known for $k = 1, \ldots, K$ (this is analogous to the assumption in Section 2 that, for k in $1, \ldots, K$, $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$ are known), Bayes theorem states that:

$$\begin{aligned} p(\theta, \mathrm{T}, \beta|\text{data}) &\propto p(\text{data}|(\theta, \mathrm{T}, \beta))p(\theta, \mathrm{T}, \beta) \\ &= \prod_{k=1}^{K} p(\hat{\beta}^{[k]}|\beta^{[k]})p(\beta^{[k]}|\theta, \mathrm{T})p(\theta)p(\mathrm{T}), \end{aligned} \tag{26}$$

where $p(\hat{\beta}^{[k]}|\beta^{[k]})$ is defined according to (25).

3.2 | Adjusting for non-differential measurement error

Suppose observed covariates are measured with non-differential error such that:

$$X_{j,-1}^{*[k]} = X_{j,-1}^{[k]} + \mathcal{MVN}(0, \Phi^{[k]}), \tag{27}$$

for j in $1, \ldots, n^{[k]}$ and for k in $1, \ldots, K$; where $\Phi^{[k]}$ is a $Q \times Q$ covariance matrix. Note that, equation (27) is analogous to equation (11) in Section 2.2. Also note that, if $\Phi^{[k]}$ is a diagonal matrix, then the measurement error in any one covariate is entirely independent of the measurement error in every other covariate.

Multivariate measurement error can bias estimators in unpredictable and unexpected ways. For instance, Abel (2017)[29] shows that, even if the q-th covariate, $X_{\times q}$, is measured without error, one may still incorrectly reject the null hypothesis of $\beta_q^{[k]} = 0$, if $X_{\times q}$ is correlated with another covariate that is itself tainted by measurement error. Since covariates may be correlated to one another in many different ways, it is difficult to anticipate the impact of multivariate measurement error for general multivariate settings[30].

The multivariate *BayesMA* model outlined in Section 3.1 can be adapted to account for measurement error in one or several of the covariates. If IPD are available[31], we can frame a flexible multivariate *BMEMA* model as follows.

We begin by assuming that the covariates are multivariate normal (but this could be modified as needed) and consider the following three-part model structure:

$$X_{j,-1}^{[k]} \sim \mathcal{MVN}\left(\mu^{[k]}, \Lambda^{[k]}\right), \tag{28}$$

$$X_{j,-1}^{[k]*} | X_{j,-1}^{[k]} \sim \mathcal{MVN}\left(X_{j,-1}^{[k]}, \Phi^{[k]}\right), \quad \text{and} \tag{29}$$

$$Y_j^{[k]} | X_{j\times}^{[k]} \sim \mathcal{N}\left(X_{j\times}^{[k]}\beta^{[k]}, \sigma^{[k]2}\right), \tag{30}$$

for j in $1, \ldots, n^{[k]}$ and k in $1, \ldots, K$.

For the unknown parameters of interest $(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda, \Phi)$, and the measurement error -tainted data ($data^*$) from the K studies (we require: $(Y_j^{[k]}, X_{\times,j}^{[k]*})$, for j in $1, \ldots, n^{[k]}$ and k in $1, \ldots, K$), Bayes theorem states that:

$$\begin{aligned} p((\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda, \Phi)|data^*) &\propto p(data^*|(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda, \Phi)p(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda, \Phi) \\ &= \prod_{k=1}^{K}\Bigg(\prod_{j=1}^{n^{[k]}}\left\{p(Y_j^{[k]}|X_{j\times}^{[k]}, \beta^{[k]}, \sigma^{[k]})p(X_{j\times}^{[k]*}|X_{j\times}^{[k]}, \Phi^{[k]})p(X_{j\times}^{[k]}|\mu^{[k]}, \Lambda^{[k]})\right\} \\ &\quad \times p(\beta^{[k]}|\theta, \mathrm{T})p(\sigma^{[k]})p(\mu^{[k]})p(\Lambda^{[k]})\Bigg) \\ &\quad \times p(\theta)p(\mathrm{T})p(\Phi), \end{aligned} \tag{31}$$

where $p(Y_j^{[k]}|X_{j\times}^{[k]}, \beta^{[k]}, \sigma^{[k]})$ is defined according to (30), $p(X_{j\times}^{[k]*}|X_{j\times}^{[k]}, \Phi^{[k]})$ is defined according to (29), and $p(X_{j\times}^{[k]}|\mu^{[k]}, \Lambda^{[k]})$ is defined according to (28).

We specify priors for θ, T, $\sigma^{[k]}$, $\mu^{[k]}$ and $\Lambda^{[k]}$ as in Section 3.1. For Φ, we select the inverse-Wishart distribution with covariance matrix $2\zeta_2 \times I_{[Q]}$ and $2\zeta_1$ degrees of freedom (the multivariate generalization of the inverse-gamma distribution in (20)) to reflect vague prior knowledge such that:

$$\Phi^{[k]} \sim \text{Inv-Wishart}(2\zeta_2 \times I_{[Q]}, 2\zeta_1); \qquad \zeta_1 \sim \text{Exp}(\rho) \quad \text{and:} \quad \zeta_2 \sim \text{Exp}(\rho), \tag{32}$$

for k in $1, \ldots, K$; where $\zeta_1 > (Q/2)$ and $\zeta_2 > 0$. When $Q = 1$, this prior is equivalent to the univariate case described in Section 2.2. Indeed, if $Q = 1$, we have that, *a priori*, $\mathrm{E}(\Phi^{[k]}) = 2\zeta_2/(2\zeta_1 - Q - 1) = \zeta_2/(\zeta_1 - 1)$; and $\mathrm{Var}(\Phi^{[k]}) = \zeta_2^2/((\zeta_1 - 1)^2(\zeta_1 - 2))$. In order to reflect vague prior knowledge, we set $\rho = 0.1$ as in the univariate model.

As in Section 2.2, we consider a scenario in which a subset of studies are known to be unaffected by measurement error. Without loss of generality, suppose this subset of "gold standard" studies is the first k' studies, such that for $k = 1, \ldots, k'$, we have $\Phi^{[k]} = 0$, or equivalently we have:

$$X_{j,-1}^{[k]*} | X_{j,-1}^{[k]} = X_{j,-1}^{[k]},$$

for j in $1, \ldots, n^{[k]}$; for k in $1, \ldots, k'$. Thus, in a situation where all studies are known to be unbiased by measurement error, k' will equal K and the *BayesMA* and *BMEMA* models will be identical.

3.3 | Example: the NELS88 dataset

We return to the NELS88 dataset example and consider $Q = 2$ covariates. Let X_1 be the reading test score and X_2 be the mathematics test score. In order to illustrate the impact of measurement error, we will create the NELS88* dataset by adding non-differential measurement error as described in equation (27) so as to corrupt both the reading test scores and the mathematics test score for 8 out of the $K = 13$ schools. As such, as in Section 2.4.1, we have $k' = 5$. We define $\Phi^{[k]}$ to be a diagonal matrix such that the measurement error in X_1 is independent of the measurement error in X_2. For $k = 6, \ldots, 13$, we set $\Phi_{1,1}^{[k]}$ equal to between 4 and 6; and set $\Phi_{2,2}^{[k]}$ equal to between 8 and 12. Table 4 - A lists regression coefficients obtained before and after adding the measurement error and also lists the set values for $\Phi_{1,1}^{[k]}$ and $\Phi_{2,2}^{[k]}$ for reference.

Table 4 - B lists parameter estimates obtained from the multivariate *BMEMA* model. The measurement error introduced to the data biases the estimate of θ_2 towards 0. With the unbiased data, and $k' = 13$, we obtain $\theta_2 = 0.34$ (see Table 4-B, line 2). In contrast, with the biased data, NELS88*, we obtain $\theta_2 = 0.26$ (see Table 4-B, line 1). The *BMEMA* model obtains a an estimate of $\theta_2 = 0.34$ (see Table 4-B, line 3, and MCMC diagnostic plots in Figure 6 in the Appendix). When $k' = 0$, the MCMC mixing is problematic; this is clear in the MCMC diagnostic plots; see Figure 7 in the Appendix. The challenging sampling is no doubt due to to the identifiability issues discussed in Section 2.3 and to the fact that different combinations of $\Phi^{[k]}$, $\theta^{[k]}$ and $\mathrm{T}^{[k]}$ can yield similar model probabilities. Unless custom samplers are configured, inference from the *BMEMA* model with $k' = 0$ may not be possible. This should not be so surprising: computation with "partially identified" models can be a "bottleneck issue" (see Section 7.1 of Gustafson (2015)[32]).

4 | CONCLUSION

A meta-analysis based on all available evidence, even if some evidence is less than perfect, may be preferable to a meta-analysis that ignores large swaths of data[33,34].

In the simplest univariate scenario, if the exposure of interest in a study is compromised due to non-differential measurement error, one must inflate its point estimates and down-weight the study's overall contribution to the meta-analysis. The proposed Bayesian model, the *BMEMA* model, provides a systematic and efficient way to do just this. If one suspects that certain studies are compromised by measurement error, one can incorporate this uncertainty regarding the bias into the hierarchical Bayesian framework and obtain appropriate point and interval estimates. Moreover, as we demonstrated with the NELS88 analysis example, incorporating these biased studies can be beneficial: credible intervals were narrower when data from all studies were included for meta-analysis relative to when only the unbiased studies were included. This is relevant in real-world settings where meta-analyses pool evidence from varied sources. For example, in epidemiology, observational studies are frequently combined with randomized control trials (RCTs) in systematic reviews and meta-analyses; see citebun2020meta. While certain studies may be biased, they may still provide value if one can appropriately account and adjust for the bias.

We also showed that, if IPD are available, a Bayesian meta-analysis model can easily adjust for multiple participant-level covariates that are measured with or without measurement error[35]. While issues of identifiability may make a Bayesian model difficult to fit for more complex multivariable data, so long as a subset of studies is known to be unbiased, these studies can "anchor" the uncertainty allowing for straightforward Bayesian inference.

On a final note, beyond the bias caused by measurement error, a meta-analysis of observational studies should, ideally, also take into account other potential biases, e.g.: publication bias and bias due to unmeasured confounding[36,37]). The solutions we put forward may be more broadly applicable and it would seem desirable, and feasible, to consider all sources of uncertainty and bias within a single comprehensive Bayesian model. Future work should investigate whether the Bayesian hierarchical framework proposed and the "heterogeneity of bias" principle can be used to derive appropriate estimates in a meta-analysis where individual studies are subject to varying degrees –and varying types– of bias.

HIGHLIGHTS

What is already known:

- It is important to adjust for known sources of measurement error when conducting a meta-analysis. If the exposure of interest in a study is compromised due to non-differential measurement error, one must simply inflate its point estimates and down-weight the study's overall contribution to the meta-analysis.

What is new:

- The proposed Bayesian model, the *BMEMA* model, provides a systematic and efficient way to conduct a meta-analysis of measurement error - tainted continuous outcome data. If individual participant data (IPD) are available, a Bayesian

meta-analysis model can adjust for multiple participant-level covariates that are measured with or without measurement error.

Potential impact for RSM readers outside the authors' field:

- Meta-analyses based on all available evidence, even if some evidence is less than perfect, may be preferable to meta-analyses that ignore large swaths of data.

DATA AVAILABILITY STATEMENT

Code to replicate all analysis in this paper is available in two R files at:

https://github.com/harlanhappydog/MEMA:

univMEMAjags.R replicates all results in Table 3.

multiMEMAjags.R replicates all results in Table 4.

References

1. Debray TP, Damen JA, Snell KI, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017; 356: i6460.

2. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine* 2008; 149(12): 889–897.

3. Wells GA, Shea B, Higgins JP, Sterne J, Tugwell P, Reeves BC. Checklists of methodological issues for review authors to consider when including non-randomized studies in systematic reviews. *Research Synthesis Methods* 2013; 4(1): 63–77.

4. Hammer GP, Prel dJB, Blettner M. Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications. *Deutsches Ärzteblatt International* 2009; 106(41): 664.

5. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *The American Journal of Clinical Nutrition* 1997; 65(4): 1179S–1186S.

6. Fosgate G. Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null. *Emerging Themes in Epidemiology* 2006; 3(1): 7.

7. Brakenhoff TB, Mitroiu M, Keogh RH, Moons KG, Groenwold RH, Smeden vM. Measurement error is often neglected in medical literature: a systematic review. *Journal of Clinical Epidemiology* 2018; 98: 89–97.

8. Thompson S, Ekelund U, Jebb S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *International Journal of Epidemiology* 2010; 40(3): 765–777.

9. Hunter JE, Schmidt FL. *Methods of meta-analysis: Correcting error and bias in research findings*. Sage . 2004.

10. Wiernik BM, Dahlke JA. Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science* 2020; 3(1): 94–123.

11. Carroll RJ, Stefanski LA. Meta-analysis, measurement error and corrections for attenuation. *North Carolina State University. Dept. of Statistics* 1991; -(-).

12. Lian Q, Hodges JS, MacLehose R, Chu H. A Bayesian approach for correcting exposure misclassification in meta-analysis. *Statistics in Medicine* 2019; 38(1): 115–130.

13. Zeisser C, Stockwell TR, Chikritzhs T. Methodological biases in estimating the relationship between alcohol consumption and breast cancer: the role of drinker misclassification errors in meta-analytic results. *Alcoholism: Clinical and Experimental Research* 2014; 38(8): 2297–2306.

14. Brenner H, Loomis D. Varied forms of bias due to nondifferential error in measuring exposure. *Epidemiology* 1994: 510–517.

15. Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. *Statistical Methods in Medical Research* 2018; 27(6): 1695–1708.

16. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; 10(4): 277–303.

17. Novick MR, Jackson PH, Thayer DT, Cole NS. Estimating multiple regressions in m groups: A cross-validation study. *British Journal of Mathematical and Statistical Psychology* 1972; 25(1): 33–50.

18. McNeish D. On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal* 2016; 23(5): 750–773.

19. Chung Y, Rabe-Hesketh S, Choi IH. Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine* 2013; 32(23): 4071–4089.

20. Kim S, Peiris TB. Meta analysis of regression: a review and new approach with application to linear-circular regression model. *Communications in Statistics-Theory and Methods* 2019: 1–9.

21. Polson NG, Scott JG, others . On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis* 2012; 7(4): 887–902.

22. Williams DR, Rast P, Bürkner PC. Bayesian meta-analysis with weakly informative prior distributions. *PsyArXiv pre-print* 2018. doi: doi:10.31234/osf.io/7tbrm

23. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 2005; 24(15): 2401–2428.

24. Berger JO. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media . 2013.

25. Burke DL, Bujkiewicz S, Riley RD. Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. *Statistical Methods in Medical Research* 2018; 27(2): 428–450.

26. Becker BJ, Wu MJ, others . The synthesis of regression slopes in meta-analysis. *Statistical Science* 2007; 22(3): 414–429.

27. Kruschke J. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press . 2014.

28. Manski CF. *Partial identification of probability distributions*. Springer Science & Business Media . 2003.

29. Abel AB. Classical measurement error with several regressors. *Wharton School of the University of Pennsylvania* 2017.

30. Levi MD. Errors in the variables bias in the presence of correctly measured variables. *Econometrica (pre-1986)* 1973; 41(5): 985.

31. Debray TP, Moons KG, Valkenhoef vG, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research Synthesis Methods* 2015; 6(4): 293–309.

32. Gustafson P. *Bayesian inference for partially identified models: Exploring the limits of limited data*. 140. CRC Press . 2015.

33. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009; 172(1): 21–47.

34. Stone J, Glass K, Munn Z, Tugwell P, Doi SA. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modelling may have less limitations than other approaches. *Journal of Clinical Epidemiology* 2019.

35. Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; 20(15): 2219–2241.

36. Maier M, Bartoš F, Wagenmakers EJ. Robust Bayesian Meta-Analysis: Addressing Publication Bias with Model-Averaging. *PsyArXiv* 2020.

37. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in medicine* 2007; 26(11): 2331–2347.

5 | APPENDIX

Consider the evaluation of (18) for $\overline{\tau} > 0$, i.e., where a limited heterogeneity in the $\beta^{[k]}$ is permitted. Figure 2 (lower panel) shows how the global identification interval is wider when $\overline{\tau} = 0.50$ relative to when $\overline{\tau} = 0$ for the "fixed-effects" case. This interval outlined by the green rectangle can be easily obtained via quadratic programming.

Recall that quadratic programming constitutes the minimization of a quadratic function subject to linear constraints, and these may be a mix of equality and inequality constraints. Let x be a candidate value, which we will test for membership in the identification interval. To perform this test, we use a standard quadratic programming package (quadprog) to minimize the quadratic function $Var(\beta^{[k]})$, subject to the equality constraint $\theta = x$ and the inequality constraints which restrict β_k to the interval I_k for each k. The x value belongs in the identification interval if and only if the minimized variance does not exceed $\overline{\tau}^2$. Thus a simple grid search over values of x numerically determines the identification interval. Two numerical searches can be undertaken. One starts at the underlying value and tests successively larger x until a failing value is obtained. The other starts at the underlying value and does the same, but moving downwards.

□

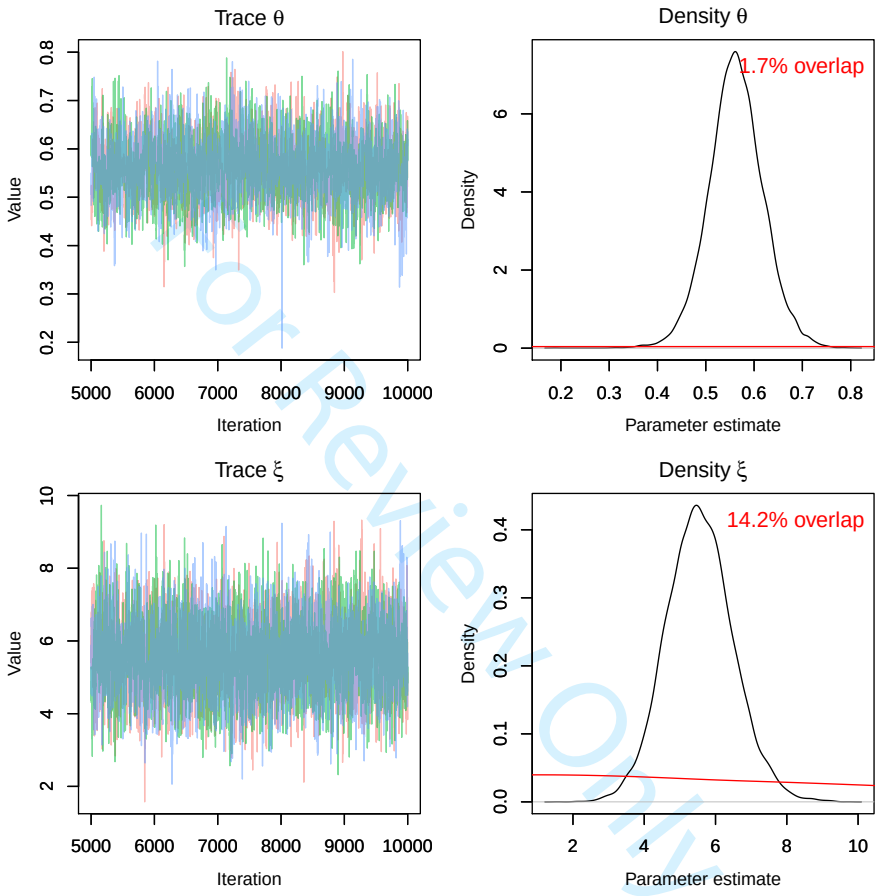


FIGURE 4 Diagnostic plots for parameters θ and ξ, for the MCMC simulation of the univariate *BMEMA* model with $k' = 5$ (this corresponds to the results in Table 3, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior-posterior overlap).

The NELS88 aggregate data

k	$n^{[k]}$	$\hat{\alpha}^{[k]}$	$\hat{\beta}^{[k]}$	$\hat{\sigma}^{[k]2}$	$\hat{\mu}^{[k]}$	$\hat{\lambda}^{[k]}$
1	45	4.99	0.64	12.28	14.35	5.29
2	64	7.91	0.42	19.50	12.00	6.51
3	47	8.46	0.51	11.48	17.06	3.20
4	45	2.95	0.49	15.71	11.26	6.54
5	45	1.75	0.67	15.15	8.57	5.59
6	59	5.53	0.54	19.45	12.33	6.48
7	56	5.48	0.56	17.31	14.52	4.84
8	45	9.69	0.39	13.51	17.04	3.47
9	51	6.60	0.61	6.93	17.86	3.13
10	67	6.33	0.67	13.88	16.71	4.29
11	48	4.03	0.57	11.62	11.58	5.09
12	45	6.39	0.62	20.55	13.54	5.39
13	47	4.02	0.50	15.16	13.67	4.47

TABLE 1 The NELS88 data- Aggregate data obtained from the NELS88 dataset for meta-analysis.

The NELS88* aggregate data								
k	$n^{[k]}$	$\hat{\alpha}^{[k]*}$	$\hat{\beta}^{[k]*}$	$\hat{\sigma}^{[k]2*}$	$\hat{\mu}^{[k]*}$	$\hat{\lambda}^{[k]*}$	$\phi^{[k]}$	$\gamma^{[k]}$
1	45	4.99	0.64	3.50	14.35	5.29	0.00	1.00
2	64	7.91	0.42	4.42	12.00	6.51	0.00	1.00
3	47	8.46	0.51	3.39	17.06	3.20	0.00	1.00
4	45	2.95	0.49	3.96	11.26	6.54	0.00	1.00
5	45	1.75	0.67	3.89	8.57	5.59	0.00	1.00
6	59	5.73	0.54	4.40	12.06	6.57	1.00	0.98
7	56	5.53	0.56	4.11	14.55	5.02	2.57	0.78
8	45	13.51	0.16	3.82	17.04	5.30	4.14	0.41
9	51	14.54	0.16	3.12	18.66	5.95	5.71	0.23
10	67	13.94	0.22	4.20	16.24	9.69	7.29	0.26
11	48	9.93	0.08	4.43	9.19	10.20	8.86	0.25
12	45	13.88	0.07	5.58	12.84	12.90	10.43	0.21
13	47	10.22	0.05	4.46	13.90	12.88	12.00	0.12

TABLE 2 A. The NELS88* data- Aggregate data obtained from the NELS88* dataset for meta-analysis.

dataset	k'	ξ	θ	$CI(\theta)_{95\%}$	τ	ω
1. NELS88*	13	8.97	0.32	0.20 , 0.45	0.19	3.95
2. NELS88	13	5.40	0.57	0.51 , 0.64	0.04	1.81
3. NELS88*	5	5.52	0.56	0.46 , 0.68	0.06	1.93
4. NELS88$^{1:5}$	5	4.90	0.56	0.41 , 0.72	0.08	2.40
5. NELS88*	0	0.39	0.94	0.66 , 1.24	0.08	1.13
6. NELS88	5	5.19	0.59	0.51 , 0.66	0.05	1.74
7. NELS88	0	4.90	0.60	0.52 , 0.80	0.05	1.66

TABLE 3 The data analysis results obtained - posterior medians with 95% equal-tailed credible intervals. R-code to replicate table: https://tinyurl.com/y5sq68mp.

A. NELS88* data

k	$n^{[k]}$	$\hat{\beta}_1^{[k]}$	$\hat{\beta}_1^{[k]*}$	$\hat{\beta}_2^{[k]}$	$\hat{\beta}_2^{[k]*}$	$\hat{\beta}_3^{[k]}$	$\hat{\beta}_3^{[k]*}$	$\sqrt{\Phi_{1,1}^{[k]}}$	$\sqrt{\Phi_{2,2}^{[k]}}$
1	45	4.38	4.38	0.39	0.39	0.18	0.18	0.00	0.00
2	64	5.47	5.47	0.26	0.26	0.22	0.22	0.00	0.00
3	47	4.31	4.31	0.28	0.28	0.26	0.26	0.00	0.00
4	45	2.35	2.35	0.20	0.20	0.18	0.18	0.00	0.00
5	45	0.23	0.23	0.34	0.34	0.28	0.28	0.00	0.00
6	59	3.59	6.70	0.27	0.19	0.25	0.14	6.00	12.00
7	56	2.29	4.53	0.31	0.31	0.29	0.20	6.00	8.00
8	45	3.60	15.94	0.26	0.04	0.25	-0.01	6.00	10.00
9	51	2.16	14.20	0.50	0.18	0.19	0.01	5.00	10.00
10	67	5.62	8.32	0.64	0.30	0.04	0.14	6.00	8.00
11	48	3.62	5.87	0.41	0.25	0.13	0.08	5.00	10.00
12	45	3.14	7.62	0.38	0.34	0.25	0.11	5.00	12.00
13	47	3.78	6.37	0.15	0.06	0.25	0.19	5.00	8.00

B. Analysis results with *BMEMA*, $Q = 2$

dataset	k'	θ_1	$CI(\theta_1)$	θ_2	$CI(\theta_2)$	θ_3	$CI(\theta_3)$	$\sqrt{T_{22}}$	$\sqrt{T_{33}}$
1. NELS88*	13	6.17	4.51, 7.97	0.25	0.19, 0.32	0.15	0.11, 0.2	0.05	0.06
2. NELS88	13	3.14	2.11, 4.28	0.34	0.26, 0.42	0.22	0.18, 0.26	0.07	0.02
3. NELS88*	5	2.41	0.86, 4.06	0.35	0.23, 0.49	0.25	0.18, 0.33	0.07	0.04
4. NELS88$^{1:5}$	5	2.99	0.97, 5.33	0.30	0.13, 0.46	0.23	0.14, 0.33	0.08	0.05
5. NELS88*	0	–	–, –	–	–, –	–	–, –	–	–
6. NELS88	5	2.28	1.13, 3.46	0.36	0.25, 0.48	0.24	0.18, 0.31	0.07	0.03
7. NELS88	0	–	–, –	–	–, –	–	–, –	–	–

TABLE 4 A. Aggregate data obtained from the NELS88 dataset and the NELS88* dataset for comparison. Values for $\Phi_{1,1}^{[k]}$ and $\Phi_{2,2}^{[k]}$ are also listed for reference. B. The data analysis results obtained - posterior medians with 95% equal-tailed credible intervals. R-code to replicate the table: https://tinyurl.com/y5w9jxg7.

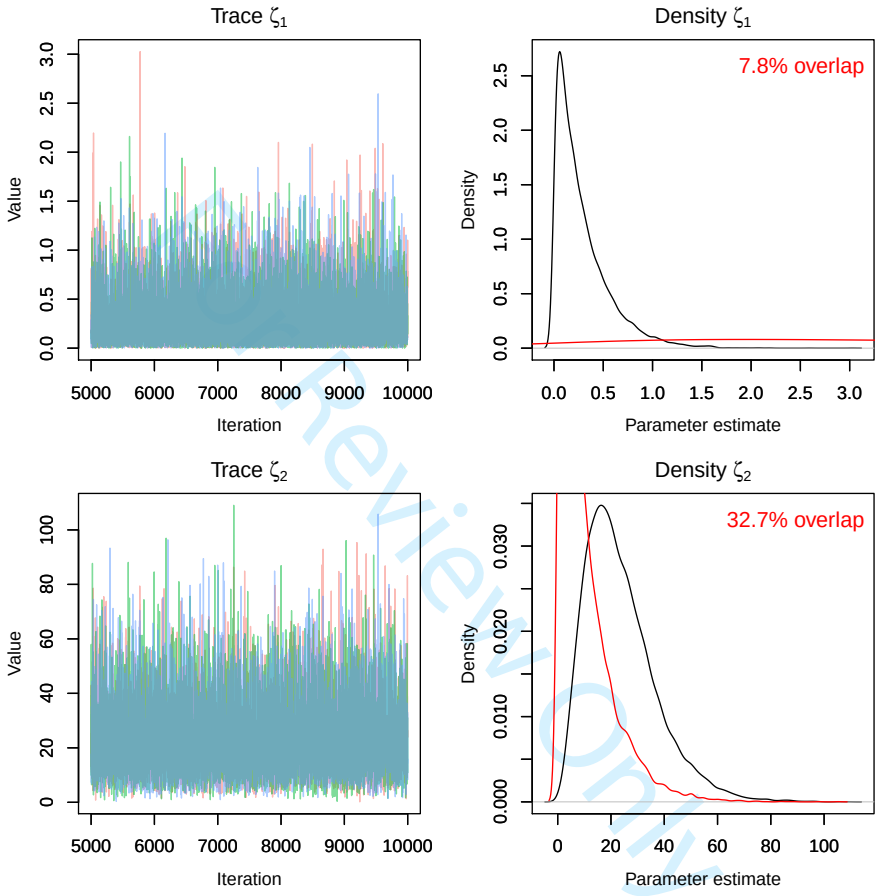


FIGURE 5 Diagnostic plots for parameters ζ_1 and ζ_2, for the MCMC simulation of the univariate *BMEMA* model with $k' = 5$ (this corresponds to the results in Table 3, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior-posterior overlap).

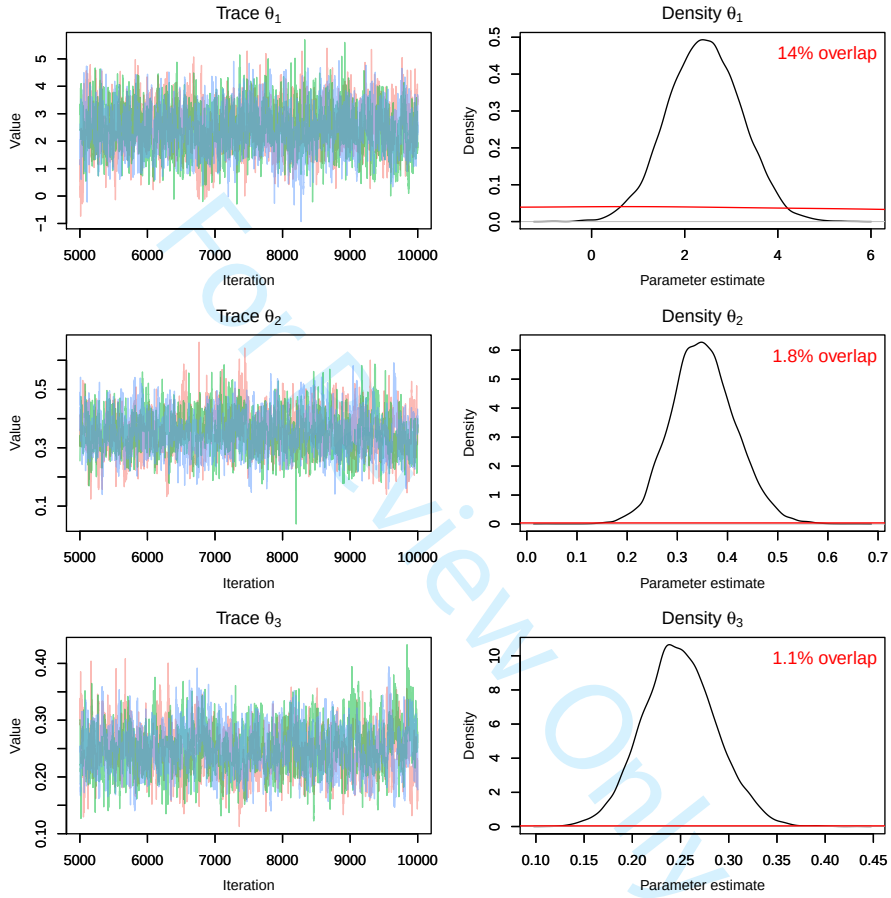


FIGURE 6 Diagnostic plots for the θ parameter, for the MCMC simulation of the multivariate *BMEMA* model with $k' = 5$ (this corresponds to the results in Table 4-B, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior-posterior overlap).

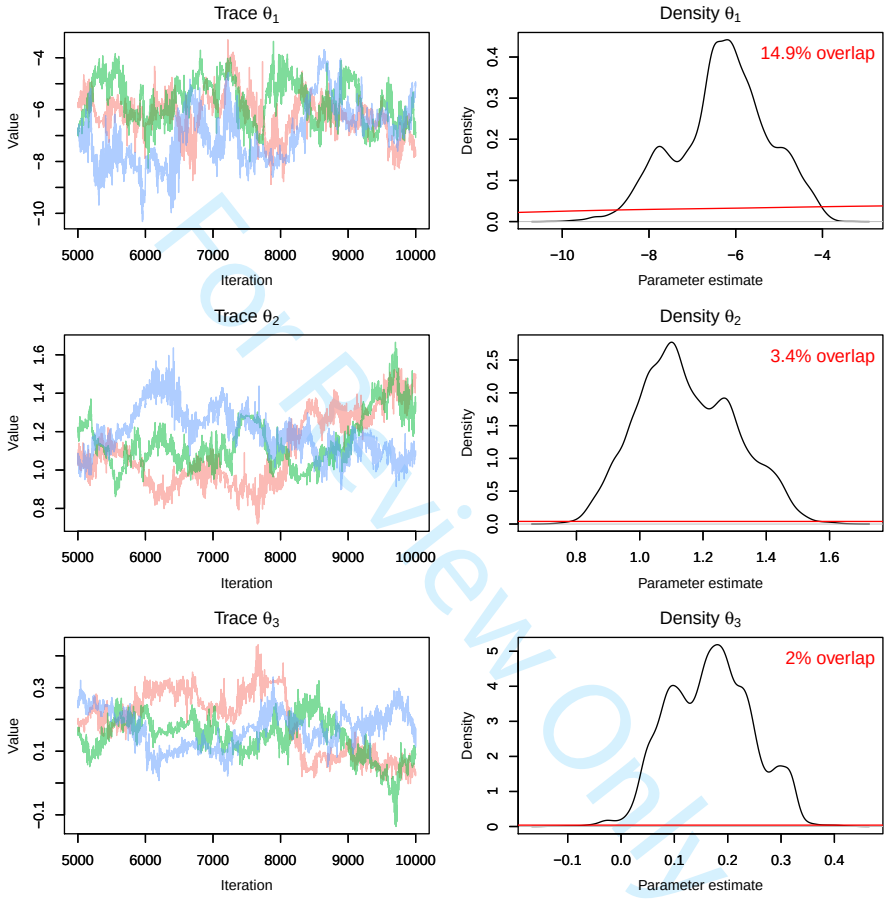


FIGURE 7 Diagnostic plots for the θ parameter, for the MCMC simulation of the multivariate *BMEMA* model with $k' = 0$ (this corresponds to the results in Table 4-B, line 5). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior distribution estimate (black solid line) together with the prior distribution for that parameter (red solid line). The % overlap reported in red is the PPO (prior-posterior overlap).