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

Journal:	*Research Synthesis Methods*
Manuscript ID	RSM-12-2020-0197.R1
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 Jaenisch, Thomas; University Hospital Heidelberg, Heidelberg Institute for Global Health; Colorado School of Public Health, Department of Epidemiology 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 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 look into the less than ideal situation in which contributing studies may be compromised by non-differential measurement error in the exposure variable. Specifically, we consider a meta-analysis for the association between a continuous outcome variable and one or more continuous exposure variables, where the associations may be quantified as regression coefficients of a linear regression model.} A flexible Bayesian framework is developed 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, these being measured with or without measurement error.

Response to reviewers
Manuscript ID RSM-12-2020-0197

Associate Editor

Comments to the Author:

On top of all the issues raised by the reviewers, a common problem in meta-analyses of observational studies based on aggregate data is that it is very rare the different studies to provide results from exactly the same regression model. Pooling coefficients from regression models including, for example, different covariates would not be appropriate. The authors should discuss this point in their manuscript and clarify under which conditions the proposed model can be validly applied.

Understood. We have simplified the Section 2 models and have now clarified, throughout the paper, the practical applicability of our models. Specifically at the start of Section 3, we write:

"Note that, while this generalization is of theoretical interest, in practice it may be unlikely to have multiple different studies provide results from exactly the same regression model. Should different studies adjust for different subsets of covariates, pooling their coefficients together in a meta-analysis may not be appropriate."

Furthermore, due to the limited practical applicability of the Section 3 models, we have cut down Section 3 substantially so that it includes only what is absolutely necessary.

Reviewer: 1

Major comments

The adjustment for measurement error (the (multivariable) BMEMA model) is really the heart of the paper, since the multivariable BayesMA model is completely standard. The method is nice, but I am not convinced that it is widely applicable in practice. You give only one example of a scenario where it nicely works, but that scenario does not apply in most practical cases. To convince me, you should devote much more discussion on this, giving more scenarios and examples.The univariable BayesMA model: What is wrong with a simple, straightforward, univariate two-stage meta-analysis on the beta_hat_k and its standard error? Please motivate why your much more complicated bivariate approach is needed. The same question applies to the multivariable BayesMA model.

Thank you for your feedback. These are all valid points. We have re-written much of Section 2 with your comments in mind and have added new paragraphs and references in the conclusion. We have also re-written all the Section 2 models so that they can be written out as simple, straightforward meta-analysis models on the beta_hat_k and its standard error.

The BayesMA model: The authors follow a two-stage meta-analysis approach, assuming that some extracted estimates from the first stage are fixed in the second stage- Motivate why you choose this approach, because it is not really needed. As you have the sufficient statistics, you can easily create a data set of pseudo IPD having the same likelihood as the unknown true IPD. You can work with these pseudo IPD data as if it were the true, the result is exactly the same. See for the details Papadimitropoulou et al. (RSM 2020). Then you can fit the model in one stage, using any linear mixed model program, without the need to assume these estimates fixed.

This is a very interesting idea. While we now discuss this idea and cite Papadimitropoulou et al. (RSM 2020), we have decided to proceed with the simpler approach which assumes that, as you say, “some extracted estimates from the first stage”. Based on our understanding, this is not that uncommon. Williams et al. (2018) note that the values for sigma2_i are “generally obtained from the individual studies and are assumed to be known” citing Chung, Rabe-Hesketh & Choi (2013) which we now also cite.

In the whole manuscript it is not clear when formulas follow from known standard statistical theory and when the authors have derived new results themselves. The principle is that you mention it clearly when known results are used (with a reference) and when you mention a result that you have derived yourself for the special purpose of the article. For instance formulas (1) to (4) are results from standard simple linear model theory, but that might not be obvious to all readers of RSM. So give a reference to a textbook. This comment applies to several other passages in the manuscript.

We now explicitly note which models and equations are known/standard. Changes have been made with this in mind throughout the manuscript (e.g., we write “follow according to standard theory” and “can be estimated in a standard univariate random-effects meta-analysis”) and we have now added references to textbooks and tutorial-type papers.

Formulas (5) and (6): It is very unrealistic to assume that alpha and beta are independent. This is a standard random intercept - random slope mixed linear model, and in every textbook it is said that you should not forget the correlation, since in practice it is almost always there. Also think of control risk regression, where Schmid et al. (Stat Med 1998) have shown that correlation is almost always there in practice. You do not make the analysis more difficult in theory or practice by allowing the correlation. This also applies to the multivariable BayesMA and (multivariable) BMEMA models, where you assume all regression coefficients independent.

Very good observation. We have changed the models with this in mind and introduced the rho correlation parameter. As you say, it does not make the analysis any more difficult.

Formulas (10), (19), (24) and (26) and (31) : These are unnecessary formulas and could be deleted. Everybody knows that that the posterior is proportional to the product of the likelihood and prior density. This saves a lot space and make the manuscript look less mathchy.

Good suggestion. We have eliminated all but 1 of these formulas.

The (multivariable) BMEMA model: How sensitive are the results for different choices of the prior distribution on the phi's? Typically in practice there is not much information in the data on the phi's, the phi's can even be (almost) unidentifiable. So I suspect that the choice of this prior can be very tricky. Discussion of this is and sensitivity analyses are desirable. Even if the overlap in prior and posterior is small, the choice of the prior can still be important.

We now run a second version of the BMEMA model with a different prior specification (delta=0.1 vs. delta =0.5) in order to get a better idea as to the impact of the priors.

Somewhere here you might stress the advantages of the Bayesian approach primarily apply to the BMEMA models. The BayesMA models are easily fitted in a frequentist manner with standard statistical packages, but for the BMEMA model that is not the case.

We have added two notes on this point:

"That being said, frequentist models are often easily implemented with standard statistical packages whereas Bayesian models may require a certain amount of customization."

and:

"Note that while the BayesMA model (or something similar) could be easily fit in a frequentist manner with standard statistical packages. However, this is not the case..."

Minor comments

Abstract
The beginning of the abstract can be made more specific. Begin the abstract with a sentence like the one below and then the rest of your abstract follows.
"The situation considered is meta-analysis of the association between a continuous outcome variable and one or more continuous exposure variables, where the associations are quantified as the regression coefficients of a linear regression model."

We have made edits to the abstract to make it more specific.

Page 3 of 27, line 52-53: This is an illogical sentence. You simply mean: "The assumed model is the simple linear regression model ..... for Y given X and a normal distribution for …."

(What do you mean by the "sufficient statistics" here? The sufficient statistics are the sample means of Y and X together with the sample covariance matrix of Y and X.)

Fixed.

Page 4 of 27
line 2: “sufficient statistics”. You should have 5 sufficient statistics: basically you have two means and three variance/covariance parameters. So I miss one.
I should expect hats on sigma and lambda. They are statistics, not parameters here, even if, in the second stage, these estimates are assumed to be fixed.

Fixed.

Line 7-8: “should make little practical difference [16]”. I am not so sure of this, because the referenced article is not appropriate. It considers the most simple situation in meta-analysis where you have observed measure plus standard error as input data in a standard two-stage meta-analysis. But the situation here is more complicated. So why would this claim be true?

We have changed the wording on this. We now also discuss the pseudo-data idea and have added references to Papadimitropoulou et al. (RSM 2020) and Chung, Rabe-Hesketh & Choi (2013).

Formulas (5) and (6): You use an awful lot of symbols in this article and the reader will have difficulties to keep track of them. You might make it easier to replace the symbols ksi and omega by alpha and beta.

We understand that there are a lot of greek letters! However, alpha and beta are already defined as the study-specific parameters.

Formula (7): You might mention that the model as yet is nothing special. It is just the well known bivariate case of the general two-stage multivariate meta-analysis model (e.g. Chapter 10 of Handbook of Meta-Analysis).

Good point. Changes have been made and we have added a reference to the Handbook of Meta-Analysis.

Line 37-38: “is only one out of four different sources of variances”. What do you mean by sources of variance? In fact, you have two sources: between studies and within studies variance, which are both of bivariate nature characterized by two or three (co)variance parameters.

This section has been re-written.

Line 55-56: “is only a minor inconvenience”. I think you forget to mention the major inconvenience that very often (at least in the medical literature) the residual standard deviation sigma is not given in a publication. What to do then?

Very good question. This section has been re-written so as to clarify that the models can be fit with only the beta_hat_k and its standard error.

Page 5 of 27, line 20-21: There are four unknown parameters, but you give six. Maybe the (non-defined) alpha and beta are reminiscences from an early version where you used different notation?

The alpha and beta are unknown parameters in the random-effects model.

Page 6 of 27, line 39-40: “fact”. This is no fact but an assumption.

Fixed.

Section 2.2: I think the pseudo IPD approach works also in the measurement error case.

Perhaps, although things do get progressively more complex as the number of covariates increases and the measurement errors are correlated. We shall consider this idea in future work.

Page 7, line 32-32: Somewhere here insert a few sentences explaining the standard theory for measurement error in the independent variable of a simple linear regression. Maybe many readers will not know the attenuation factor.

Good suggestion. We now add a reference to Hutcheon et al. (2010) which provides a very nice introduction to the main ideas.

Line 28 and following: I would like much more discussion about the practical applicability of the BMEMA model. You discuss one specific scenario, which is nice, but could you give other scenarios in which the BMEMA model works well? When is it not feasible? Also discuss here the identifiability problems (in the multivariable case you do that a bit, but it belongs better here). What are conditions under which the method will work? When does it not work well? I have the feeling that the BMEMA model needs always a good “anchor”. If you wish that people are going to use your method, more discussion and examples are needed.

The practical applicability of the BMEMA model is no doubt limited. However, we believe this is an important step in the right direction when it comes to addressing the important problem of measurement error in evidence synthesis. We have added a few references to “meta-analysis of linear regressions” (Crouch (1995,1996), and Lau et al. (1999)) and well as references to meta-analyses, that without “tools to correct for measurement error”... “are left to simply list

measurement error as a study limitation (e.g. Wu et al. (2016): ``this study has several limitations [...] none of the studies corrected for measurement error.''; and Merino et al. (2009): ``although every effort was made to maximize the validity of the study, minimize bias, and incorporate heterogeneity and uncertainty, the estimated hazard ratios of dietary components could be affected by measurement error'').

Page 13 of 27, line 48-49: Why diagonal? I think this is an unrealistic assumption. There is a whole bunch of models available for this covariance matrix that allow correlation.

We now state that: “If we are able to assume that the regression coefficients are a priori independent, then \Tau will be a diagonal matrix.”

Section 3.1: If you have multivariable IPD, then your one-stage approach is better replaced by a one-stage approach, needing less assumptions. If you don’t have IPD, but only the sufficient statistics, you can use pseudo IPD with exactly the same results as if you really had the IPD.

We discuss the pseudo-data idea and have added a references to Papadimitropoulou et al. (RSM 2020).

Section 3.2: Is this multivariable BMEMA also amenable to the pseudo IPD if you only have the sufficient statistics and not the true IPD? I think so.

Perhaps, although things get progressively more complex as the number of covariates increases and the measurement errors are correlated. Further research on this is required.

Page 17 of 27, lines 14-18: I doubt whether this is a good example. It only makes sense if you are willing to assume that the differences between the randomised and observational studies must be due to measurement error in the observational studies. But this is not justified; there might be real differences, due to other causes. You could add a covariate (randomized versus observational) to the between studies model, but that probably makes the phi's unidentifiable. Could you please give another example?

We have added a comment on this:
“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. That being said, the BMEMA model can only do so much. We stress that the BMEMA model will assume that all differences between the subset of unbiased studies and the biased studies are due to measurement error when in reality there may be other systematic differences at play.”

All tables: The columns are now headed with only the symbols. Please give also the meaning: study number, sample size, estimated intercept, estimated regression coefficient of reading test score, …

We have made some minor changes to the tables (captions and column headers) to help clarify things.

Reviewer: 2

Comments to the Author
This paper addresses an important topic in meta-analysis. I am unclear as to the novel contribution of this work, as although the authors point to previous work they do not explain what this previous work included and how their work adds to / improves on this. The paper is a difficult read, and would benefit from moving some of the more technical details to appendices and giving the reader a more intuitive explanation in the text. Improving the structure of the paper may help, giving the motivating example early on, collecting all methods together, followed by a dedicated results section. The introduction and discussion sections in particular could be improved.

We agree that this is an "important topic in meta-analysis"- thank you for your feedback. We have made substantial revisions in an effort to make the paper an easier read and have removed much of the more technical material.

Major comments
1. General: this is highly technical and a difficult read. I think the concepts are easy enough for a general reader to follow, but as currently written the general reader would struggle. I suggest some of the more technical material could be moved to appendices and a more intuitive description given in the main text. This would also help to shorten what is a long paper (at least the methods section is very long).

Understood. We have removed many of the most complex equations and have added many additional references to help readers better understand. Also, we have cut down SEction 3 (perhaps the most technical section) to the bare minimum.

2. Methods equation (5). The alpha parameters are nuisance parameters and should be given independent priors, rather than come from a random effects model. The primary focus for meta-analysis are the beta's (as you acknowledge), and it is these that should be given the hierarchical model ie equation (6). Putting a random effects model on the alphas can introduce bias in the pooled estimate of the beta's, which we want to avoid. The same comment applies to the multivariable regression in equation (23) for the intercept term beta0.

We have re-written the models so that they can be fit with only the betas and their standard errors (i.e., without the alphas).

3. Partial identifiability. I apologise but I couldn't follow what you were saying here. I think it needs to be described in more intuitive language. It relies on setting a bound for tau, but how is

this chosen? Or would you always explore sensitivity to different values of tau_bar? Or use tau_bar to constrain the prior for tau? I don't know how to interpret Fig 2. It's also not clear how this links to the rest of the paper / subsequent section on BMEMA. It is not mentioned again until the very end of section 3. Perhaps the partial identifiability section could be moved to an appendix? Or omitted from the paper and just mentioned in the discussion.

We added some text to better introduce (and motivate) this section:

"For a scenario in which k'=0, things are not so straightforward. Without any ``gold standard'' studies and without any specific knowledge regarding the degree of measurement error for each of the K contributing studies, the parameters of interest may not be identifiable. With this in mind, before moving on, we briefly consider an asymptotic argument for so-called ``partial identifiability'' by considering the degree to which \theta can be estimated in the presence of unspecified measurement error."

4. Discussion section is weak. There is lots more to discuss. There is no mention about work of previous authors and what this work has added. I didn't get a feeling for the novel contribution of this article. There should be some discussion as to how you would incorporate differential measurement error. Also, the authors mention combining RCT and observational evidence after adjusting for measurement error, and although they mention some other types of bias that need to be considered, selection bias isn't mentioned or discussed. How big an effect do you think measurement error has compared to these other biases? How do you propose researchers use the methods in practise? Adjust for all studies? Or just those where you suspect issues with measurement error? What about measurement error in the outcomes? For multivariable regression what if IPD isn't available? Or available for just some studies? What if only some summary results are reported (eg beta's may be reported but not alpha's)? What to do about issues with lack of identifiability? Etc, etc.

Thank you for all the suggestions! We have added several of these important points of discussion in the conclusion.

Minor comments

5. Both the title and abstract should make it clear that this paper is concerned with non-differential measurement error.

Done.

6. Introduction. I think it would be easier for the reader if you list some of the "wide range of biases" (p2 of 27, line 46) in observational studies, and then say that an important bias that receives relatively little attention is measurement error, so it's clear that this is one of many biases and why you are focussing on it in this paper.

Good point - done.

It would also be helpful to distinguish between differential and non-differential measurement error early in the introduction. It's mentioned later in the introduction, but not fully explained. When you say "observational studies" I think you mean prospective comparative studies looking at exposure outcome relationships? Or are you also considering retrospective studies? I think the issues with measurement error may differ substantially for retrospective and prospective studies.

Good point - we have added a comment on this with reference to White (2003).

7. p3 of 27, top par. The comments about RCTs are a bit confused and hard to follow. I think this will be easier to explain when you have defined differential and non-differential measurement error, as you wouldn't expect the measurement error to be differential in an RCT, but it could be in an observational study. At this point in the manuscript I wasn't sure what kind of measurement error you were considering … but it later becomes clear that it's for the exposure variable, which is why the issue is more problematic for observational studies (as exposure is well defined in RCTs). A clearer introduction to terminology and what you are addressing in this paper would be helpful here.

We have added a phrase at the very beginning to make this clear: "Our focus will be on measurement error which we define as the error due to inaccurate measuring of the exposure variable(s)."

8. Introduction. The reference to other work in the intro didn't give me a feeling for the methods already proposed, their limitations and the motivation for the current work and what the additional novel contribution of this work is. The introduction needs to motivate the work, so by the end of it the reader realises the limitations in previous work and the need for better approaches. The quotes on the advantages of Bayesian approaches demonstrate my frustration … you say Bayesian approaches "have significant advantages" and leave it at that! What are the advantages? Briefly summarise what they are and hence motivate why you are taking a Bayesian approach here. A Bayesian approach has been taken before (Lian), so what does your approach add to this? You do in fact discuss this on p.5 of 27 in the middle of the methods section. But it would live better in the introduction.

Fair. We have now expanded the discussion of previous work in the introduction to clarify the specific gaps in the existing literature.

9. Methods. Superscript [k] notation is quite difficult. Why not include this as a subscript, so notation is (X_{j,k},Y_{j,k})?

We like this notation but recognize that it might not be for everyone. If the journal has guidelines for notation, or if the AE has advice for standard usage, we would be happy to make changes.

10. p4 of 27, line 2. Should mu be listed here too?

This section has been re-written.

11. Fig 1. Legend says that the credible intervals are symmetrical, but posterior distributions may be skewed?? What should be plotted are the 2.5 and 97.5 quantiles. Perhaps that is what you have done, but it's not clear from the description. These study results look very heterogeneous, and I'm not sure how sensible it is to pool them at all (even after accounting for measurement error …). Measurement error is probably the least of your problems in the analysis of this data set!

Good observation- the plot was the problem. We had simply made a basic mistake in the coding for the uncertainty intervals. Our apologies for this unneeded confusion! We have fixed this now and we have also added a note in the captions to clarify that we are plotting the 2.5 and 97.5 quantiles.

12. Section 2.4. Spell out what BMEMA stands for

Added: "BMEMA (Bayesian model for Measurement Error in Meta-Analysis)"

13. Fig 3. Figure should be understandable on its own. The (k', K) notation is opaque and should be explained in the legend. The reader is referred to Table 3, but I don't know how to interpret the lines in Table 3 either. I'm really quite confused!

We have clarified the figure captions and Table 3 as well.

Also, here you purposefully manipulated some studies to have measurement error and not others, and analysed the studies using this information. In practise you would not know this, and would adjust all studies for measurement error. So, I think that although there is a gain in precision from including all studies, the reduction in bias from adjusting for measurement error will depend on being able to identify those studies where there is measurement error. What if you adjusted all studies (even those without it). What would the estimate look like then?

Good question. The results for this scenario are listed in Table 3, line 5. The estimates are biased upwards in this situation. In other words, the model overcompensates.

What do the open square boxes represent? I know these are the studies that have been manipulated, but why isn't the uncertainty in these estimates plotted? Or are they super-precise? Did you purposely add measurement error to studies that had very low

estimates? If so, perhaps you have exaggerated the impact of accounting for measurement error in your example.

Good observation- the plot was the problem. We have fixed this now.

14. My intuition is that although it is important to account for measurement error, there are other factors that need accounting for with observational studies that may have bigger implications in practise. I didn't get a sense of this from the paper. It should be expanded on in the discussion. In particular what about differential measurement error?

Your intuition is good and we do note that: "beyond the bias caused by measurement error, a meta-analysis of observational studies should, ideally, also take into account other (potentially bigger) biases, e.g.: publication bias and bias due to unmeasured confounding" We have also added a comment on differential measurement error with reference to White (2003).

15. **Illustrative example. I think it would be helpful to introduce this earlier in the paper …** before methods. So you can motivate the problem/methods with the example. Then come back to it (as you do) later in the manuscript. Currently you intersperse methods with results. It might be easier for the reader to digest if the methods are all given first, then the implication and results.'

We have re-worked the sub-sections so as to have less interspersing of the methods and the example.

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 subject to non-differential measurement error

Harlan Campbell*[1] | Valentijn M.T. de Jong[2] | Lauren Maxwell[3] | Thomas Jaenisch[3,5] | Thomas P.A. Debray[2,4] | 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 look into the less than ideal situation in which contributing studies may be compromised by non-differential measurement error in the exposure variable. Specifically, we consider a meta-analysis for the association between a continuous outcome variable and one or more continuous exposure variables, where the associations may be quantified as regression coefficients of a linear regression model. A flexible Bayesian framework is developed 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, these being 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 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, including selection bias and bias due to unobserved confounding [4]. One important bias that receives relatively little attention is measurement bias. 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] Our focus will be on measurement error which we define as the error due an inaccurate measuring of the exposure variable(s).

If the measurement error affecting a particular study is of known magnitude, adjustment for measurement 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]). While practical, these approaches fall short if the degree of measurement error is unknown. 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. These methods are developed "[a]ssuming that data are available for consistent estimation of [the attenuation factors]"[11]. When such data are unavailable, the proposed attenuation factors fail to provide adequate adjustment.

More recently, Lian et al. (2019)[12] introduce Bayesian meta-analysis models for binary outcomes accounting for exposure misclassification. These models are well designed but do not consider the possibility of measurement error in a continuous exposure or how to address continuous outcome data. Finally, in the applied literature, Zeisser (2014)[13] discuss the impact of likely exposure misclassification (i.e., measurement error of a binary exposure variable) 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 a potentially unknown degree of 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] due to their inherent flexibility to handle more complicated data structures. Bayesian methods also offer a "number of specific advantages" for meta-analysis; see Sutton and Abrams (2001)[16]. In Section 2, we outline a proposed Bayesian

framework for the case of meta-analysis with measurement error in 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 and, for each of these studies, the exposure and the outcome are continuous variables. We begin by describing some basic distributional assumptions for the underlying data.

Let $\left(X_j^{[k]}, Y_j^{[k]}\right)$ be the exposure and outcome for the j-th observation in the k-th study. We assume that the exposure and outcome are related by means of a conditional Normal distribution such that: $Y_j^{[k]}|X_j^{[k]} \sim \mathcal{N}\left(\alpha^{[k]} + \beta^{[k]}X_j^{[k]}, \sigma^{[k]2}\right)$; for $k = 1, \ldots, K$ and j in $1, \ldots, n^{[k]}$. Furthermore, we assume that each study has its own exposure distribution governed by a Normal distribution: $X_j^{[k]} \sim \mathcal{N}\left(\mu^{[k]}, \lambda^{[k]2}\right)$. Finally, the study specific parameters, $\alpha^{[k]}$ and $\beta^{[k]}$, are related to one another such that:

$$\begin{matrix}\alpha^{[k]} \\ \beta^{[k]}\end{matrix} \sim \mathcal{N}\left(\begin{pmatrix}\xi \\ \theta\end{pmatrix}, \begin{pmatrix}\omega^2 & \rho\omega\tau \\ \rho\omega\tau & \tau^2\end{pmatrix}\right). \tag{1}$$

for $k = 1, \ldots, K$, 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 the main "parameter of interest"), τ^2 is the variance of slopes across studies, and ρ is the correlation of the regression coefficients.

For each of the K studies, a standard simple linear regression model could be fit to the outcome and exposure data to obtain least-squares parameter estimates, $\hat{\alpha}^{[k]}$ and $\hat{\beta}^{[k]}$, which will follow, according to standard theory [17], a bivariate Normal distribution such that, for $k = 1, \ldots, K$:

$$\begin{matrix}\hat{\alpha}^{[k]} \\ \hat{\beta}^{[k]}\end{matrix}\Bigg|\begin{matrix}\alpha^{[k]} \\ \beta^{[k]}\end{matrix} \sim \mathcal{N}\left(\begin{pmatrix}\alpha^{[k]} \\ \beta^{[k]}\end{pmatrix}, \begin{pmatrix}\Sigma_{11}^{[k]} & \Sigma_{12}^{[k]} \\ \Sigma_{12}^{[k]} & \Sigma_{22}^{[k]}\end{pmatrix}\right), \tag{2}$$

$$\text{where:} \quad \Sigma_{11}^{[k]} = \frac{(\lambda^{[k]2} + \mu^{[k]2}) \times \sigma^{[k]2}}{\lambda^{[k]2} \times n^{[k]}}; \tag{3}$$

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

$$\text{and} \quad \Sigma_{22}^{[k]} = \frac{\sigma^{[k]2}}{\lambda^{[k]2} \times n^{[k]}}. \tag{5}$$

A meta-analysis will typically combine the summary statistics reported in each contributing study to obtain an overall estimate for the parameter(s) of interest. If the value of $\hat{\beta}^{[k]}$ and its standard error, $\text{se}(\hat{\beta}^{[k]})$, are available for $k = 1, \ldots, K$, the primary parameter of interest, θ, can be estimated in a standard univariate random-effects meta-analysis[18] in which:

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

Or if, in an admittedly rare situation (Becker et al. (2007)[19] point to Crouch (1995,1996)[20,21] and Lau et al. (1999)[22] as examples), data are also available for $\hat{\alpha}^{[k]}$, $\text{se}(\hat{\alpha}^{[k]})$ and $\hat{\sigma}^{[k]2}$ for $k = 1, \ldots, K$, one may fit a bivariate meta-analysis model[23,24] in which:

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

where $\Sigma_{11}^{[k]}$, $\Sigma_{12}^{[k]}$, and $\Sigma_{22}^{[k]}$ are given by equations (3), (4), and (5) with the $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$ parameters assumed to be known (i.e., measured without error) and equal to:

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

and

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

Note that with a sufficient amount of data, the uncertainty surrounding the $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$ parameters will be quite small[16], and the simplifying assumption that the $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$ parameters are known should make little practical difference[25]. However, note that if one wished to properly account the additional uncertainty of $\mu^{[k]}$, $\sigma^{[k]}$, and $\lambda^{[k]}$, a suitable strategy would be to create pseudo-values for $\left(X_j^{[k]}, Y_j^{[k]}\right)$, for $k = 1, \ldots, K$ and j in $1, \ldots, n^{[k]}$, using the observed sufficient statistics (if these were all available). This pseudo individual participant level data would have the same likelihood as the true unknown underlying IPD and could then be fit –as if it were the true data– with an IPD-meta-analysis model (as in Section 3.1); see Papadimitropoulou et al. (2020)[26].

Inference for either the univariate or bivariate meta-analysis (i.e., for either (6) or (7)) can be done within either a frequentist or a Bayesian framework[19,27,28]. However, there are several reasons why a Bayesian approach may be advantageous. For instance, unlike Bayesian models, frequentist meta-analysis models are known to have difficulty estimating variance parameters if these parameters are near-zero, particularly when sample sizes are small[29,25]. Also, Bayesian models also offer substantial flexibility for handling complicated data structures that may arise with multiple covariates and the possibility of measurement error[16,15,30]. That being said, frequentist models are often easily implemented with standard statistical packages whereas Bayesian models may require a certain amount of customization[31].

A Bayesian model requires defining priors for all of the unknown parameters and, for better or worse, the performance of any Bayesian estimator will depend on the choice of these priors. Particularly when few data are available, the choice of priors can

substantially influence the posterior[32,33,34]. In the examples considered throughout this paper, our strategy will be to adopt wide Normal distributions (with variance of 100) for the mean parameters, weakly-informative half-Cauchy priors (with scale parameter of 2) for the variance parameters, and a uniform distribution for the correlation parameter; following the recommendations of Polson et al. (2012)[35] and the simulation results of Williams et al. (2018)[28].

Before going on to discuss measurement error, let us briefly demonstrate how standard Bayesian univariate and bivariate meta-analysis models (*BayesMA*) can be used in an analysis of some simple illustrative data.

2.1.1 | Example: the NELS88 dataset

The NELS88 dataset has been used previously as an example dataset by Becker et al. (2007)[19] and is from a survey of U.S. grade 10 high-school students in 1988 from over 1,000 schools. Becker et al. (2007)[19] 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 $N = \sum_{k=1}^{K} n^{[k]} = 664$ students from $K = 13$ different schools.

Table 1 displays the aggregate data from the NELS88 dataset required for both the univariate and bivariate meta-analyses. We fit these data with *BayesMA* models defined in Section 2.1 with the following priors:

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

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

$\tau \sim$ half-Cauchy$(0, 2)$, $\tau > 0$; (location of 0, scale of 2);

$\omega \sim$ half-Cauchy$(0, 2)$, $\omega > 0$; and

$\rho \sim$ Uniform$(-1, 1)$.

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[36]. Each model is fit based on 100,000 Monte Carlo draws from each of three chains (thinning of 10), and for each we report posterior medians and equal-tailed 95% credible intervals. Note that while the *BayesMA* model (or something similar) could be easily fit in a frequentist manner with standard statistical packages. However, this is not the case for the *BMEMA* model we introduce in Section 2.2.

Results for the univariate and bivariate models are very similar but not identical. For the univariate model, we obtain posterior medians: $\hat{\theta} = 0.56$ with 95% equal-tailed credible interval of $CI(\theta)_{95\%} = [0.49, 0.62]$; and $\hat{\tau} = 0.04$. For the bivariate model, we obtain the posterior medians: $\hat{\theta} = 0.57$ with 95% equal-tailed credible interval of $CI(\theta)_{95\%} = [0.50, 0.64]$; $\hat{\tau} = 0.04$; $\hat{\xi} = 5.41$ with 95% equal-tailed credible interval of $CI(\xi)_{95\%} = [4.05, 6.85]$; $\hat{\omega} = 1.85$; and $\hat{\rho} = 0.07$. Figure 1 displays a forest plot summarizing the primary analysis results for the bivariate model.

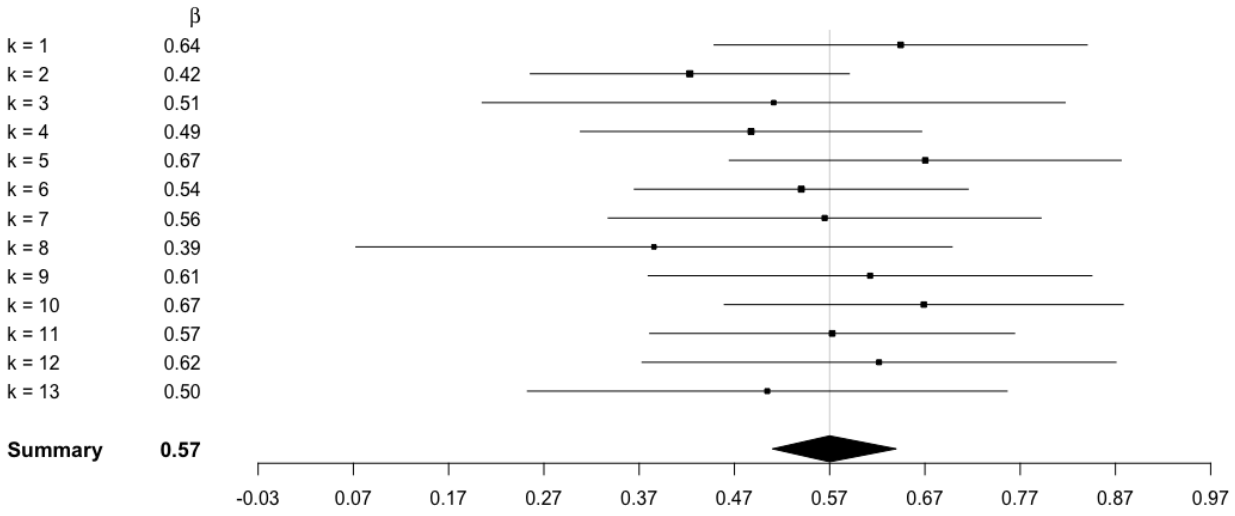

FIGURE 1 Forest plot for the meta-analysis of the NELS88 aggregate data (Table 1). The $K = 13$ black squares correspond to the $\hat{\beta}^{[k]}$ values, for $k = 1, \ldots, K$; with horizontal lines corresponding to symmetrical $1.96 \times \text{se}(\hat{\beta}^{[k]})$ confidence intervals. The *BayesMA* posterior estimate of θ is plotted as a diamond (labelled "Summary"), the lateral points of which indicate the equal-tailed 95% credible interval (i.e., the 2.5% and 97.5% quantiles) for this estimate.

2.2 | Adjusting for non-differential measurement error

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

In this situation, we wish to determine the relationship between the outcome, Y, and the exposure, X, with data based instead on measuring Y and X^*. Ignoring the measurement error in such a situation will bias the estimates of the regression slope coefficients towards the null (i.e., will bias $\hat{\beta}$ towards 0); see Hutcheon et al. (2010)[37] for an excellent review. We assume the vector of independent surrogates, X^*, arises from a classical additive measurement error model:

$$X_j^{[k]*}|X_j^{[k]} \sim \mathcal{N}\left(X_j^{[k]}, \phi^{[k]2}\right), \text{ for } j \text{ in } 1, \ldots, n^{[k]}, \tag{10}$$

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

$$X_j^{[k]*} \sim \mathcal{N}\left(\mu^{[k]}, \lambda^{[k]*2}\right) \tag{11}$$

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

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 $k = 1, \ldots, K$. The range of values for $\phi^{[k]}$ is therefore restricted to: $0 \leq \phi^{[k]} \leq \lambda^{[k]*}$. Linear regression coefficients estimated for each of the individual studies will be biased and governed by:

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

$$\text{where:} \quad \Sigma_{11}^{[k]*} = \frac{(\lambda^{[k]*2} + \mu^{[k]*2}) \times \sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}}; \tag{13}$$

$$\Sigma_{12}^{[k]*} = -\mu^{[k]*} \times \frac{\sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}}; \tag{14}$$

$$\text{and } \Sigma_{22}^{[k]*} = \frac{\sigma^{[k]*2}}{\lambda^{[k]*2} \times n^{[k]}}; \tag{15}$$

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 $k = 1, \ldots, K$.

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

$$\theta^* = \mathrm{E}\left(\gamma^{[k]}\beta^{[k]}\right) = \mathrm{E}\left\{\left(1 + \frac{\phi^{[k]^2}}{\lambda^{[k]^2}}\right)^{-1} \times \beta^{[k]}\right\}, \tag{16}$$

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

$$\theta^* = \mathrm{E}\left(\gamma^{[k]}\right)\theta. \tag{17}$$

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 $k = 1, \ldots, K$. 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 $k = 1, \ldots, K$. Invoking independence of $\gamma^{[k]}$ and $\beta^{[k]}$, we have:

$$\tau^{*2} = \left\{\mathrm{E}\left(\gamma^{[k]}\right)\right\}^2 \tau^2 + \mathrm{Var}\left(\gamma^{[k]}\right)\left(\tau^2 + \theta^2\right), \tag{18}$$

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 $k = 1, \ldots, K$. However, the second term could counteract this. More precisely, if the magnitude of the cross-study variation in measurement error (i.e., $\mathrm{Var}(\gamma^{[k]})$) and/or the average effect size (i.e., θ) 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 (18) reduces to $\tau^{*2} = \text{Var}\left(\gamma^{[k]}\right)\theta^2$. 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 ($\tau^2 = 0$).

For the bivariate model, the bias in estimating ξ and ω^2 must also be considered. If we ignore measurement error, then our estimation procedures will mistakenly target, for $k = 1, \ldots, K$:

$$\xi^* = \text{E}\{\alpha^{[k]} + (1 - \gamma^{[k]})\beta^{[k]}\mu^{[k]}\} = \xi + \text{E}\{(1 - \gamma^{[k]})\beta^{[k]}\mu^{[k]}\}, \tag{19}$$

and

$$\omega^{*2} = \text{Var}\left(\alpha^{[k]*}\right) = \text{Var}\left(\alpha^{[k]} + (1 - \gamma^{[k]})\beta^{[k]}\mu^{[k]}\right). \tag{20}$$

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.

The standard *BayesMA* models described in Section 2.1, can be adapted to account for non-differential measurement error in a relatively straightforward manner. The univariate *BMEMA* (Bayesian model for Measurement Error in Meta-Analysis) can be defined in two parts as:

$$\hat{\beta}^{[k]*}|\beta^{[k]} \sim \mathcal{N}(\gamma^{[k]}\beta^{[k]}, (\text{se}(\hat{\beta}^{[k]}))^{*2}) \qquad \text{and} \qquad \beta^{[k]} \sim \mathcal{N}(\theta, \tau^2), \tag{21}$$

where $\text{se}(\hat{\beta}^{[k]})$ is the standard error for $\hat{\beta}$ as reported by a study potentially biased due to measurement error; and where, if data for $\lambda^{[k]*2}$ are available, we can define:

$$\gamma^{[k]} = \left(1 + \frac{\phi^{[k]2}}{(\lambda^{[k]*2} - \phi^{[k]2})}\right)^{-1}. \tag{22}$$

The bivariate *MEMA* model can also defined in two-parts with the conditional bivariate normal likelihood for $(\hat{\alpha}^{[k]*}, \hat{\beta}^{[k]*}|\alpha^{[k]}, \beta^{[k]})$ as specified by equation (12), and the bivariate normal likelihood for $\alpha^{[k]}$ and $\beta^{[k]}$ as specified by equation (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 $\gamma^{[k]}$ is known and equal to 1 (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.

For a scenario in which $0 < k^{'} < K$, our strategy will depend on whether or not, for $k = k^{'} + 1, \ldots, K$, data are available for $\hat{\lambda}^{[k]*}$, as this can serve as an upper bound on $\phi^{[k]}$ (recall that: $0 \leq \phi^{[k]} \leq \lambda^{[k]*}$). If $\hat{\lambda}^{[k]*}$ data is not available, we can simply specify a uniform prior on $\gamma^{[k]}$, such that, for $k = (k^{'} + 1), \ldots, K$:

$$\gamma^{[k]} \sim \text{Uniform}(0, 1). \tag{23}$$

Alternatively, if $\hat{\lambda}^{[k]*}$ is available for $k = (k^{'}+1), \ldots, K$, we place an inverse-gamma prior on the study-specific $\phi^{[k]2}$ parameters, such that:

$$\phi^{[k]2} \sim \text{Inv-Gamma}(\zeta_1, \zeta_2); \qquad \text{and} \quad \zeta_1 \sim \text{Exp}(\delta), \quad \zeta_2 \sim \text{Exp}(\delta), \tag{24}$$

for $k = (k^{'} + 1), \ldots, K$; 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, 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 $\delta = 0.1$.

For a scenario in which $k^{'} = 0$, things are not so straightforward. Without any "gold standard" studies and without any specific knowledge regarding the degree of measurement error for each of the K contributing studies, the parameters of interest may not be identifiable. With this in mind, before moving on, we 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.

2.2.1 | Issues of identifiability

Our logic is based on similar arguments for partial identifiability considered in Campbell et al. (2020)[38]. 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:

$$I_k(\underline{\gamma}^{[k]}) = \left[\frac{\beta^{*[k]}}{\underline{\gamma}^{[k]}}, \beta^{*[k]}\right]. \tag{25}$$

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[39]. 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:

$$I(\underline{\gamma}, \bar{\tau}) = \left\{\theta : \tau \leq \bar{\tau}, \beta^{[k]} \in I_k(\underline{\gamma}^{[k]}), \forall k \in \{1, \ldots, K\}\right\}. \tag{26}$$

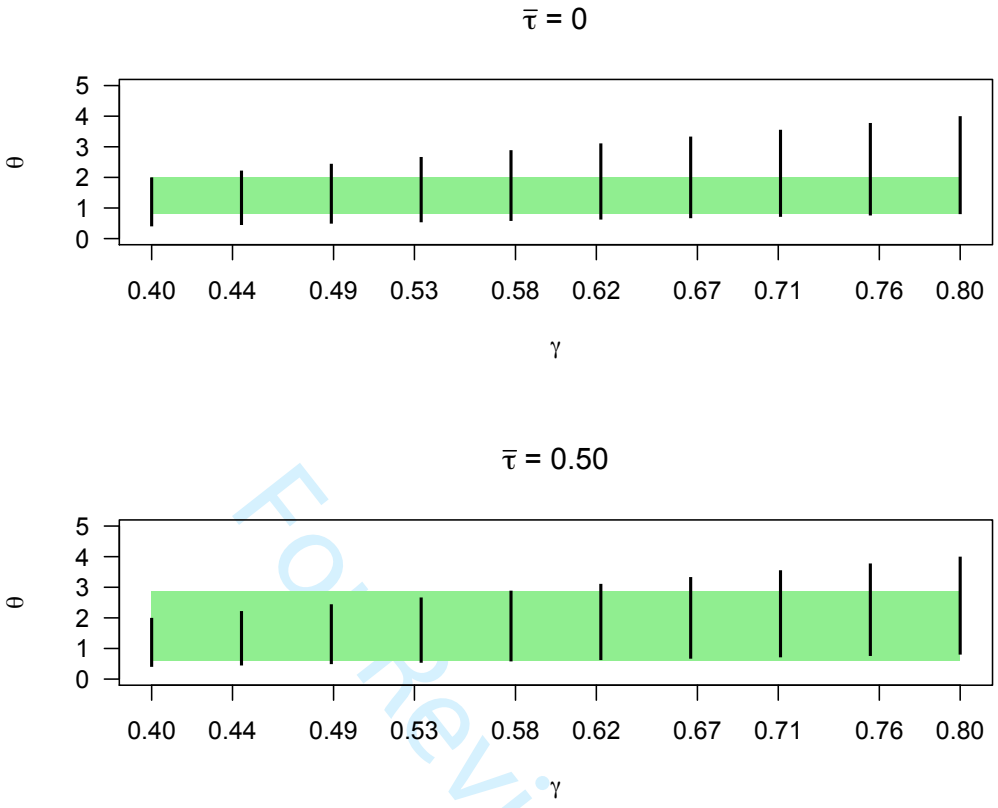

FIGURE 2 Black vertical lines correspond to jurisdiction-specific identification intervals and the green rectangle corresponds to the global identification interval. The upper panel corresponds to assumption of $\bar{\tau} = 0$ such that the global identification interval is simply the intersection of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

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 (26) 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 (26) 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 $k = 1, \ldots, 10$. Suppose the observed $\hat{\beta}^{[k]*}$ values for these studies lie equally spaced between 0.40 and 0.80, (since the unknown $\gamma^{[k]}$ 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 by the black vertical lines 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 (26) 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 much 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.

2.2.2 | Example: the NELS88* dataset

Returning now to our example with the NELS88 dataset, we illustrate the impact of measurement error by intentionally adding non-differential measurement error to the data as described in equation (10) 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 are "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.

We will fit both the univariate and bivariate *BMEMA* models. For the univariate model, we suppose that data for $\lambda^{[k]*}$s are unavailable and place a uniform prior on the $\gamma^{[k]}$ parameters as in (23). We also fit an additional bivariate *BMEMA* model with $\delta = 0.5$ instead of $\delta = 0.1$ (for all analyses where $K \neq k'$) to see how sensitive results may be to the chosen priors.

With the NELS88* dataset, we have that $\mathrm{E}(\gamma^{[k]}) = 0.63$ and $\mathrm{Var}(\gamma^{[k]}) = 0.15$. Based on equation (17), we have that $\theta^* = 0.66 \times 0.57 = 0.36$; and based on equation (18), 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 (17) and (18) (see Table 3 line 1): when measurement error is added to the data, estimates for θ are biased downwards, while estimates for τ are biased upwards.

For the NELS88* data, $\mathrm{cor}(\hat{\beta}^{*[k]}, \hat{\lambda}^{*[k]}) = -0.66$, whereas for the NELS88 data, $\mathrm{cor}(\hat{\beta}^{[k]}, \hat{\lambda}^{[k]}) = -0.02$. 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 estimates of $\hat{\theta} = 0.56$ and $\hat{\theta} = 0.57$ obtained with the univariate and bivariate *BayesMA* models respectively, with the "clean" NELS88 dataset (see Table 3, line 2) serve as approximate targets. For the NELS88* dataset, the bivariate *BMEMA* model, with $\delta = 0.1$, and with $k' = 5$, obtains an estimate of $\hat{\theta} = 0.58$, with 95% credible interval of $CI(\theta)_{95\%} = [0.48, 0.68]$(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. With the alternative prior specified by $\delta = 0.5$, the bivariate model obtains an estimate of $\hat{\theta} = 0.52$. The univariate model obtains a similar estimate of $\hat{\theta} = 0.53$, with 95% credible interval of $CI(\theta)_{95\%} = [0.41, 0.65]$.

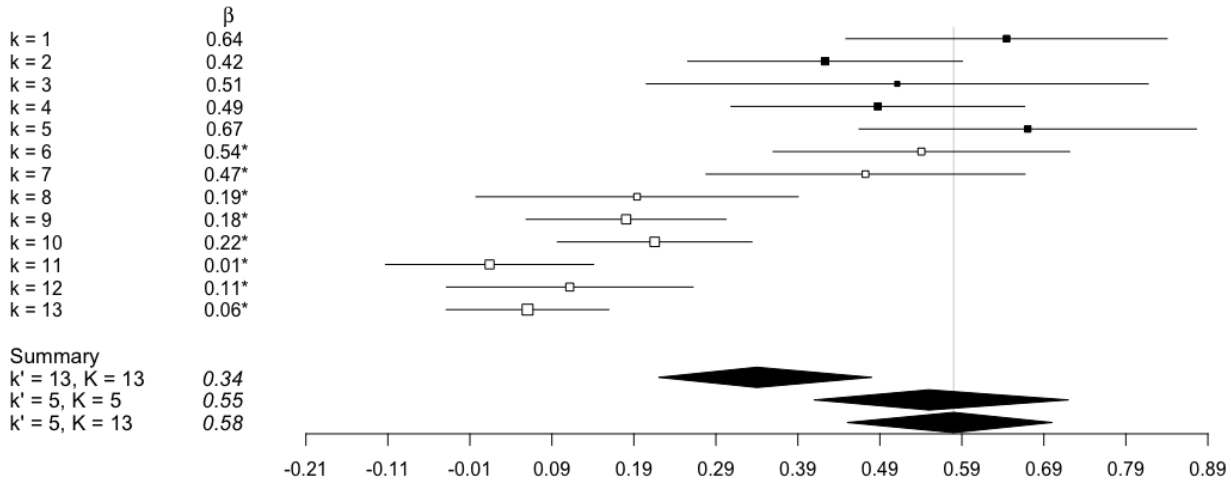

FIGURE 3 Forest plot for the meta-analysis of the NELS88* aggregate data, with three summaries corresponding to line 1 (in which all studies are assumed to be measurement error-free;$k' = 13$, $K = 13$), line 4 (in which only the 5 studies known to be measurement error-free are included in the analysis; $k' = 5$, $K = 5$), and line 3 (in which the 5 studies known to be measurement error-free are included in the analysis along with the 8 studies comprised by measurement error; $k' = 5$, $K = 13$) 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 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 (i.e., the 2.5% and 97.5% quantiles) for these estimates.

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 $\hat{\theta} = 0.55$ with a notably wider credible interval of $CI(\theta)_{95\%} = [0.41, 0.72]$ (identical results with the both the univariate and bivariate models). This suggests that the data from the additional eight schools, while compromised by measurement error, may provide "added value" and sharpen our inference.

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* models, with $k' = 0$, obtain estimates of θ much too high: $\hat{\theta} = 0.69$ (for the univariate), 0.90 (for the bivariate with $\delta = 0.1$), and 0.71 (for the bivariate with $\delta = 0.5$) (see Table 3, line 5). These three estimates are quite different and this suggests that, when there are no known "gold standard studies" to "anchor" the estimates, the chosen priors may yield significant leverage.

Finally, note that if the *BMEMA* model is fit to the original NELS88 data, we will end up slightly overestimating the θ parameter. For the bivariate model (with $\delta = 0.1$), we obtain $\hat{\theta} = 0.58$, with $k' = 5$, and $\hat{\theta} = 0.61$, 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

We can generalize the *BayesMA* models described in Section 2 (in which each study can be summarized as a simple linear regression) to a general case where each study can be summarized as a multivariable linear regression. Note that, while this generalization is of theoretical interest, in practice it may be unlikely to have multiple different studies provide results from exactly the same regression model. Should different studies adjust for different subsets of covariates, pooling their coefficients together in a meta-analysis may not be appropriate.

We denote $X_{j\times}^{[k]}$ 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 $X_{\times q}^{[k]}$ 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: $X_{j\times}^{[k]}$, 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 $X_{\times,-1}^{[k]}$ to denote the design matrix excluding the column of 1s.

Suppose data from each study can be summarized as a multivariable linear regression model such that, for $k = 1, \ldots, K$, and j in $1, \ldots, n^{[k]}$:

$$Y_j^{[k]} | X_{j\times}^{[k]} \sim \mathcal{N}\left(X_{j\times}^{[k]} \beta^{[k]}, \sigma^{[k]2}\right), \tag{27}$$

where $\beta^{[k]} = (\beta_0^{[k]}, \beta_1^{[k]}, \ldots, \beta_Q^{[k]})'$ is the column-vector of regression coefficients. We assume that each study has its own exposure distribution governed by a multivariate Normal distribution:

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

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:

$$\beta^{[k]} | \theta, \mathrm{T} \sim \mathcal{MVN}(\theta, \mathrm{T}), \tag{29}$$

for all $k = 1, \ldots, K$; where θ is a $(Q+1)$-length vector, and T is a $(Q+1) \times (Q+1)$ covariance matrix. If we are able to assume that the regression coefficients are *a priori* independent, then T will be a diagonal matrix.

If individual participant data (IPD) are available, the model will have many moving parts. 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 $k = 1, \ldots, K$), Bayes' theorem states that:

$$\begin{aligned} p(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda | \text{data}) &\propto p(\text{data} | (\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda)) p(\theta, \mathrm{T}, \beta, \sigma, \mu, \Lambda) \\ &= \prod_{k=1}^{K} \Big(\prod_{j=1}^{n^{[k]}} \Big\{ p(Y_j^{[k]} | X_{j\times}^{[k]}, \beta^{[k]}, \sigma^{[k]}) p(X_{j\times}^{[k]} | \mu^{[k]}, \Lambda^{[k]}) \Big\} \\ &\qquad \times p(\beta^{[k]} | \theta, \mathrm{T}) p(\mu^{[k]}) p(\Lambda^{[k]}) p(\sigma^{[k]}) \Big) p(\theta) p(\mathrm{T}). \end{aligned} \tag{30}$$

If IPD are not available, and only aggregate data are available, an simpler two-part model can be defined whereby:

$$\hat{\beta}^{[k]}|\beta^{[k]} \sim \mathcal{MVN}(\beta^{[k]}, COV^{[k]}), \text{ where } COV^{[k]} = (X^{[k]T}X^{[k]})^{-1}\sigma^{[k]2}; \text{ and} \tag{31}$$

$$\beta^{[k]}|\theta, \mathrm{T} \sim \mathcal{MVN}(\theta, \mathrm{T}), \tag{32}$$

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

3.2 | Adjusting for non-differential measurement error

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

$$X^{*[k]}_{j,-1} = \mathcal{MVN}(X^{[k]}_{j,-1}, \Phi^{[k]}), \tag{33}$$

for j in $1, \ldots, n^{[k]}$ and for $k = 1, \ldots, K$; where $\Phi^{[k]}$ is a $Q \times Q$ covariance matrix. Note that, equation (33) is analogous to equation (10) 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)[40] shows that, even if the q-th covariate, $X_{\times q}$, is measured without error, one may still incorrectly reject the null hypothesis of $\beta^{[k]}_q = 0$, 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[41].

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[42], we can frame a flexible multivariate *BMEMA* model by assuming that the covariates are multivariate normal (but this could be modified as needed) and defining the following three-part model structure:

$$X^{[k]}_{j,-1} \sim \mathcal{MVN}\left(\mu^{[k]}, \Lambda^{[k]}\right), \tag{34}$$

$$X^{[k]*}_{j,-1}|X^{[k]}_{j,-1} \sim \mathcal{MVN}\left(X^{[k]}_{j,-1}, \Phi^{[k]}\right), \tag{35}$$

$$\text{and} \quad Y^{[k]}_j|X^{[k]}_{j\times} \sim \mathcal{N}\left(X^{[k]}_{j\times}\beta^{[k]}, \sigma^{[k]2}\right), \tag{36}$$

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

As a prior for Φ, the inverse-Wishart distribution with covariance matrix $2\zeta_2 \times I_{[Q]}$ and $2\zeta_1$ degrees of freedom is a multivariate generalization of the prior we considered in Section 2.2 for $\phi^{k]}$. Consider:

$$\Phi^{[k]} \sim \text{Inv-Wishart}(2\zeta_2 \times I_{[Q]}, 2\zeta_1); \qquad \zeta_1 \sim \text{Exp}(\delta) \quad \text{and:} \quad \zeta_2 \sim \text{Exp}(\delta), \tag{37}$$

for $k = 1, \ldots, K$; where $\zeta_1 > (Q/2)$ and $\zeta_2 > 0$. 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 $\mathrm{Var}(\Phi^{[k]}) = \zeta_2^2/((\zeta_1 - 1)^2(\zeta_1 - 2))$.

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 (33) 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 2.2.2, 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 $\Phi^{[k]}_{1,1}$ equal to between 4 and 6; and set $\Phi^{[k]}_{2,2}$ 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 $\Phi^{[k]}_{1,1}$ and $\Phi^{[k]}_{2,2}$ for reference.

For this simple example analysis, we specify multivariate normal priors for $\mu^{[k]}$ and θ: $\mu^{[k]} \sim \mathcal{MVN}(0_{[Q]}, 100 \times I_{[Q]})$, for $k = 1, \ldots, K$; 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. Also, we will assume that the regression coefficients are *a priori* independent and specify half-Cauchy priors for $\sigma^{[k]}$ and the $Q + 1$ diagonal elements of T:$\sigma^{[k]} \sim$ half-Cauchy$(0, 2)$, for $k = 1, \ldots, K$; and $\sqrt{\mathrm{T}_{qq}} \sim$ half-Cauchy$(0, 2)$, for q in $1, \ldots, Q + 1$. Finally, for $\Lambda^{[k]}$ and Φ, we specify inverse-Wishart priors such that: $\Lambda^{[k]} \sim$ Inv-Wishart$(I_{[Q]}, Q)$, for $k = 1, \ldots, K$; and $\Phi^{[k]} \sim$ Inv-Wishart$(2\zeta_2 \times I_{[Q]}, 2\zeta_1)$; $\zeta_1 \sim$ Exp(0.1) and: $\zeta_2 \sim$ Exp(0.1).

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 2.2.1 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)[43]).

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[44,45]. However, one should always correct for sources of bias if this is possible. Currently, tools to correct for measurement error in a meta-analysis are not available and as a consequence, researchers are left to simply list measurement error as a study limitation (e.g. Wu et al. (2016)[46]: "this study has several limitations [...] none of the studies corrected for measurement error."; and Merino et al. (2009)[47]: "although every effort was made to maximize the validity of the

study, minimize bias, and incorporate heterogeneity and uncertainty, the estimated hazard ratios of dietary components could be affected by measurement error").

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 for continuous outcome data. 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 Bun et al. (2020)[48]. While certain studies may be biased, they may still provide value if one can appropriately account and adjust for the bias. Bayesian inference is well suited to the task.

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[49]. 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. That being said, the proposed *BMEMA* model can only do so much. We stress that the *BMEMA* model will assume that all differences between the unbiased studies and the biased studies are due to measurement error when in reality there may be other systematic differences at play. Furthermore, the model, as it is currently described, can only deal with non-differential measurement. In practice, there may be substantial bias due to differential measurement error[50]. Differential measurement error is a major concern, for instance, in retrospective studies; see White (2003)[51]. Future research should investigate how to address these difficult issues. In addition, future research should generalize the proposed *BMEMA* model for binary and time-to-event outcomes and could also extend the model to network meta-analysis.

On a final note, beyond the bias caused by measurement error, a meta-analysis of observational studies should, ideally, also take into account other (potentially bigger) biases, e.g.: publication bias and bias due to unmeasured confounding [52,53]). 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. Myers RH, Myers RH. *Classical and modern regression with applications*. 2. Duxbury press Belmont, CA . 1990.

18. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in medicine* 2001; 20(6): 825–840.

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

20. Crouch GI. A meta-analysis of tourism demand. *Annals of tourism research* 1995; 22(1): 103–118.

21. Crouch GI. Demand elasticities in international marketing: A meta-analytical application to tourism. *Journal of Business Research* 1996; 36(2): 117–136.

22. Lau RR, Sigelman L, Heldman C, Babbitt P. The effects of negative political advertisements: A meta-analytic assessment. *American Political Science Review* 1999: 851–875.

23. 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.

24. Schmid CH, Stijnen T, White I. *Handbook of Meta-analysis (Chapter 10)*. CRC Press . 2020.

25. 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.

26. Papadimitropoulou K, Stijnen T, Riley RD, Dekkers OM, Cessie lS. Meta-analysis of continuous outcomes: Using pseudo IPD created from aggregate data to adjust for baseline imbalance and assess treatment-by-baseline modification. *Research synthesis methods* 2020; 11(6): 780–794.

27. 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.

28. 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

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

30. Hossain S, Gustafson P. Bayesian adjustment for covariate measurement errors: a flexible parametric approach. *Statistics in Medicine* 2009; 28(11): 1580–1600.

31. Röver C. Bayesian random-effects meta-analysis using the bayesmeta R package. *arXiv preprint arXiv:1711.08683* 2017.

32. 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.

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

34. 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.

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

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

37. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *Bmj* 2010; 340.

38. Campbell H, Valpine dP, Maxwell L, et al. Bayesian adjustment for preferential testing in estimating the COVID-19 infection fatality rate: Theory and methods. *arXiv preprint arXiv:2005.08459* 2020.

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

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

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

42. 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.

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

44. 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.

45. 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.

46. Wu L, Sun D. Meta-analysis of milk consumption and the risk of cognitive disorders. *Nutrients* 2016; 8(12): 824.

47. Merino J, Guasch-Ferré M, Ellervik C, et al. Quality of dietary fat and genetic risk of type 2 diabetes: individual participant data meta-analysis. *Bmj* 2019; 366.

48. Bun RS, Scheer J, Guillo S, Tubach F, Dechartres A. Meta-analyses frequently pooled different study types together: a meta-epidemiological study. *Journal of Clinical Epidemiology* 2020; 118: 18–28.

49. 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.

50. Smeden vM, Lash TL, Groenwold RH. Reflection on modern methods: five myths about measurement error in epidemiological research. *International journal of epidemiology* 2020; 49(1): 338–347.

51. White E. Design and interpretation of studies of differential exposure measurement error. *American Journal of Epidemiology* 2003; 157(5): 380–387.

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

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

54. Weingessel A, Turlach BA. quadprog: Functions to solve Quadratic Programming Problems.. *URL https://CRAN.R-project.org/package=quadprog*. 2013: 1–5.

5 | APPENDIX

Consider the evaluation of (26) for $\overline{\tau} > 0$, i.e., where a limited heterogeneity in the $\beta^{[k]}$ parameters 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 (e.g., quadprog[54]) to minimize the quadratic function $Var(\beta^{[k]})$, subject to the equality constraint $\theta = x$ and the inequality constraints which restrict $\beta^{[k]}$ 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.

□

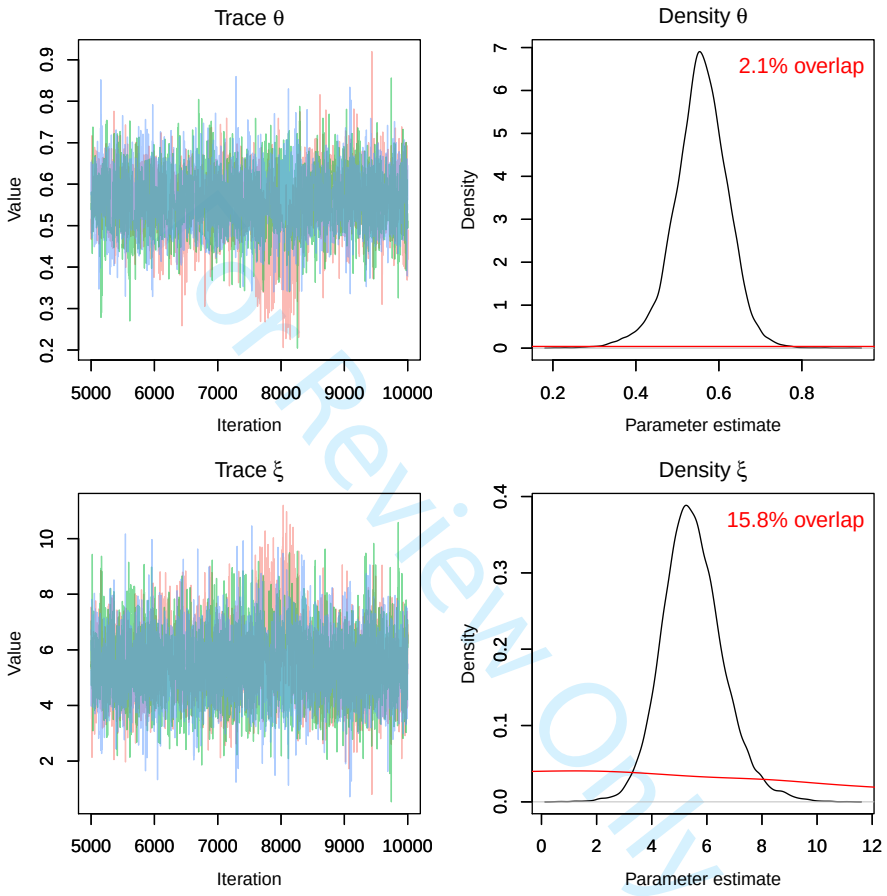

FIGURE 4 Diagnostic plots for parameters θ and ξ, for the MCMC simulation of the univariate *BMEMA* model with $k' = 5$ and $\delta = 0.1$ (this corresponds to the results in Table 3, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding 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

	sample size	data required for univariate model		data required for bivariate model				
k	$n^{[k]}$	$\hat{\beta}^{[k]}$	$se(\hat{\beta}^{[k]})$	$\hat{\alpha}^{[k]}$	$se(\hat{\alpha}^{[k]})$	$\hat{\beta}^{[k]}$	$se(\hat{\beta}^{[k]})$	$\hat{\sigma}^{[k]2}$
1	45	0.64	0.10	4.99	1.53	0.64	0.10	12.28
2	64	0.42	0.09	7.91	1.16	0.42	0.09	19.50
3	47	0.51	0.16	8.46	2.71	0.51	0.16	11.48
4	45	0.49	0.09	2.95	1.19	0.49	0.09	15.71
5	45	0.67	0.10	1.75	1.07	0.67	0.10	15.15
6	59	0.54	0.09	5.53	1.24	0.54	0.09	19.45
7	56	0.56	0.12	5.48	1.77	0.56	0.12	17.31
8	45	0.39	0.16	9.69	2.78	0.39	0.16	13.51
9	51	0.61	0.12	6.60	2.15	0.61	0.12	6.93
10	67	0.67	0.11	6.33	1.84	0.67	0.11	13.88
11	48	0.57	0.10	4.03	1.23	0.57	0.10	11.62
12	45	0.62	0.13	6.39	1.85	0.62	0.13	20.55
13	47	0.50	0.13	4.02	1.85	0.50	0.13	15.16

TABLE 1 The NELS88 data- Aggregate data required for the univariate and bivariate meta-analyses in Section 2.1.1, as obtained from the NELS88 dataset.

The NELS88* aggregate data

	sample size	measurement error	attenuation factor	data required for univariate model		data required for bivariate model				
k	$n^{[k]}$	$\phi^{[k]}$	$\gamma^{[k]}$	$\hat{\beta}^{[k]*}$	$se(\hat{\beta}^{[k]*})$	$\hat{\alpha}^{[k]*}$	$se(\hat{\alpha}^{[k]*})$	$\hat{\beta}^{[k]*}$	$se(\hat{\beta}^{[k]*})$	$\hat{\sigma}^{[k]2*}$
1	45	0.00	1.00	0.64	0.10	4.99	1.53	0.64	0.10	12.28
2	64	0.00	1.00	0.42	0.09	7.91	1.16	0.42	0.09	19.50
3	47	0.00	1.00	0.51	0.16	8.46	2.71	0.51	0.16	11.48
4	45	0.00	1.00	0.49	0.09	2.95	1.19	0.49	0.09	15.71
5	45	0.00	1.00	0.67	0.10	1.75	1.07	0.67	0.10	15.15
6	59	1.00	0.98	0.56	0.09	5.33	1.25	0.56	0.09	19.19
7	56	2.57	0.78	0.49	0.12	6.70	1.76	0.49	0.12	18.80
8	45	4.14	0.41	0.17	0.12	13.19	2.11	0.17	0.12	14.57
9	51	5.71	0.23	0.17	0.08	14.50	1.46	0.17	0.08	9.74
10	67	7.29	0.26	0.15	0.06	15.12	1.20	0.15	0.06	20.62
11	48	8.86	0.25	0.12	0.06	9.24	0.95	0.12	0.06	18.71
12	45	10.43	0.21	0.08	0.06	13.74	1.21	0.08	0.06	30.94
13	47	12.00	0.12	0.10	0.05	9.84	0.82	0.10	0.05	18.59

TABLE 2 The NELS88* data- Aggregate data obtained from the NELS88* dataset required for the *BMEMA* univariate and bivariate models in 2.2.2. Values of $\phi^{[k]}$ which correspond to the amount of measurement error intentionally added to each study and values of the attenuation factor, $\gamma^{[k]}$, are also listed for reference.

	dataset	model	K	k'	θ	$CI(\theta)_{95\%}$	τ
line 1.	NELS88*	univariate	13	13	0.34	0.22 , 0.48	0.21
		bivariate ($\delta = 0.1$)	13	13	0.34	0.23 , 0.49	0.2
line 2.	NELS88	univariate	13	13	0.56	0.49 , 0.62	0.04
		bivariate ($\delta = 0.1$)	13	13	0.57	0.51 , 0.64	0.04
line 3.	NELS88*	univariate	13	5	0.53	0.41 , 0.65	0.08
		bivariate ($\delta = 0.1$)	13	5	0.58	0.48 , 0.68	0.05
		bivariate ($\delta = 0.5$)	13	5	0.52	0.34 , 0.64	0.10
line 4.	NELS88$^{1:5}$	univariate	5	5	0.55	0.41 , 0.72	0.08
		bivariate ($\delta = 0.1$)	5	5	0.55	0.41 , 0.72	0.08
line 5.	NELS88*	univariate	13	0	0.69	0.50 , 1.05	0.13
		bivariate ($\delta = 0.1$)	13	0	0.90	0.60 , 1.21	0.10
		bivariate ($\delta = 0.5$)	13	0	0.71	0.48 , 0.99	0.09
line 6.	NELS88	univariate	13	5	0.63	0.53 , 0.78	0.09
		bivariate ($\delta = 0.1$)	13	5	0.58	0.51 , 0.66	0.04
		bivariate ($\delta = 0.5$)	13	5	0.58	0.51 , 0.66	0.04
line 7.	NELS88	univariate	13	0	0.81	0.65 , 1.16	0.08
		bivariate ($\delta = 0.1$)	13	0	0.61	0.52 , 0.78	0.05
		bivariate ($\delta = 0.5$)	13	0	0.61	0.52 , 0.78	0.05

TABLE 3 The data analysis results obtained - posterior medians with 95% equal-tailed credible intervals.

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

TABLE 4 A. Aggregate data obtained from the NELS88 dataset and the NELS88* dataset for comparison. Values for $\Phi_{1,1}^{[k]}$ and $\Phi_{2,2}^{[k]}$ which correspond to the amount of measurement error intentionally added to the reading test score and the mathematics test score, respectively, for the k-th study 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.

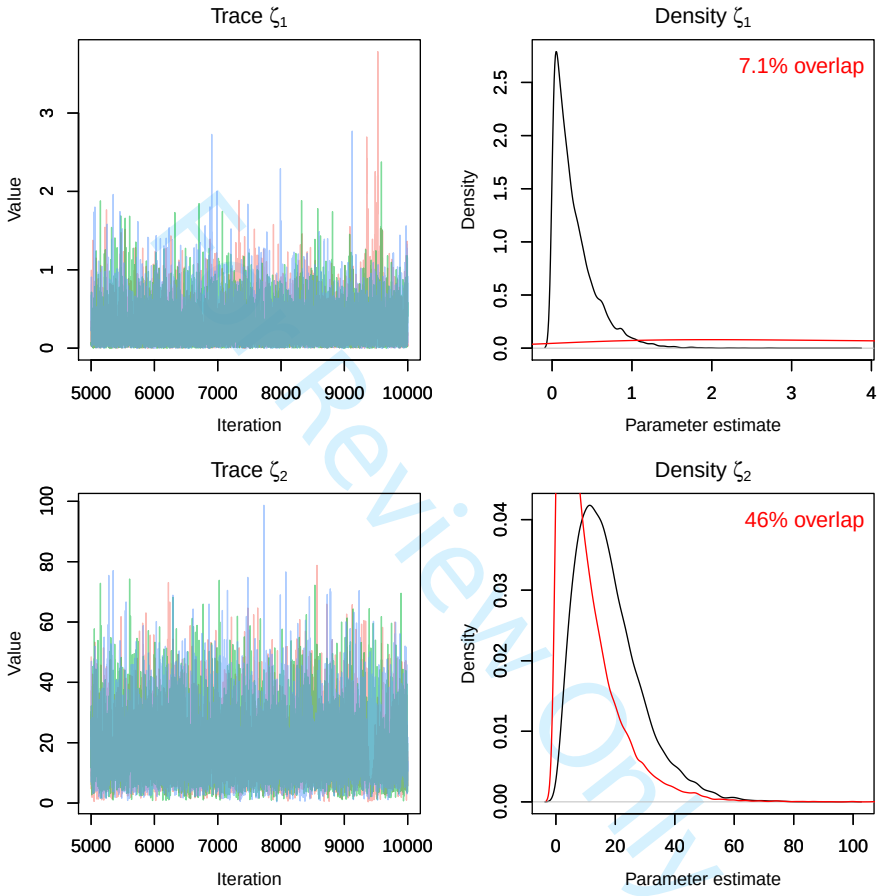

FIGURE 5 Diagnostic plots for parameters ζ_1 and ζ_2, for the MCMC simulation of the univariate *BMEMA* model with $k' = 5$ and $\delta = 0.1$ (this corresponds to the results in Table 3, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the 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).

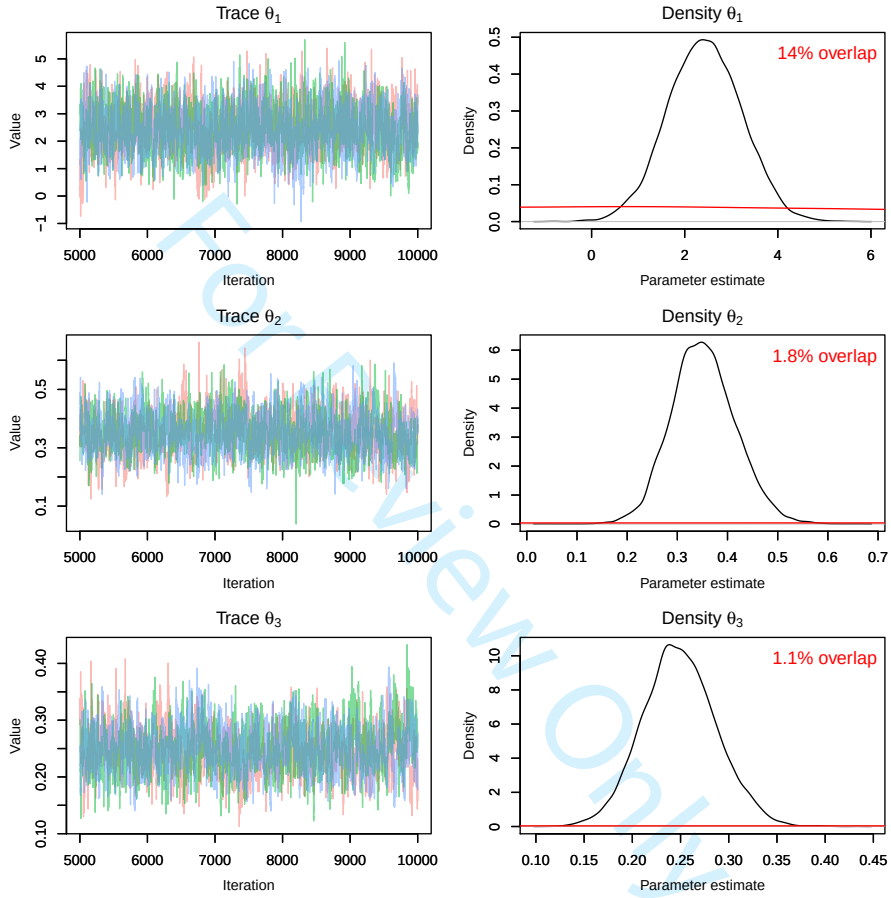

FIGURE 6 Diagnostic plots for the θ parameter, for the MCMC simulation of the multivariate *BMEMA* model with $k' = 5$ (this corresponds to the results in Table 4-B, line 3). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior 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).

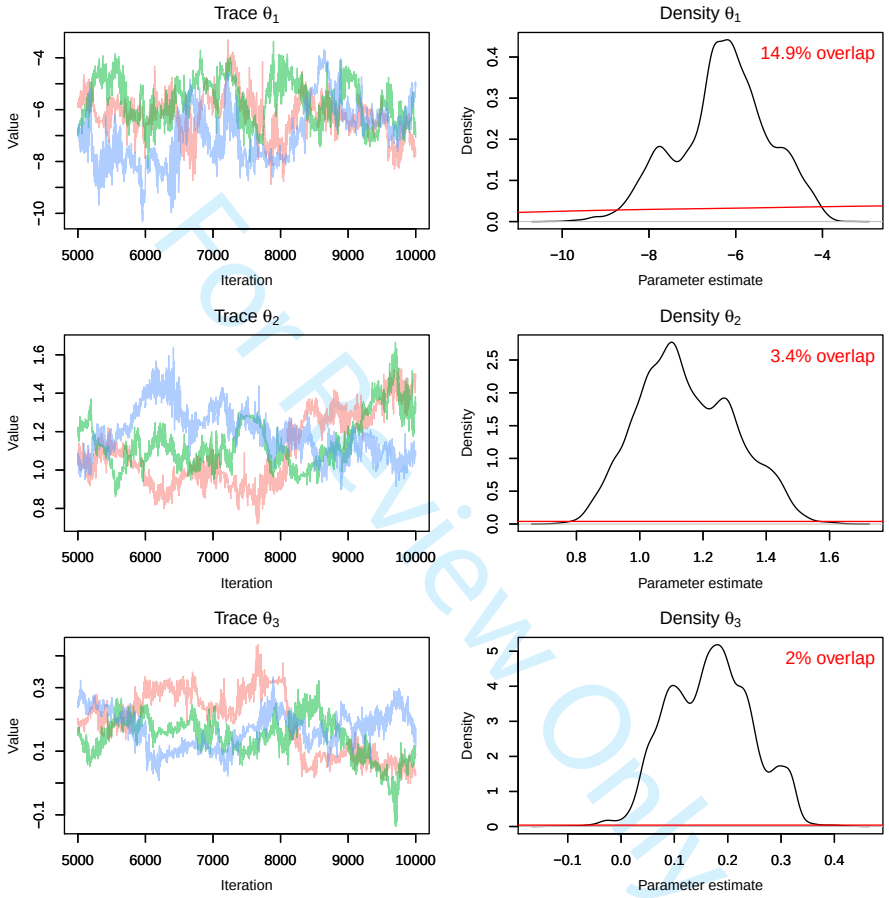

FIGURE 7 Diagnostic plots for the θ parameter, for the MCMC simulation of the multivariate *BMEMA* model with $k' = 0$ (this corresponds to the results in Table 4-B, line 5). The left panels report trace plots from the posterior to check convergence. The right panels report the corresponding posterior 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).