

# **Bayesian bivariate meta-analysis of correlated effects: impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences**

## **Author affiliations**

Danielle L Burke<sup>1\*</sup>, Sylwia Bujkiewicz<sup>2</sup> and Richard D Riley<sup>1</sup>

\* Correspondence: [d.burke@keele.ac.uk](mailto:d.burke@keele.ac.uk)

<sup>1</sup> Research Institute for Primary Care and Health Sciences, Keele University, Staffordshire, UK

<sup>2</sup> Biostatistics Group, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK

## **Acknowledgements**

Richard D Riley was supported by funding from a multivariate meta-analysis grant from the MRC Methodology Research Programme (grant reference number: MR/J013595/1). Danielle Burke was partly supported by funding from the MRC Midlands Hub for Trials Methodology Research, at the University of Birmingham (Medical Research Council Grant ID G0800808). Sylwia Bujkiewicz is supported by the Medical Research Council (MRC) Methodology Research Programme [New Investigator Research Grant MR/L009854/1]

We would like to thank Dan Jackson and Ian White for helpful feedback during the initial stages of the project, and also two anonymous reviewers for their constructive feedback on how to improve the article.

## **Author contributions**

RR and SB conceived the research idea. DB undertook all the simulation analyses under the supervision of RR and feedback from SB. DB drafted the paper and revised following comments from RR and SB.

## Abstract

Multivariate random-effects meta-analysis allows the joint synthesis of correlated results from multiple studies, for example for multiple outcomes or multiple treatment groups. In a Bayesian univariate meta-analysis of one endpoint, the importance of specifying a sensible prior distribution for the between-study variance is well-understood. However, in multivariate meta-analysis there is little guidance about the choice of prior distributions for the variances or, crucially, the between-study correlation,  $\rho_B$ ; for the latter, researchers often use a Uniform(-1,1) distribution assuming it is vague.

In this article, an extensive simulation study and a real illustrative example is used to examine the impact of various (realistically) vague prior distributions for  $\rho_B$  and the between-study variances within a Bayesian bivariate random-effects meta-analysis of two correlated treatment effects. A range of diverse scenarios are considered, including complete and missing data, to examine the impact of the prior distributions on posterior results (for treatment effect and between-study correlation), amount of borrowing of strength, and joint predictive distributions of treatment effectiveness in new studies.

Two key recommendations are identified to improve the robustness of multivariate meta-analysis results. Firstly, the routine use of a Uniform(-1,1) prior distribution for  $\rho_B$  should be avoided, if possible, as it is not necessarily vague. Instead, researchers should identify a sensible prior distribution, for example by restricting values to be positive or negative as indicated by prior knowledge. Secondly, it remains critical to use sensible (e.g. empirically-based) prior distributions for the between-study variances, as an inappropriate choice can adversely impact the posterior distribution for  $\rho_B$ , which may

then adversely affect inferences such as joint predictive probabilities. These recommendations are especially important with a small number of studies and missing data.

## **Keywords**

Bayes, bivariate/multivariate meta-analysis, correlation, prior distributions, simulation study, multiple outcomes.

## **Short title**

Bayesian bivariate meta-analysis and choice of prior distributions

# 1 Introduction

The multivariate meta-analysis approach has been advocated to jointly synthesise multiple correlated results from related research studies (1, 2). For example, in a meta-analysis of multiple outcomes, a cancer patient's overall survival time is likely to be correlated with their progression-free survival time, and therefore treatment effect estimates for both outcomes are likely correlated within a study. Similarly, in a network meta-analysis of multiple treatment groups, the treatment effect for A vs B is likely correlated with that for A vs C. Compared to separate univariate meta-analyses, the multivariate approach utilises such correlation to gain additional information toward the estimation of summary meta-analysis results (3, 4). This is especially advantageous when there are missing effect estimates in some studies (such as missing direct comparisons in network meta-analysis) and when there is potential outcome reporting bias (5, 6), as the correlation can lead to more precise inferences and/or a reduction in bias (2), which has been referred to as 'borrowing of strength' (7).

The Bayesian framework for multivariate meta-analysis is a natural way to account for all parameter uncertainty, to make predictions regarding the possible effects in new studies, and to derive joint probability estimates regarding the multiple effects of interest. However, it requires the specification of prior distributions for all unknown parameters, which may be considered a disadvantage when genuine prior information does not exist. A previous simulation study of Bayesian univariate meta-analyses (8) found that the pooled effect estimates can be particularly sensitive to the choice of prior distribution for the between-study variance, even when seemingly 'vague' prior distributions are specified. To address this, previous work has utilised a large collection

of existing meta-analyses to generate empirical prior distributions for the unknown between-study variance in a new univariate meta-analysis of intervention effects for continuous outcomes (9) and binary outcomes (10, 11), across a wide-range of healthcare settings, such as where the outcome of interest is all-cause mortality.

In addition to prior distributions for the between-study variances, a multivariate meta-analysis also requires prior distribution(s) for the between-study correlation(s). One might address this using the conjugate prior distribution for the entire between-study variance-covariance matrix, which is the inverse-Wishart prior distribution, and this has been used by previous authors, such as bivariate meta-analyses of test accuracy studies (12-14). However, others argue it is preferable to place separate prior distributions on each component of the between-study variance-covariance matrix because the Wishart prior distribution can be very influential toward the posterior estimates of the between-study variances (14-17); the Wishart distribution is a generalisation of the gamma distribution, which is known to be influential in univariate meta-analysis when used as a prior distribution for the between-study variances, especially when the true between-study variances are close to zero (8). Separation of the between-study variance-covariance matrix also allows more flexibility in the choice of prior distributions for each component, for instance if genuine prior information was available for some, but not all, of the components.

In situations where separate prior distributions are placed on the between-study variances and correlations, an unanswered question remains: what is the impact of the choice of prior distributions for the between-study correlations and variances in a multivariate meta-analysis, especially in situations where little or no prior information is

available? Appropriate estimation of the between-study variance-covariance matrix is important to making valid inferences, and thus undesired influence of prior distributions is unwanted when prior information is unavailable. For instance, appropriate estimation of the between-study correlation is desired because it dictates the magnitude of the borrowing of strength (1), and is therefore potentially influential toward pooled effects, credible intervals and prediction intervals; it is also pivotal when estimating functions of the pooled estimates, or when deriving joint probability estimates (such as the probability that the treatment is effective for all outcomes). However, in our experience, most previous Bayesian applications of multivariate meta-analysis (including some of our own) adopt a  $U(-1,1)$  prior distribution for the between-study correlation, but do not conduct sensitivity analyses to check whether it is appropriate or influential (1, 17, 18).

The aim of this paper is to examine the impact of seemingly vague and realistically vague prior distributions for the between-study correlations and variances in a bivariate meta-analysis, to extend previous work in the univariate setting (8). Real application and an extensive simulation study are described, focusing on a Bayesian bivariate meta-analysis of treatment effects for two correlated outcomes, and investigating how the choice of prior distributions impacts upon posterior estimates of the pooled treatment effects and between-study covariance matrix, the accuracy of 95% credible and prediction intervals, and joint probabilistic inferences. Both complete and missing outcome data situations are examined, and the impact on the amount of borrowing of strength (that is, the change in pooled results and credible intervals from univariate to bivariate analysis) is also considered.

The remainder of this paper is structured as follows. Section 2 introduces the bivariate random-effects meta-analysis model and potential prior distributions for the between-study variances and correlation. Section 3 describes the methods and results of the simulation study. The key findings are then illustrated in the context of a real meta-analysis dataset in Section 4. Section 5 concludes with some discussion and recommendations.

## 2 General model for bivariate random-effects meta-analysis

This section summarises the general framework for bivariate meta-analysis, and introduces possible prior distributions for the between-study variances and correlation. We focus on the use of bivariate meta-analysis for two correlated outcomes, but the issues remain similarly pertinent in other situations of correlated effects, such as multiple treatment groups (network meta-analysis) and multiple performance statistics (such as sensitivity and specificity) (19, 20).

### 2.1 Model specification

Suppose that each of  $i=1$  to  $n$  studies examines an effect of interest (such as a treatment effect) for two outcomes ( $j=1,2$ ), such as systolic and diastolic blood pressure, or overall and progression-free survival. Let each study provide the estimated effects,  $Y_{i1}$  and  $Y_{i2}$ , and their associated standard errors,  $s_{i1}$  and  $s_{i2}$ , where each  $Y_{ij}$  is an estimate of an underlying true value,  $\theta_{ij}$ , and these true values may vary between studies due to heterogeneity. Assuming the  $Y_{ij}$  and  $\theta_{ij}$  are drawn from a bivariate normal distribution,



and that the within-study variance-covariance matrix ( $\mathbf{S}_i$ ) is known, then the bivariate random-effects meta-analysis model can be specified as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \mathbf{S}_i \right\} \quad (1)$$

$$\mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & \rho_{W_i} s_{i1} s_{i2} \\ \rho_{W_i} s_{i1} s_{i2} & s_{i2}^2 \end{pmatrix}$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D} \right\}$$

$$\mathbf{D} = \begin{pmatrix} \tau_1^2 & \rho_B \tau_1 \tau_2 \\ \rho_B \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}$$

The true values ( $\theta_{ij}$ ) therefore have a mean value  $\beta_j$  (referred to as the ‘pooled’ effect for outcome  $j$ ) and between-study variance,  $\tau_j^2$ . The within-study covariance matrix,  $\mathbf{S}_i$ , contains the known within-study variances,  $s_{ij}^2$ , and within-study covariances,  $\rho_{W_i} s_{i1} s_{i2}$ , for each trial, where  $\rho_{W_i}$  represents the within-study correlation of  $Y_{i1}$  and  $Y_{i2}$ . The between-study covariance matrix,  $\mathbf{D}$ , contains the unknown between-study variances,  $\tau_j^2$ , and the unknown between-study correlation,  $\rho_B$ , of the  $\theta_{i1}$ s and  $\theta_{i2}$ s. Multivariate extensions to the bivariate model (1) follow naturally, though are more complex due to the increasing number of between-study variances and correlations that require estimation (2, 14, 21, 22).

### 2.1.1 Within-study and between-study correlation

Within-study and between-study correlation are two measures of correlation in a multivariate random-effects meta-analysis model. The within-study correlation is a measure of the association between the effect estimates in each study, and is caused by the same patients contributing correlated data toward both outcomes. Estimation of model (1) typically assumes these are known (just as the within-study variances are assumed known) (1), and for the purposes of this paper we also make this assumption. Authors such as Riley et al. and Trikalinos et al. detail how to derive within-study correlations when individual participant data are available (23, 24), but they can also be approximated using aggregate data in some other situations (25). Alternatively, it is possible to construct prior distributions from previous studies (21, 22).

The between-study correlation is a measure of how the true underlying effects are related across studies, and occurs because of between-study heterogeneity in, for example, the dosage of a drug or patient characteristics of the study populations, such as age. The between-study correlation is unknown and must be estimated in the meta-analysis model, alongside the between-study variances.

Both within- and between-study correlation can influence the amount of borrowing of strength in a bivariate meta-analysis (5, 7). Within-study correlations are more influential when the within-study variances are large relative to the between-study variance, whereas the between-study correlation is more influential when the between-study variances are large relative to the within-study variances. Further, accounting for such correlation is essential when an aim is to make joint inferences about the two effects of interest, such as the probability that they are both above a particular value.

## 2.2 Model estimation

In a frequentist framework, model (1) can be estimated by methods of moments or restricted maximum likelihood (2). Within a Bayesian framework, the likelihood pertaining to model (1) is combined with prior distributions for the unknown parameters of  $\beta_j$ ,  $\tau_j^2$  and  $\rho_B$ , and then posterior inferences are derived by sampling from the marginal posterior distributions using, for example, Markov chain Monte Carlo (MCMC) via Gibbs sampling. The convergence of parameters must be checked, which can be done visually using history and trace plots, and possible autocorrelation must be examined, which can be reduced by thinning the samples.

The prior distributions for the pooled effects ( $\beta_j$ ) are not evaluated and are given a vague  $N(0, 1000^2)$  prior distribution throughout. Here, the focus is on examining different choices of the prior distributions for  $\tau_j^2$  and, especially,  $\rho_B$ , and these are now discussed.

## 2.3 Choice of prior distribution for $\tau_j$

In univariate meta-analysis, the prior distribution for  $1/\tau^2$  was once commonly chosen to be the  $\text{Gamma}(\epsilon, \epsilon)$  distribution with the misperception that if  $\epsilon$  were very small (i.e. 0.001) then this distribution would be ‘vague’ (8). However, previous work by Lambert et al (8) (and more generally outside the meta-analysis field by Gelman (26)) demonstrated that the Gamma distribution is not appropriate, as posterior inferences for the between-study variance and pooled effects are sensitive to  $\epsilon$ . Here,  $\epsilon$  must be set to a reasonable value, or meta-analysts should rather use one of a number of different weakly informative prior distributions discussed by Lambert et al. and Gelman (8, 26).

These refer to distributions that are set up so that the information they provide is weak, but contain only realistic values for the variance. These include the half-Normal( $0, a$ ) distribution (27, 28), and the half- $t$  family of distributions, such as the half-Cauchy distribution (26). In particular, for the half-Normal( $0, a$ ) distribution, the value of  $a$  can be chosen to cover all realistic values of the between-study variance, for example as identified from other previous meta-analyses of the same outcome type in the same disease field.

The latter idea leads naturally to empirically-based prior distributions for the between-study variances (29). Indeed, previous work has used a large collection of existing meta-analyses to generate empirical prior distributions for the unknown between-study variance in a new univariate meta-analysis of intervention effects for continuous outcomes (9) and binary outcomes (10, 11), across a wide-range of healthcare settings, such as where the outcome of interest is all-cause mortality.

Here, in the setting of bivariate meta-analysis, we interrogate some inappropriate and sensible/weakly informative prior distributions for the between-study variances, to explore their impact on bivariate meta-analysis estimates and conclusions. In particular, in the simulation study (Section 3) two contrasting prior distributions for the between-study variances are compared: an inappropriate Gamma distribution and a more suitable truncated normal distribution that was suggested by Lambert et al (8). Then, in the illustrative example in Section 4, a relevant empirical prior distribution is chosen and compared to an inappropriate Gamma prior distribution.

We include an inappropriate Gamma distribution for  $1/\tau^2$  in both simulations and the example to highlight the danger of using this (or its extension, the Wishart distribution) as a prior distribution for the between-study variances in the context of bivariate meta-analysis applications, with particular emphasis on how it can adversely affect the posterior distribution for  $\rho_B$  and the amount of borrowing of strength toward the pooled effects. Though it is well-documented that inverse-Gamma and Wishart prior distributions for variance terms are inappropriate, unfortunately they are still adopted in the meta-analysis field. For example, Menke (30), Riley et al. (12) and Zwinderman and Bussuyt (13) use a Wishart prior distribution in bivariate meta-analyses of sensitivity and specificity from multiple test accuracy studies. Yang et al. use a Wishart prior distribution in their network meta-analysis of multiple therapies for acute ischemic stroke (31), as does Jansen in a network meta-analysis of multiple treatments of lung cancer (32). In their seminal paper on the Bayesian approach to multivariate meta-analysis of multiple outcomes, Nam et al. use an inverse Gamma prior on each of the between-study variances (18). Therefore, given its continued use, herein it is important to demonstrate the drawback of the Gamma prior distribution within multivariate meta-analysis, with a novel angle on its impact on  $\rho_B$ , the amount of borrowing of strength and joint inferences.

## 2.4 Choice of prior distribution for $\rho_B$

A range of (realistically) vague prior distributions for the between-study correlation are considered to account for varying levels of hypothetical prior knowledge. Below are five possible prior distributions in which options 1 to 3 allow the between-study

correlation to be positive or negative, and options 4 and 5 only allow the between-study correlation to be positive. The five prior distributions are shown in Figure 1.

(INSERT FIGURE 1)

*Option 1*

$$\rho_B \sim U(-1,1)$$

This prior distribution gives equal weight to all possible positive and negative values of correlation. This distribution is often used in practice [for example, see the following references (1, 17, 18)], and is usually considered when there is no prior information regarding the true value of the between-study correlation.

*Option 2*

$$z = \frac{1}{2} \log \left( \frac{1+\rho_B}{1-\rho_B} \right) \sim N(0, \text{sd}=0.4)$$

This prior distribution is referred to as a Fisher prior and it is similar to option 1, as it has the same mean and allows both positive and negative values (21), but gives more weight around the mean and less weight at the extremes.

*Option 3*

$$\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5, 1.5)$$

Similar to options 1 and 2, this Beta prior distribution also allows for positive and negative values of the between-study correlation. It is similar to option 1 in that it is relatively flat across the range of values, with the exception that values at the extreme ends of the distribution are considered extremely unlikely. The scale and shape parameter values of 1.5 are chosen here to ensure a prior distribution that is noticeably different to both options 1 and 2.

#### *Option 4*

$$\rho_B \sim U(0,1)$$

This prior distribution gives equal weight to all possible positive values of correlation.

#### *Option 5*

$$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$$

Similar to option 4, this logit prior distribution allows only positive values, however, more weight is given around the mean and less weight is given in the tails of the distribution.

Although these five prior distributions reflect a key range of options, we recognise that other choices of prior distributions could be specified. In particular, it may be that negative values of the correlation are very unlikely but not impossible and therefore a prior distribution might be specified that, unlike priors 4 and 5, allows for some small

probability of negative values. An example of such a prior distribution is shown in the Supplementary Material. Clearly the choice will be context specific, but here onwards the five prior distributions described above are our key focus.

### **3 Simulation study to examine choice of prior distributions**

We now describe the methods and results of the simulation study to examine the impact of (realistically) vague prior distributions for the between-study variances and correlation in a Bayesian estimation of bivariate meta-analysis model (1). The simulation focuses mainly on  $N=10$  studies per meta-analysis, but both complete data (both outcomes available in all 10 studies) and missing data (some studies only provide one outcome) situations are considered. Alternative  $N$  is also considered briefly in Section 3.2.5.

#### **3.1 Methods of the simulation study**

The simulation study involves three key steps, as follows.

##### **Step 1: Generation of bivariate meta-analysis datasets for a range of settings.**

We use the simulation data previously generated by Riley et al. (33), where full details of the simulation process are documented. Briefly, for each simulation scenario (see below), a true between-study and within-study bivariate Normal model was specified according to equation (1). Then, allowing for the specified within- and between-study variances and correlations, two effect estimates ( $Y_{i1}$  and  $Y_{i2}$ ) were generated (one for



each outcome) for each of the 10 studies in the meta-analysis. This was repeated 1000 times, so to generate 1000 meta-analysis datasets for each simulation. A range of simulation settings are considered (Table 1).

**Table 1: Settings for which simulated meta-analysis datasets were generated.**

Setting	True parameter value					
	$\rho_{Wi}$	$\rho_B$	$\beta_1$	$\beta_2$	$\tau_1$	$\tau_2$
Complete data						
1	0	0	0	2	0.5	0.5
2	0	0.8	0	2	0.5	0.5
3	0.8	0	0	2	0.5	0.5
4	0.8	0.8	0	2	0.5	0.5
5	0.8	0.8	0	2	0.05	0.05
Missing data						
6	0	0	0	2	0.5	0.5
7	0	0.8	0	2	0.5	0.5
8	0.8	0	0	2	0.5	0.5
9	0.8	0.8	0	2	0.5	0.5

Within-study variances ( $s_{ij}^2$ ) were drawn from a log normal distribution, and had an average value of 0.5. Therefore, settings 1 to 4 and 6 to 9 had similarly sized within- and between-study variances on average, whilst settings 5 and 10 had relatively large within-study variances.

Settings 1 to 5 involve complete data (i.e.  $Y_{i1}$  and  $Y_{i2}$  are available for all studies) but settings 6 to 9 involve missing data, where some studies were made to have only  $Y_{i2}$ . Missing data scenarios are very important, as borrowing of strength may be large in such situations. We chose to generate non-ignorable missingness. In each complete data meta-analysis dataset, the treatment effect estimate for outcome 1 ( $Y_{i1}$ ) was selectively removed if it was larger than the unweighted mean of  $Y_{i1}$  within each set of 10 trials, i.e.:

$$\text{Remove } Y_{il} \text{ if } Y_{il} > \frac{1}{10} \sum_{i=1}^{10} Y_{il}$$

On average, this process removed half of the treatment effect estimates and their standard deviations for outcome 1 in the simulated datasets. This missing data process was chosen to reflect selective outcome reporting bias in which an outcome is measured and analysed but not reported on the basis of the results (34, 35). Though this missing data mechanism is missing-not-at-random (MNAR), the utilisation of correlation from reported outcomes can still reduce (though not entirely remove) bias in univariate meta-analysis results in this situation, as shown elsewhere (6), and is now a key reason for applying the multivariate model (36). Therefore, it is of particular interest whether chosen prior distributions affect the bivariate meta-analysis results for outcome 1 in this setting.

**Step 2: Fit model (1) to each dataset in each setting, for all the different sets of prior distributions.**

To each of the 1000 meta-analysis datasets within each of the nine settings, model (1) was fitted using MCMC with a particular set of chosen prior distributions. This was then repeated for each different set of prior distributions. The different sets of prior distributions were as follows.

*Pooled effects ( $\beta_j$ )*

The prior distributions for  $\beta_j$  were always given a vague  $N(0, 1000^2)$ .

### *Between-study variances ( $\tau_j^2$ )*

Two prior distributions for  $\tau_j^2$  were chosen (one that appeared appropriate and one that appeared inappropriate) based on the results of a univariate meta-analysis of the simulated datasets where  $\rho_{W_i}=\rho_B=0$  in model (1) (see Table S1 in the Supplementary material). Because the true between-study standard deviations in the simulations were 0.5, a  $\tau_j \sim N(0,2)$  ( $\tau_j > 0$ ) prior distribution appeared most suitable (realistically vague) amongst six prior distributions previously explored by Lambert et al. (8). In contrast, the Gamma(0.1,0.1) prior distribution for  $1/\tau_j^2$  was, as expected, by far the poorest in terms of estimating  $\tau_j$  accurately. However, as this inappropriate prior distribution is still often adopted in the multivariate meta-analysis literature (see earlier) we include it here to highlight its impact. Thus, in each setting of the simulation study, both these prior distributions were evaluated to compare the impact of a seemingly suitable prior distribution with a seemingly inappropriate prior distribution for  $\tau_j$ .

### *Between-study correlation ( $\rho_B$ )*

The prior distributions evaluated for  $\rho_B$  were the five prior distributions detailed in Section 2.4.

This led to 10 combinations of the prior distributions for between-study variances and correlations shown in Table 2.

**Table 2: All combinations of prior distributions for between-study correlation and between-study variance.**

Combination	Prior distribution for $\rho_B$	Prior distribution for $\tau_j$
(i)	$\rho_B \sim U(-1,1)$	
(ii)	$z = \frac{1}{2} \log \left( \frac{1+\rho_B}{1-\rho_B} \right) \sim N(0, \text{sd}=0.4)$	
(iii)	$\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5, 1.5)$	$\tau_j \sim N(0, 2), \tau_j > 0$
(iv)	$\rho_B \sim U(0,1)$	
(v)	$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$	
(vi)	$\rho_B \sim U(-1,1)$	
(vii)	$z = \frac{1}{2} \log \left( \frac{1+\rho_B}{1-\rho_B} \right) \sim N(0, \text{sd}=0.4)$	
(viii)	$\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5, 1.5)$	$1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$
(ix)	$\rho_B \sim U(0,1)$	
(x)	$\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$	

In each analysis, the posterior parameter estimates were obtained using the Gibbs Sampler MCMC method, which was implemented in SAS 9.3 using the PROC MCMC procedure (37). For each dataset, the analyses were performed with 300,000 iterations after allowing for a 200,000 iteration burn in and the samples were thinned by 100 to reduce autocorrelation (see Supplementary Material for SAS code). The convergence of parameters was checked using history and trace plots.

### Step 3: Summarise results.

In each setting, to summarise and compare the posterior results for each set of prior distributions, the following were calculated from the set of 1000:

- The mean posterior mean of pooled effects across all simulations, the mean and median posterior median of between-study standard deviation across all simulations; and the mean and median posterior median of between-study correlation across all simulations (to check for bias),
- The mean and median standard deviation of the posterior pooled effects across simulations,
- The mean-squared error (MSE) of the pooled effects, calculated by the average squared difference from the true value across the 1000 simulated datasets,
- The coverage performance of the 95% credible intervals for the pooled effects, calculated by the percentage of the 1000 95% credible intervals that contain the true effect.

Further, we also evaluated performance in terms of predictive inferences about treatment effects in new trials. The predictive distribution of treatment effects in a new trial was assumed to be:

$$\begin{pmatrix} \theta_{i1_{new}} \\ \theta_{i2_{new}} \end{pmatrix} \sim N \left( \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{D} \right) \quad (2)$$

In each analysis, values of  $\theta_{i1_{new}}$  and  $\theta_{i2_{new}}$  were sampled from this distribution during the MCMC process, which naturally accounts for the uncertainty in the pooled average effects,  $\beta_1$  and  $\beta_2$ , and the uncertainty in the between-study covariance matrix,  $\mathbf{D}$ .

Across all datasets in each setting for each set of prior distributions, we used these to derive:

- The average marginal probability that  $\theta_{i1new} > 0$ , the average marginal probability that  $\theta_{i2new} > 2$ , and the average *joint* probability that both  $\theta_{i1new} > 0$  and  $\theta_{i2new} > 2$ .

In settings 1, 3, 6 and 8, where  $\rho_B = 0$ , the two true marginal probabilities that  $\theta_{i1new} > 0$  and  $\theta_{i2new} > 2$  was both 0.5, and the true joint probability that  $\theta_{i1new} > 0$  and  $\theta_{i2new} > 2$  was 0.25. When  $\rho_B = 0.8$  in settings 2, 4, 5, 7 and 9, the true joint probability was 0.4.

## 3.2 Results of the simulation study

### 3.2.1 Complete case data when using prior distribution for between-study variance of $\tau_j \sim N(0,2)$ ( $\tau_j > 0$ )

Table 3 and Table 4 display the simulation results for setting 1 and setting 4, respectively, for the different prior distributions for the between-study correlation where the sensible prior distribution for  $\tau_j$  is used ( $N(0,2)$  truncated at zero). The equivalent results for settings 2 and 3 are presented in Tables S2 and S3 in the Supplementary Material.

**Table 3: Simulation results for 10 studies with complete data (setting 1). The within-study correlation,  $\rho_{wi}$  was zero and the same for each study. The prior distribution for  $\tau_j$  is  $N(0,2)I(0,)$  and for  $\beta_j$  is  $N(0,1000^2)$ .**

Prior for $\rho_B$	Mean posterior mean of $\beta_1$ (SD of mean)	Mean/ median SD of $\beta_1$	Mean MSE of $\beta_1$	% of 95% CrIs for $\beta_1$ including $\beta_1$ %	Mean prob ( $\theta_{i1new}>0$ )	Mean posterior mean of $\beta_2$ (SD of mean)	Mean/ median SD of $\beta_2$	Mean MSE of $\beta_2$	% of 95% CrIs for $\beta_2$ including $\beta_2$ (%)	Mean prob ( $\theta_{i2new}>2$ )	Mean/ median posterior median $\tau_1$	Mean/ median posterior median $\tau_2$	Mean/ median posterior median $\rho_B$	Mean prob ( $\theta_{i1new}>0$ & $\theta_{i2new}>2$ )
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
$\rho_B \sim U(-1,1)$	-0.0020 (0.1955)	0.2185/ 0.2159	0.0382	95.6	0.4969	2.0011 (0.2198)	0.2616/ 0.2583	0.0483	96.6	0.4989	0.5006/ 0.4985	0.5344/ 0.5280	0.0070/ 0.0045	0.2483
Fisher $z \sim$ $N(0, sd=0.4)$	-0.0021 (0.1952)	0.2168/ 0.2136	0.0381	95.8	0.4966	2.0011 (0.2193)	0.2606/ 0.2569	0.0480	96.6	0.4989	0.4965/ 0.4953	0.5293/ 0.5281	0.0026/ 0.0012	- 0.2478
$(\rho_B+1)/2 \sim$ $Beta(1.5,1.5)$	-0.0025 (0.1957)	0.2179/ 0.2149	0.0383	95.6	0.4965	2.0014 (0.2195)	0.2614/ 0.2600	0.0481	96.6	0.5000	0.4995/ 0.4996	0.5327/ 0.5295	0.0049/ 0.0055	0.2480
$\rho_B \sim U(0,1)$	-0.0020 (0.1954)	0.2190/ 0.2160	0.0382	95.5	0.4963	2.0022 (0.2203)	0.2616/ 0.2579	0.0485	96.4	0.4994	0.5006/ 0.5050	0.5326/ 0.5302	0.4121/ 0.4114	0.2958
$Logit(\rho_B) \sim$ $N(0, sd=0.8)$	-0.0019 (0.1955)	0.2198/ 0.2175	0.0382	95.5	0.4955	2.0017 (0.2204)	0.2619/ 0.2573	0.0485	96.5	0.4991	0.5033/ 0.5041	0.5359/ 0.5285	0.4682/ 0.4738	0.3012

MSE is mean-square error; CrI is credible interval; SD is standard deviation; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

**Table 4: Simulation results for 10 studies with complete data (setting 4). The within-study correlation,  $\rho_{wi}$  was 0.8 and the same for each study. The prior distribution for  $\tau_j$  is  $N(0,2)I(0,)$  and for  $\beta_j$  is  $N(0,1000^2)$ .**

Prior for $\rho_B$	Mean posterior mean of $\beta_1$ (SD of mean)	Mean/ median SD of $\beta_1$	Mean MSE of $\beta_1$	% of 95% CrIs for $\beta_1$ including $\beta_1$	Mean prob ( $\theta_{i1new}>0$ )	Mean posterior mean of $\beta_2$ (SD of mean)	Mean/ median SD of $\beta_2$	Mean MSE of $\beta_2$	% of 95% CrIs for $\beta_2$ including $\beta_2$	Mean prob ( $\theta_{i2new}>2$ )	Mean/ median posterior median $\tau_1$	Mean/ median posterior median $\tau_2$	Mean/ median posterior median $\rho_B$	Mean prob ( $\theta_{i1new}>0$ & $\theta_{i2new}>2$ )
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
$\rho_B \sim U(-1,1)$	-0.0091 (0.1850)	0.2081/ 0.2085	0.0343	95.6	0.4962	2.0009 (0.2029)	0.2319/ 0.2320	0.0411	96.3	0.4973	0.5134/ 0.5203	0.4965/ 0.5019	0.5160/ 0.5770	0.3279
Fisher $z \sim$ $N(0, sd=0.4)$	-0.0088 (0.1848)	0.2055/ 0.2070	0.0342	95.5	0.4968	2.0002 (0.2055)	0.2329/ 0.2301	0.0422	96.0	0.4974	0.5047/ 0.5150	0.4813/ 0.4829	0.2363/ 0.2452	0.2915
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.0088 (0.1846)	0.2071/ 0.2081	0.0341	95.6	0.4963	2.0012 (0.2034)	0.2315/ 0.2301	0.0413	96.3	0.4975	0.5094/ 0.5179	0.4893/ 0.4933	0.4226/ 0.4582	0.3159
$\rho_B \sim U(0,1)$	-0.0090 (0.1844)	0.2074/ 0.2067	0.0341	95.3	0.4958	1.9994 (0.2116)	0.2299/ 0.2283	0.0448	96.2	0.4975	0.5105/ 0.5153	0.5036/ 0.5056	0.6458/ 0.6562	0.3460
Logit( $\rho_B$ ) $\sim$ $N(0, sd=0.8)$	-0.0092 (0.1844)	0.2045/ 0.2045	0.0341	95.4	0.4957	2.0012 (0.2026)	0.2285/ 0.2270	0.0410	95.9	0.4975	0.5021/ 0.5081	0.4891/ 0.4908	0.5545/ 0.5538	0.3312

MSE is mean-square error; CrI is credible interval; SD is standard deviation; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.



In all settings the choice of prior distribution for  $\rho_B$  is informative of the posterior estimate of  $\rho_B$ . This is expected since there are only 10 studies per meta-analysis, so there are only 10 data points to estimate a correlation, and thus the posterior mean is similar to the prior mean. For example, in setting 1 ( $\rho_{W1}=\rho_B=0$ , Table 3) where  $\rho_B \sim U(-1,1)$ , the mean posterior median for  $\rho_B$  across simulations is 0.007. When  $\rho_B \sim U(0,1)$ , the mean posterior median for  $\rho_B$  across simulations is 0.412. A similar result is observed in settings 2-4. In setting 2 and setting 4, the true value of  $\rho_B$  is 0.8, however, none of the selected prior distributions led to average medians of  $\rho_B$  across simulations close to its true value. For example, in setting 4 ( $\rho_{W1}=\rho_B=0.8$ , Table 4) where  $\rho_B \sim U(0,1)$ , the average posterior median of  $\rho_B$  is only 0.646.

The performance of the 95% credible intervals is also close to 95% for  $\beta_j$  regardless of the choice of prior distribution for  $\rho_B$ . Further, the choice of prior distribution for  $\rho_B$  has little impact on the posterior means of  $\beta_1$  and  $\beta_2$  across simulations, and their mean standard deviations. In other words, there appears to be very little borrowing of strength, which agrees with previous work that shows the borrowing of strength in a bivariate meta-analysis toward the estimates of  $\beta_j$  is usually very small when complete data are available for both outcomes (5, 33).

However, the prior distribution for  $\rho_B$  does have a larger impact upon average *joint* inferences across both outcomes. The average joint probability that  $\hat{\theta}_{i1new} > 0$  &  $\hat{\theta}_{i2new} > 2$  is slightly higher for the prior distributions for  $\rho_B$  that allow only positive values. Also, since no prior distribution leads to an average posterior median of the between-study correlation close to 0.8, the average joint probability is always lower than the true value of 0.4.

### 3.2.2 Complete case data when using prior distribution for between-study variance of $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$

Table 5 displays the simulation results for setting 3 where the inappropriate  $\text{Gamma}(0.1, 0.1)$  prior distribution for  $1/\tau_j^2$  is used. The equivalent results for settings 1, 2 and 4 are in Tables S4, S5 and S6, respectively, in the Supplementary Material. The posterior means for  $\beta_1$  and  $\beta_2$  remain approximately unbiased for all choices of the prior distributions for  $\rho_B$ , for settings 1 to 4. However, the posterior distributions of the  $\tau_j^2$ s are centred on much larger values than 0.25 for both outcomes. Therefore, the standard deviations of the pooled effects are much larger than those when  $\tau_j \sim N(0, 2)I(0, \cdot)$ . Thus, the credible intervals for the pooled effect estimates are too wide, leading to inappropriate coverage of 100% in all settings, regardless of the choice of prior distribution for  $\rho_B$ .

The simulation results also show that when the values of  $\tau_j$  are larger,  $\rho_B$  is likely to increase. This can lead to a huge upward bias in the posterior distribution of  $\rho_B$ , even with the  $U(-1, 1)$  prior distribution for  $\rho_B$ . For example, using prior distributions of  $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$  and  $\rho_B \sim U(-1, 1)$  in setting 3 (true  $\rho_B = 0$ , Table 5), the mean posterior median  $\rho_B$  across simulations is 0.605. However, using the same prior distribution for  $\rho_B$  but a prior distribution for  $\tau_j$  of  $N(0, 2)I(0, \cdot)$  the average posterior median for  $\rho_B$  is -0.035 (Table 4). This is due to much higher average estimates of  $\tau_j$  with the Gamma prior distribution (mean posterior median  $\tau_1 = 1.926$ , mean posterior median  $\tau_2 = 2.157$ ) compared to the half Normal prior distribution (mean posterior median  $\tau_1 = 0.532$ , mean posterior median  $\tau_2 = 0.536$ ).

The estimates of the joint probability (that  $\hat{\theta}_{i1new} > 0$  &  $\hat{\theta}_{i2new} > 2$ ) are again influenced by the estimate of correlation between the outcomes. In the same example as above, where the correlation is dramatically overestimated, the true joint probability is 0.25 but the mean joint probability estimate across simulations is 0.342. This highlights that seemingly vague prior distributions for the  $\tau_j$ 's and  $\rho_B$  may have undesired impact on the posterior conclusions, which may lead to incorrect (joint) inferences.

**Table 5: Simulation results for 10 studies with complete data (setting 3). The within-study correlation,  $\rho_{wi}$  was 0.8 and the same for each study. The prior distribution for  $1/\tau_j^2$  is gamma(0.1,0.1) and for  $\beta_j$  is N(0,1000<sup>2</sup>).**

Prior for $\rho_B$	Mean posterior mean of $\beta_1$ (SD of mean)	Mean/ median SD of $\beta_1$	Mean MSE of $\beta_1$	% of 95% CrIs for $\beta_1$ including $\beta_1$	Mean prob ( $\theta_{i1new}>0$ )	Mean posterior mean of $\beta_2$ (SD of mean)	Mean/ median SD of $\beta_2$	Mean MSE of $\beta_2$	% of 95% CrIs for $\beta_2$ including $\beta_2$	Mean prob ( $\theta_{i2new}>2$ )	Mean/ median posterior median $\tau_1$	Mean/ median posterior median $\tau_2$	Mean/ median posterior median $\rho_B$	Mean prob ( $\theta_{i1new}>0$ & $\theta_{i2new}>2$ )
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.0	0.25
$\rho_B \sim U(-1,1)$	0.0024 (0.1999)	0.6813/ 0.6785	0.0399	100	0.5008	1.9923 (0.2412)	0.7940/ 0.7822	0.0582	100	0.4986	1.9255/ 1.9173	2.1574/ 2.1207	0.6047/ 0.8646	0.3419
Fisher z~ N(0,sd=0.4)	0.0031 (0.2006)	0.5985/ 0.5968	0.0402	100	0.5010	1.9937 (0.2506)	0.6925/ 0.6886	0.0628	100	0.4986	1.7049/ 1.6961	1.8709/ 1.8571	0.2378/ 0.2482	0.2811
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	0.0029 (0.1997)	0.6511/ 0.6491	0.0399	100	0.5008	1.9923 (0.2418)	0.7595/ 0.7513	0.0584	100	0.4992	1.8459/ 1.8391	2.0611/ 2.0399	0.6045/ 0.7817	0.3338
$\rho_B \sim U(0,1)$	0.0023 (0.2005)	0.6822/ 0.6770	0.0401	100	0.5005	1.9931 (0.2446)	0.7980/ 0.7863	0.0598	100	0.4988	1.9280/ 1.9166	2.1711/ 2.1359	0.8858/ 0.8982	0.4132
Logit( $\rho_B$ )~ N(0,sd=0.8)	0.0029 (0.2013)	0.6077/ 0.6049	0.0405	100	0.5010	1.9941 (0.2480)	0.7095/ 0.7037	0.0615	100	0.4995	1.7370/ 1.7305	1.9309/ 1.9167	0.6942/ 0.6970	0.3637

MSE is mean-square error; CrI is credible interval; SD is standard deviation; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

### 3.2.3 Results with missing data when prior distribution for $\tau_j$ is $N(0,2)$ ( $\tau_j > 0$ )

For the missing data settings, it was of particular interest whether the prior distributions affect the outcome 1 results (for which missing data was selectively missing) and the amount of borrowing of strength. Both the  $N(0,2)$  ( $\tau > 0$ ) prior distribution for  $\tau_j$  and the  $\text{Gamma}(0.1,0.1)$  prior distribution for  $1/\tau_j^2$  were considered again, but for brevity the results are only presented for settings 8 and 9 where there are within-study correlations of 0.8.

The simulation results are shown in Table 6 for setting 9 ( $\beta_1=0, \beta_2=2, \tau_1=\tau_2=0.5, \rho_{W1}=\rho_B=0.8$ ) (setting 8 is in Table S7 in the Supplementary Material). As expected, due to the selective missingness, the average posterior mean for  $\beta_1$  is consistently lower than the true value for all prior distributions, and in all settings. For example, where the true  $\beta_1=0$ , the mean  $\beta_1$  is, on average, -0.432 (s.d. is 0.250) where  $\rho_B \sim U(-1,1)$ . However, if the posterior mean for  $\rho_B$  is higher, the bias in the posterior distribution of  $\beta_1$  is lower; in other words, the borrowing of strength increases as the posterior mean for  $\rho_B$  increases. For example, in the same set of results, for  $\rho_B \sim U(0,1)$  the average posterior median  $\rho_B$  is 0.545 and the mean  $\beta_1$  across simulations is -0.390. The estimated effect for outcome two remains approximately unbiased across all settings as there is complete data for this outcome (6).

Though bias remains in the mean  $\beta_1$  across simulations, crucially it is closer to the true value of 0 than a separate univariate meta-analysis for outcome 1. In the same example, where  $\rho_B \sim U(-1,1)$  the average mean  $\beta_1$  is -0.432 (s.d. 0.250) whereas the average mean

from the univariate analysis is -0.483 (s.d. is 0.251). The MSE of  $\beta_I$  is also lower in the bivariate model compared to the univariate model for all prior distributions for  $\rho_B$ . In the same scenario, the MSE of  $\beta_I$  from the bivariate analysis is 0.249 but 0.296 in the univariate analysis. Further, if a more appropriate prior distribution is used for  $\rho_B$ , the greater the reduction in the MSE. The more appropriate prior distributions for  $\rho_B$  also leads to better coverage. Where  $\rho_B \sim U(0,1)$ , the number of 95% CrIs that contain the true  $\beta_I$  is 73.5%, compared to 67.2% when  $\rho_B \sim U(-1,1)$ , and just 61.2% in the univariate analysis. Therefore, the amount of borrowing of strength is heavily influenced by the choice of prior distribution for  $\rho_B$ .

**Table 6: Simulation results for 10 studies with missing data for outcome 1 (setting 9). The within-study correlation,  $\rho_{wi}$  was 0.8 and the same for each study. The prior distribution for  $\tau_j$  is  $N(0,2)I(0,)$  and for  $\beta_j$  is  $N(0,1000^2)$ .**

Prior for $\rho_B$	Mean posterior mean of $\beta_1$ (SD of mean)	Mean/ median SD of $\beta_1$	Mean MSE of $\beta_1$	% of 95% CrIs for $\beta_1$ including $\beta_1$	Mean prob ( $\theta_{i1new}>0$ )	Mean posterior mean of $\beta_2$ (SD of mean)	Mean/ median SD of $\beta_2$	Mean MSE of $\beta_2$	% of 95% CrIs for $\beta_2$ including $\beta_2$	Mean prob ( $\theta_{i2new}>2$ )	Mean/ median posterior median $\tau_1$	Mean/ median posterior median $\tau_2$	Mean/ median posterior median $\rho_B$	Mean prob ( $\theta_{i1new}>0$ & $\theta_{i2new}>2$ )
True values	0.0	-	-	-	0.5	2.0	-	-	-	0.5	0.5	0.5	0.8	0.4
Univariate	-0.4826 (0.2513)	0.2820/ 0.2553	0.2960	61.2	0.1501	2.0009 (0.2185)	0.2593/ 0.2580	0.0477	96.5	0.4989	0.2890/ 0.2485	0.5249/ 0.5293	-	0.0749
$\rho_B \sim U(-1,1)$	-0.4324 (0.2496)	0.2787/ 0.2535	0.2492	67.2	0.1800	2.0041 (0.2324)	0.2615/ 0.2596	0.0540	95.2	0.5031	0.3216/ 0.2801	0.6108/ 0.6085	0.1552/ 0.1531	0.1129
Fisher $z \sim$ $N(0, sd=0.4)$	-0.4438 (0.2506)	0.2727/ 0.2468	0.2597	64.1	0.1714	2.0085 (0.2243)	0.2602/ 0.2579	0.0503	95.6	0.5040	0.3126/ 0.2654	0.6048/ 0.6029	0.0497/ 0.0406	0.0987
$(\rho_B+1)/2 \sim$ Beta(1.5,1.5)	-0.4364 (0.2494)	0.2756/ 0.2506	0.2526	66.4	0.1765	2.0050 (0.2329)	0.2608/ 0.2584	0.0542	95.6	0.5040	0.3176/ 0.2719	0.6092/ 0.6068	0.1109/ 0.1052	0.1061
$\rho_B \sim U(0,1)$	-0.3920 (0.2370)	0.2718/ 0.2480	0.2098	73.7	0.1948	2.0005 (0.2225)	0.2604/ 0.2582	0.0495	95.8	0.5000	0.3210/ 0.2819	0.6112/ 0.6101	0.5450/ 0.5452	0.1453
Logit( $\rho_B$ ) $\sim$ $N(0, sd=0.8)$	-0.3965 (0.2367)	0.2695/ 0.2480	0.2132	73.5	0.1918	2.0017 (0.2226)	0.2589/ 0.2569	0.0495	95.7	0.5004	0.3161/ 0.2762	0.6078/ 0.6073	0.5132/ 0.5132	0.1397

MSE is mean-square error; CrI is credible interval; SD is standard deviation; the means and medians represent the posterior means and medians from the distribution of summary estimates from the 1000 datasets.

### **3.2.4 Results with missing data when prior distribution for $1/\tau_j^2$ is Gamma(0.1,0.1)**

The results of the missing data scenario when the prior distribution for  $1/\tau_j^2$  is Gamma(0.1,0.1) are shown in Tables S8 and S9 in the Supplementary Material. As in the complete data simulations, the main finding is that the posterior estimates of  $\tau_j$  are hugely overestimated, and this leads to overly large estimates of  $\rho_B$  for all prior distributions for  $\rho_B$  (compared to when using a  $N(0,2)I(0,)$  prior distribution for  $\tau_j$ ).

### **3.2.5 Increasing the number of trials per meta-analysis**

One finding from the simulations so far is that the prior distribution for the between-study correlation can be highly informative towards the borrowing of strength, posterior results and joint inferences for meta-analyses of 10 studies, with complete and missing data. In settings 2 and 4, where there is strong true between-study correlation ( $\rho_B=0.8$ ), most of the prior distributions for  $\rho_B$  result in this parameter being underestimated. To ascertain if this improved when the number of studies per meta-analysis increases, the simulations were repeated with 25 and 50 studies. For brevity, only the results for complete data in setting 4 (where  $\rho_{W_i}=0.8$  and  $\rho_B=0.8$ ) where  $\tau_j \sim N(0,2)I(0,)$  are discussed.

The results are shown in Tables S10 and S11 in the Supplementary Material. As expected, as the number of studies per meta-analysis increases, the posterior median of  $\rho_B$  is closer to the true value. For example, recall that given 10 studies and  $\rho_B \sim U(-1,1)$  the mean posterior median  $\rho_B$  across simulations was 0.516 (Table 4), but with 50



studies the mean posterior median is 0.734. Interestingly, the average  $\rho_B$  is still underestimating the true value of 0.8 for any of the prior distributions for  $\rho_B$ , and the choice of prior distribution is still influential even when there are 50 studies. The mean joint probability estimates are closer to 0.4 with 50 studies compared to 25 or 10 studies, but they are still lower than the true value of 0.4 for all prior distributions for  $\rho_B$ . This again is partly due to the underestimated between-study correlation, but is also due to the uncertainty in all parameters. For instance, even when repeating the simulations in setting 4 and forcing  $\rho_B$  to be 0.8, the mean joint probability is 0.372 and thus still underestimated compared to 0.4. Only in the unrealistic situation where all parameters are known (i.e.  $\rho_B$ ,  $\tau_j$ ,  $\beta_1$ , and  $\beta_2$  are fixed at their true values), is the mean joint probability approximately 0.4. Therefore, unless the meta-analysis has a very large number of studies, the uncertainty in the estimates of the pooled treatment effects, the between-study variances and the between-study correlation, will be propagated into lower joint probabilistic inferences than if these parameters were known.

This finding can perhaps be considered comparable to the use of the  $t$ -distribution for the derivation of prediction intervals for  $\theta_{new}$  by Higgins et al. in a frequentist framework (38). Here, the  $t$ -distribution is used instead of the Normal distribution to account for the uncertainty in the between-study variance. This can be extended to a bivariate setting. If 2,000,000 samples of  $x$  and  $y$  are drawn from a bivariate  $t$ -distribution (with 8 degrees of freedom since the number of trials is 10) with means zero and two, respectively, variances equal to 0.25, and correlation equal to 0.8, then the joint probability that  $x > 0$  and  $y > 2$  is just 0.366. This is similar to the mean joint probability estimate of 0.372 in the simulation study when the correlation is forced to be 0.8. The joint probability is only equal to 0.4 when the bivariate Normal distribution is

assumed. If 2,000,000  $x$  and  $y$  are sampled from the bivariate Normal distribution, with the same parameter values as those used above, then the resulting probability is very close to 0.4.

### **3.2.6 Reducing the size of the between-study variance relative to the within-study variance**

In the simulations so far, the true between-study variance was 0.25 for both outcomes, which was a similar size compared to the within-study variances. If the between-study variances are large relative to the within-study variances, it is known that the between-study correlation (rather than the within-study correlations) will be most influential toward the borrowing of strength (1). However, even when the between-study variances are small relative to the within-study variances, the magnitude of between-study correlation is crucial toward joint (predictive) inferences, and so it is important to estimate it reliably. However, in the frequentist setting, it is known to be potentially problematic to estimate between-study variances and correlations when the between-study variation is relatively small, as shown elsewhere (33). Therefore, in the Bayesian setting, prior distributions for between-study variances and correlations are likely to be even more influential toward their posterior results when the between-study variation is relatively small.

To illustrate this, bivariate meta-analysis data were additionally simulated for setting 5 using the same approach as before, but now with true between-study variances of 0.0025 compared to within-study variances as before (i.e. on average 0.5). Only within and between-study correlations of 0.8 were considered and the results are shown in

Supplementary Table S12. The results show that the prior distributions for the between-study variances and correlations are very influential, and more than in the earlier simulations. For example, the mean posterior median correlation is 0.281 (true value is 0.8) from the new simulations for setting 5 when using a  $U(-1,1)$  prior distribution; this is much closer to the prior distribution mean compared to the mean posterior median correlation of 0.516 in the earlier simulations in setting 4 (Table 4) where the between-study variation was larger.

## **4 Illustrative example**

This section illustrates the key findings from the simulation study in a meta-analysis dataset involving (potentially selectively) missing data. The example is introduced, and then the key results summarised.

### **4.1 Combining partially and fully-adjusted results**

The dataset for the illustrative example is from a previous individual participant data (IPD) meta-analysis of trials concerned with whether smoking is a prognostic factor for stroke, where smoking is a binary variable by yes (current smoker) or no (not current smoker) (23). The summary results for the 10 trials are shown in Table 7. There are two prognostic effects for smoking: a partially adjusted log hazard ratio (HR), which is adjusted for treatment, and a fully adjusted log HR, which is adjusted for treatment, age and BMI. There is missing information for age and BMI in five out of 10 trials, and so only partially adjusted HR estimates are available in these. However, in the remaining five trials, there is information to estimate both fully and partially adjusted log HRs.

These prognostic effect estimates are highly correlated with the within-study correlations close to +1 (derived from bootstrapping) (23). Interestingly, the five studies only giving partially adjusted results have, on average, smaller HR estimates than in those studies providing both partially and fully adjusted effects. Therefore, there is concern that there is selective reporting bias here for the fully adjusted results, and that a univariate meta-analysis of the fully adjusted results will be biased upwards. A bivariate meta-analysis of the partially and fully adjusted results borrows strength to reduce this bias.

**Table 7: Results for the 10 trials in the meta-analysis of partially-adjusted and fully-adjusted log hazard ratios (log HR) (23).**

<b>Trial name*</b>	<b>Control</b>	<b>Treatment</b>	<b>Partially-adjusted log HR (var)</b>	<b>Fully-adjusted log HR (var)</b>	<b>Within-study correlations (from bootstrap)</b>
ATMH	750	780	0.216 (0.752)	0.173 (0.754)	0.992
HEP	199	150	1.238 (0.182)	1.477 (0.223)	0.893
EWPHE	82	90	-1.038 (1.080)	-0.667 (1.125)	0.988
HDFP	2371	2427	0.884 (0.072)	0.894 (0.074)	0.985
MRC-1	3445	3546	1.232 (0.119)	1.209 (0.120)	0.986
MRC-2	1337	1314	0.379 (0.039)	-	-
SHEP	2371	2365	0.399 (0.027)	-	-
STOP	131	137	1.203 (1.256)	-	-
Sy-Chi	1139	1252	0.633 (0.042)	-	-
Sy-Eur	2297	2398	0.156 (0.100)	-	-

Upon applying the bivariate meta-analysis, two prior distributions for the between-study variances are considered for comparison. The first is the inappropriate Gamma prior distribution, where  $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$  (8). The second prior distribution is an empirical prior for future meta-analyses with a binary outcome (10) where

$\tau_j^2 \sim \text{lognormal}(-2.89, 1.91^2)$ . This prior distribution is proposed by Turner et al. for non-pharmacological interventions with semi-objective outcomes (an objective outcome that is not all-cause mortality). The median for  $\tau_j^2$  is 0.056 and a 95% prior interval is 0.001 to 2.35. This prior distribution is not an exact match as these are prognostic rather than intervention effects, and the outcome is survival rather than binary. However, the event (stroke) is rare in this example and hazard ratios and odds ratios are often similar in this setting (39, 40), therefore this empirical prior distribution is considered suitable for illustrative application here.

## 4.2 Results from illustrative example

The results of the meta-analyses are shown in Table 8. Utilisation of correlation leads to large borrowing of strength toward the fully adjusted pooled results in the bivariate meta-analysis. For example, in the analysis using the empirical prior distributions for the variances and the  $U(0,1)$  prior distribution for the correlation, the fully adjusted pooled estimate for the logHR is 0.68 compared to 0.98 in the univariate analysis, which corresponds to a HR of 1.97 rather than 2.66. In regards the influence of the choice of prior distributions for the variances and correlations, the key findings are now discussed, which also highlight those identified in the simulation study.

*Key finding (i): The choice of prior distribution for  $\rho_B$  influences the posterior estimates for  $\rho_B$  and thus borrowing of strength toward  $\beta_j$  and joint inferences.*

As expected, the choice of prior distribution for  $\rho_B$  influences the mean  $\rho_B$  and its 95% CrI. Using the empirical prior distribution for  $\tau_j$ , the posterior median  $\rho_B$  is 0.069 (95%

CrI -0.618 to 0.695) when using the Fisher prior distribution for  $\rho_B$  (Figure 2). However, when  $\rho_B \sim U(0,1)$ , the posterior median  $\rho_B$  is 0.561 (95% CrI 0.035 to 0.983). These large changes in the between-study correlation affect the pooled treatment effect estimates. The posterior mean fully-adjusted log HR is 0.701 (95% CrI 0.410 to 1.037) when using the Fisher prior distribution, compared to 0.681 (95% CrI 0.404 to 0.995) with the Uniform prior distribution. The latter leads to more borrowing of strength in the bivariate analysis, which gives a narrower CrI and slightly lower summary prognostic effect than identified in the other analyses. Joint inferences are also affected. For example, consider the posterior probability that both partially and fully adjusted HRs are  $>1.5$  in the analyses using the empirical-based prior for  $\tau_j$ . These range from 0.73 to 0.8 depending on the chosen prior distribution for the between-study correlation.

(INSERT FIGURE 2)

*Key finding (ii): The choice of prior distribution for  $\tau_j$  influences the posterior results for  $\rho_B$  and thus borrowing of strength toward  $\beta_j$  and joint inferences.*

As observed in the simulation study, as the estimates of  $\tau_j$  increase,  $\rho_B$  also increases, even when the prior distribution for  $\rho_B$  remains the same. For example, when  $\rho_B \sim U(-1,1)$ , the posterior median for  $\rho_B$  is 0.199 (95% CrI -0.917 to 0.974) if  $\tau_j^2 \sim \text{lognormal}(-2.89, 1.91^2)$ , and 0.842 (95% CrI -0.644 to 0.999) when  $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$ . This is because the posterior estimates of  $\tau_j$  differ for these two prior distributions for  $\tau_j$  (Figure 3) ( $\tau_1^2$  is 0.036 and  $\tau_2^2$  is 0.035 with the empirical prior distribution, compared to  $\tau_1^2=4.508$  and  $\tau_2^2=6.821$  with a  $\text{Gamma}(0.1, 0.1)$  prior distribution). This example illustrates that the choice of prior distribution for the between-study variances can

impact considerably upon the posterior distribution for  $\rho_B$ . Subsequently, it also impacts upon the borrowing of strength and joint inferences. For example, in the bivariate meta-analysis with a  $U(0,1)$  prior distribution for the between-study correlation, the inappropriate Gamma prior for  $1/\tau_j^2$  leads to a joint probability of 0.50 that both the true partially and fully adjusted HRs are  $>1.5$ . In contrast, when using the empirically-based prior distribution, the predicted probability is 0.77 and thus far larger.

(INSERT FIGURE 3)

*Key finding (iii): The prior distribution for  $\rho_B$  also influences the posterior estimates for  $\tau_j$ .*

As identified in the simulation study, the prior distribution for  $\rho_B$  can alter the posterior distributions for  $\tau_j$ . When  $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$  and  $\rho_B \sim U(-1, 1)$ , the posterior median  $\tau_1^2$  is 4.508 (95% CrI 1.570 to 11.341) and  $\tau_2^2$  is 6.821 (95% CrI 1.924 to 22.742) (Figure 4). However, when  $1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)$  but with the Fisher prior distribution for  $\rho_B$ , the posterior medians of  $\tau_1^2$  and  $\tau_2^2$  are 3.475 (95% CrI 1.266 to 8.940) and 10.201 (95% CrI 1.929 to 42.271), respectively.

(INSERT FIGURE 4)

*Key finding (iv): The Gamma prior distribution for  $1/\tau_j^2$  is inappropriate and empirically-based prior distributions are preferred for multivariate meta-analysis.*

The Gamma prior distribution for the between-study variances appears particularly inappropriate because the posterior estimates of  $\tau_j^2$  are much larger than when using the empirical prior distribution, and this increases the mean of the posterior distribution of  $\rho_B$ , which affects the joint probability estimates. This finding agrees with those in the simulation study and those already determined elsewhere, for example by Lambert et al. (8) and Wei et al. (17) about the influential impact of a Gamma (Wishart) prior distribution on the between-study variances in meta-analysis, and Gelman (26) more generally. In addition, the results of our example and the simulation study reveal the Gamma prior can be influential toward the between-study correlation, and thus borrowing of strength and joint inferences. For example, in our illustrative case study in stroke, the joint probability that the partially and fully adjusted HRs are  $>1.5$  is reduced by about 0.3 to 0.4 in the analyses using the Gamma prior distribution compared to the empirically-based prior distribution.

Therefore, empirically-based prior distributions for between-study variances are highly preferable in the multivariate meta-analysis field. Similarly, empirically-based prior distributions for the between-study correlation are needed where possible, to ensure that the borrowing of strength and joint inferences are appropriate. The commonly chosen  $U(-1,1)$  prior distribution may not always be appropriate.



**Table 8: Illustrative example - summary results from bivariate meta-analysis for various prior distributions for  $\rho_B$  and  $\tau_j$ .**

<u>Prior for <math>\tau</math></u> Prior for $\rho_B$	Mean partially- adjusted log HR (95% CrI)	Mean fully- adjusted log HR (95% CrI)	Median $\tau_1^2$ (95% CrI)	Median $\tau_2^2$ (95% CrI)	Median $\rho_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully- adjusted logHR>0.405)
<u><math>1/\tau_j^2 \sim \text{Gamma}(0.1, 0.1)</math></u>						
$\rho_B = \rho_{wi} = 0$	0.553 (-0.667 to 1.779)	0.645 (-2.409 to 3.588)	3.512 (1.262 to 9.138)	10.999 (2.000 to 46.298)	-	0.283
$\rho_B \sim U(-1, 1)$	0.575 (-0.811 to 1.938)	0.674 (-1.146 to 2.434)	4.508 (1.570 to 11.341)	6.821 (1.924 to 22.742)	0.842 (-0.644 to 0.999)	0.446
Fisher $z \sim N(0, \text{sd}=0.4)$	0.580 (-0.658 to 1.819)	0.741 (-2.061 to 3.507)	3.475 (1.266 to 8.940)	10.201 (1.929 to 42.271)	0.143 (-0.647 to 0.804)	0.322
$(\rho_B + 1)/2 \sim \text{Beta}(1.5, 1.5)$	0.572 (-0.750 to 1.885)	0.676 (-1.493 to 2.902)	3.963 (1.391 to 10.333)	8.083 (1.883 to 31.501)	0.629 (-0.765 to 0.998)	0.457
$\rho_B \sim U(0, 1)$	0.581 (-0.818 to 2.006)	0.666 (-1.044 to 2.396)	4.642 (1.598 to 11.856)	6.423 (1.884 to 21.295)	0.932 (0.414 to >0.999)	0.504
Logit( $\rho_B$ ) $\sim N(0, \text{sd}=0.8)$	0.559 (-0.672 to 1.770)	0.666 (-1.768 to 3.082)	3.488 (1.297 to 8.919)	8.955 (1.922 to 34.985)	0.622 (0.234 to 0.908)	0.401
<u><math>\tau_i^2 \sim \text{lognormal}(-2.89, 1.91^2)</math></u>						
$\rho_B = \rho_{wi} = 0$	0.585 (0.335 to 0.852)	0.978 (0.447 to 1.449)	0.057 (0.000 to 0.272)	0.120 (0.001 to 0.965)	-	0.730

<u>Prior for <math>\tau</math></u> Prior for $\rho_B$	Mean partially- adjusted log HR (95% CrI)	Mean fully- adjusted log HR (95% CrI)	Median $\tau_1^2$ (95% CrI)	Median $\tau_2^2$ (95% CrI)	Median $\rho_B$ (95% CrI)	Prob (Partially- adjusted logHR>0.405 & fully- adjusted logHR>0.405)
$\rho_B \sim U(-1,1)$	0.581 (0.362 to 0.821)	0.692 (0.399 to 1.031)	0.036 (0.001 to 0.155)	0.035 (0.001 to 0.200)	0.199 (-0.917 to 0.974)	0.775
Fisher $z \sim N(0, \text{sd}=0.4)$	0.580 (0.367 to 0.812)	0.701 (0.410 to 1.037)	0.033 (0.001 to 0.143)	0.031 (0.001 to 0.161)	0.069 (-0.618 to 0.695)	0.795
$(\rho_B+1)/2 \sim \text{Beta}(1.5,1.5)$	0.581 (0.364 to 0.817)	0.696 (0.410 to 1.027)	0.036 (0.001 to 0.160)	0.036 (0.001 to 0.160)	0.029 (0.001 to 0.150)	0.768
$\rho_B \sim U(0,1)$	0.584 (0.359 to 0.826)	0.681 (0.404 to 0.995)	0.042 (0.001 to 0.177)	0.037 (0.001 to 0.187)	0.561 (0.035 to 0.983)	0.771
Logit( $\rho_B$ ) $\sim N(0, \text{sd}=0.8)$	0.581 (0.360 to 0.821)	0.682 (0.400 to 1.001)	0.038 (0.001 to 0.149)	0.035 (0.001 to 0.177)	0.522 (0.183 to 0.841)	0.772

## 5 Discussion

In a meta-analysis of multiple effects, a multivariate approach can jointly synthesise the endpoints and account for any correlation between the effects that may exist both within and between studies (5, 33). This leads to borrowing of strength and thus potentially different and stronger conclusions than separate univariate analyses, and therefore within a Bayesian bivariate meta-analysis framework it is crucial for prior distributions to be selected with care. This paper has explored the choice of prior distributions for the between-study variances and correlation in a Bayesian bivariate random-effects meta-analysis within a simulation study and a real example. The key recommendations are summarised in Box 1, and now briefly discussed.

### 5.1 Key findings

In current applications of multivariate meta-analysis, the  $U(-1,1)$  distribution is often selected for the between-study correlation without a sensitivity analysis (1, 17, 18), perhaps assuming it is vague. However, this work illustrates that the choice of prior distribution for  $\rho_B$  is often highly informative of posterior conclusions for all parameters of interest, especially when there are few studies in the meta-analysis, or missing outcome data. Even with large numbers of studies, such as 50, the prior distribution still noticeably influences the posterior distribution for the between-study correlation, which can impact upon the amount of borrowing of strength, joint inferences and subsequently clinically important measures such as the summary treatment effects and probabilistic statements. Therefore, a major recommendation is that the prior distribution for the between-study correlation must be

chosen carefully in future multivariate meta-analyses, and the commonly chosen  $U(-1,1)$  prior distribution is not always appropriate.

Though appropriate estimation of the between-study correlation is important in complete data settings (especially when making joint inferences across the multiple outcomes), it is even more critical in missing data settings. The prior distribution is more informative of the posterior distribution for this parameter since there is less data to estimate the between-study correlation, and the correlation itself has more impact on the borrowing of strength, which is usually greater in missing data settings (5). Therefore a sensible prior distribution for the between-study correlation is desired. External sources of data, such as similar systematic reviews, could be used to construct plausible prior distributions for this parameter (21, 26, 29). If related data is unavailable, a clinically relevant range of values for the prior distribution could still be specified. For example, if the meta-analysis pools overall and progression-free survival, it may be clinically plausible that the correlation is restricted to positive values only, and therefore a  $U(0,1)$  prior distribution may be more realistic than a  $U(-1,1)$  distribution. Alternatively, if a meta-analysis is used to pool sensitivity and specificity estimates from diagnostic test studies, the correlation could be restricted to negative values, and the  $U(-1,0)$  prior distribution may be a sensible choice. If negative (or positive) values are highly unlikely but not implausible, then a distribution might be used that allows all values but with most probability given to positive (or negative) values (see Supplementary Material). If there is no prior information to inform a more realistically vague prior distribution then the  $U(-1,1)$  distribution appears the most sensible choice. However, a sensitivity analysis that considers alternative prior distributions would then be especially important.

The choice of prior distribution for the between-study correlation and the between-study variances are not independent, and therefore wise choices must be made for all parameters in the bivariate meta-analysis model. Where previous simulation work has illustrated the importance of the prior distribution for the between-study variance in a univariate meta-analysis (8), the simulation studies in this paper reveal this is also true for a bivariate meta-analysis. If an inappropriate prior distribution is selected for the between-study variance, this not only has an impact on the posterior estimates of  $\tau_j$  themselves, but also on the posterior estimate of between-study correlation, the pooled treatment effect estimates, the amount of borrowing of strength and subsequently joint inferences. Therefore, previously derived empirical prior distributions (9-11) should be considered for the between-study variance parameters in a multivariate setting. The use of Gamma or Wishart prior distributions should be avoided; our simulation study shows this may introduce bias in the posterior estimates of the between-study variances and correlation, which then may influence the subsequent meta-analysis results and borrowing of strength. This was previously noted as a potential concern by Achana et al.(41) in a single application of network meta-analysis of multiple treatments and outcomes. However, Wishart prior distributions are still being suggested by some researchers, for example in a recent tutorial for undertaking Bayesian bivariate meta-analyses (42).

## 5.2 Limitations

Whilst many different prior distributions were examined, there are, of course, numerous other prior distributions that could be used, but were not considered here. Furthermore, the simulation study was specifically for bivariate meta-analysis, but there may be more than two

correlated outcomes. In this case, there are several more between-study variance-covariance parameters that require prior distributions. However, the findings are likely to generalise.

Finally, the simulation results (e.g. bias, MSE, coverage) are essentially a frequentist evaluation of a Bayesian analysis, which some may argue may not be appropriate. In particular, Senn previously suggested that it is perhaps philosophically incorrect to conduct a simulation study to assess the performance of Bayesian prior distributions because it is “irrelevant to any Bayesian who truly believed what the prior distribution represented” (43). Though this is an important statement, the rationale for the simulation study here is similar to that of Lambert et al. who justify that, “if a statistician desires to have a model with good bias and coverage properties, but needs/wants to use Bayesian methods, then we believe that simulation is a very good way of establishing this” (44).

### **5.3 Conclusions**

The simulation study and the illustrative example revealed that the choice of prior distribution for the between-study correlation in a Bayesian multivariate random-effects meta-analysis is important and must be chosen with caution, and in conjunction with suitable choices of prior distributions for the between-study variances. Ideally, empirical prior distributions should be utilised for the between-study variance parameters, and external clinical evidence used to inform realistically vague prior distribution of the between-study correlation. This is especially important for multivariate meta-analysis involving missing data, where the correlation dictates the amount of borrowing of strength from indirect information, and when joint inferences are desired across the multiple effects of interest. Often, sensitivity analyses to the choice of all prior distributions will be essential.

Box 1 summarizes recommendations for future Bayesian multivariate random-effects meta-analyses.

**Box 1: Key recommendations for specification of prior distributions for between-study variances and correlation in a Bayesian bivariate meta-analysis.**

- The use of a Wishart prior distribution on the entire between-study variance-covariance matrix is best avoided; it can be highly influential toward posterior meta-analysis results. Rather, a separate prior distribution should be specified for the between-study variances and the correlation.
- The prior distributions for between-study variances need to be chosen sensibly as they strongly impact on parameter estimates including the between-study correlation, and thus can influence the amount of borrowing of strength and subsequent joint inferences. For this purpose, empirical prior distributions may be most useful, such as those by Turner et al. and Rhodes et al. (9, 10). The use of an inverse Gamma prior distribution is best avoided.
- The prior distribution for between-study correlation also needs to be chosen sensibly, as it may have large influence toward the amount of borrowing of strength and joint inferences, especially when the number of studies providing both outcomes is few and the between-study variation is relatively small. A  $U(-1,1)$  prior distribution for  $\rho_B$  is not always vague and thus should not be routinely used without due thought, even when the number of studies is large (say, 50) it can have an important influence when the true correlation is large.
- Clinical, biological or methodological rationale might provide external evidence to inform a more realistically vague prior distribution for the between-study correlation. For example, a  $U(0,1)$  prior distribution could be specified if only positive values are plausible, such as prognostic effects that are partially and fully adjusted, or treatment effects on two highly correlated outcomes like systolic and diastolic blood pressure. A  $U(-1,0)$  prior distribution

might be specified if only negative values are plausible, for example for sensitivity and specificity from multiple studies that use different thresholds.

- Sensitivity analysis for the choice of prior distributions on between-study variances and correlations may be needed, especially when external evidence to inform the prior distributions is not available, borrowing of strength is potentially large (due to missing data), and there is relatively small information from the likelihood to inform their posterior distributions (for example, when the number of studies in the meta-analysis is small, and the between-study variance is small relative to the within-study variances).



## BIBLIOGRAPHY

1. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009;172(4):789-811.
2. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *StatMed*. 2011;30(20):2481-98.
3. Riley RD, Thompson JR, Abram KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172-86.
4. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *StatMethods MedRes*. 2013;22(2):133-58.
5. Riley RDA, K.R.; Lambert, P.C.; Sutton,A.J.; Thompson,J.R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*. 2007;26:78-97.
6. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *StatMed*. 2012;31(20):2179-95.
7. Jackson D, White IR, Price M, et al. Borrowing of strength and study weights in multivariate and network meta-analysis. *Stat Methods Med Res*. 2015.
8. Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of prior distributions in MCMC using WinBUGS. *Statistics in Medicine*. 2005;24:2401-28.
9. Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol*. 2014;68(1):52-60.
10. Turner RM, Davey J, Clarke MJ, et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *IntJEpidemiol*. 2012;41(3):818-27.
11. Pullenayegum EM. An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Stat Med*. 2011;30(26):3082-94.
12. Riley RD, Dodd SR, Craig JV, et al. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med*. 2008;27(29):6111-36.

13. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. *Stat Med*. 2008;27(5):687-97.
14. Bujkiewicz S, Thompson JR, Spata E, et al. Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints. *Stat Methods Med Res*. 2015.
15. Leonard TH, J.S. Bayesian inference for a covariance matrix. *Annals of Statistics*. 1992;20(4):1669-96.
16. Daniels MJK, R.E. Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*. 1999;94(448):1254-63.
17. Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*. 2013;32:1191-205.
18. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *StatMed*. 2003;22(14):2309-33.
19. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
20. Snell KI, Hua H, Debray TP, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol*. 2015.
21. Bujkiewicz S, Thompson JR, Sutton AJ, et al. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Statistics in Medicine*. 2013;32:3926-43.
22. Bujkiewicz S, Thompson JR, Riley RD, et al. Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Stat Med*. 2015.
23. Riley RP, MJ.; Jackson, D.; Wardle, M.; Gueyffier, F.; Wang, J.; Staessen, JA.; White, IR. Multivariate meta-analysis using individual participant data. *Research Synthesis Methods*. 2015;6:157-74.
24. Trikalinos TA, Hoaglin DC, Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. *StatMed*. 2014;33(9):1441-59.
25. Wei Y, Higgins JPT. Estimating the within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*. 2013;32:1191-205.
26. Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis*. 2006;1:515.

27. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: John Wiley & Sons; 2004.
28. Neuenschwander B, Capkun-Niggli G, Branson M, et al. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5-18.
29. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*. 1996;15(24):2733-49.
30. Menke J. Bayesian bivariate meta-analysis of sensitivity and specificity: summary of quantitative findings in 50 meta-analyses. *J Eval Clin Pract*. 2014.
31. Yang B, Shi J, Chen X, et al. Efficacy and safety of therapies for acute ischemic stroke in China: a network meta-analysis of 13289 patients from 145 randomized controlled trials. *PLoS One*. 2014;9(2):e88440.
32. Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med Res Methodol*. 2011;11:61.
33. Riley RD, Abrams KR, Sutton AJ, et al. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC MedResMethodol*. 2007;7:3.
34. Smyth RM, Kirkham JJ, Jacoby A, et al. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ*. 2011;342:c7153.
35. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349:g6501.
36. Frosi G, Riley RD, Williamson PR, et al. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. *J Clin Epidemiol*. 2015;68(5):542-50.
37. Inc SI. SAS/STAT<sup>®</sup> 9.3 User's Guide. Cary, NC: SAS Institute Inc; 2011.
38. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2009;172(Part 1):137-59.
39. Symons MJ, Moore DT. Hazard rate ratio and prospective epidemiological studies. *J Clin Epidemiol*. 2002;55(9):893-9.
40. Pernerger TV. Estimating the relative hazard by the ratio of logarithms of event-free proportions. *Contemp Clin Trials*. 2008;29(5):762-6.

41. Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol.* 2014;14:92.
42. Gajic-Veljanoski O, Cheung AM, Bayoumi AM, et al. A tutorial on Bayesian bivariate meta-analysis of mixed binary-continuous outcomes with missing treatment effects. *Stat Med.* 2015.
43. Senn S. Trying to be precise about vagueness. *Stat Med.* 2007;26(7):1417-30.
44. Lambert PC, Sutton AJ, Burton PR, et al. Comments on 'trying to be precise about vagueness' by Stephen Senn, *Statistics in Medicine* 2007; 26:1417-1430. *Stat Med.* 2008;27(4):619-22, author reply 22-4.

## Figure legends

Figure 1: Density plots for prior distributions for between-study correlation:

(a)  $\rho_B \sim U(-1,1)$  (Option 1); (b)  $\frac{1}{2} \log \left( \frac{1+\rho_B}{1-\rho_B} \right) \sim N(0, \text{sd}=0.4)$  (Option 2);

(c)  $\frac{(\rho_B+1)}{2} \sim \text{Beta}(1.5,1.5)$  (Option 3); (d)  $\rho_B \sim U(0,1)$  (Option 4);

(e)  $\text{logit}(\rho_B) \sim N(0, \text{sd}=0.8)$  (Option 5).

Figure 2: Posterior mean and 95% CrI for between-study correlation for various prior distributions in the illustrative example.

Figure 3: Posterior median and 95% CrI for between-study variances and between-study correlation for the two selected prior distributions for the between-study variances.

Figure 4: Posterior median and 95% CrI for between-study variance for fully-adjusted logHR for various priors for between-study correlation.

Figure S1: Density plot for prior distribution for between-study correlation:  $\frac{1}{2} \log \left( \frac{1+\rho_B}{1-\rho_B} \right) \sim N(0.47, \text{sd}=0.27)$ .