Multivariate meta-analysis of mixed outcomes: a Bayesian approach

Sylwia Bujkiewicz,[a*†] **John R. Thompson,**[b] **Alex J. Sutton,**[a] **Nicola J. Cooper,**[a] **Mark J. Harrison,**[c] **Deborah P. M. Symmons**[d] **and Keith R. Abrams**[a]

Multivariate random effects meta-analysis (MRMA) is an appropriate way for synthesizing data from studies reporting multiple correlated outcomes. In a Bayesian framework, it has great potential for integrating evidence from a variety of sources. In this paper, we propose a Bayesian model for MRMA of mixed outcomes, which extends previously developed bivariate models to the trivariate case and also allows for combination of multiple outcomes that are both continuous and binary. We have constructed informative prior distributions for the correlations by using external evidence. Prior distributions for the within-study correlations were constructed by employing external individual patent data and using a double bootstrap method to obtain the correlations between mixed outcomes. The between-study model of MRMA was parameterized in the form of a product of a series of univariate conditional normal distributions. This allowed us to place explicit prior distributions on the between-study correlations, which were constructed using external summary data. Traditionally, independent 'vague' prior distributions are placed on all parameters of the model. In contrast to this approach, we constructed prior distributions for the between-study model parameters in a way that takes into account the inter-relationship between them. This is a flexible method that can be extended to incorporate mixed outcomes other than continuous and binary and beyond the trivariate case. We have applied this model to a motivating example in rheumatoid arthritis with the aim of incorporating all available evidence in the synthesis and potentially reducing uncertainty around the estimate of interest. © 2013 The Authors. Statistics in Medicine Published by John Wiley & Sons, Ltd.

Keywords: Bayesian analysis; multivariate meta-analysis; multiple outcomes; rheumatoid arthritis

1. Introduction

When the synthesis of data from studies reporting multiple outcomes is required, multivariate random effects meta-analysis (MRMA) can be used instead of performing meta-analyses on each outcome separately, which has an advantage of taking into account the correlations between the outcomes. Methods for MRMA have been developed for a number of purposes, for example, to estimate multiple outcomes from clinical trials [1], to model relationships between surrogate endpoints [2], or to evaluate diagnostic tests [3]. Attention has mainly focussed on frequentist approaches, but methods, which use a Bayesian framework, have also been developed [2, 4]. In a recent review, Jackson *et al.* [5] describe advances in the development of the methodology of multivariate meta-analysis and discuss the advantages and disadvantages of the use of these methods. One of the advantages of integrating data on multiple outcomes

[a]*Biostatistics Research Group, Department of Health Sciences, University of Leicester, University Road, Leicester, LE1 7RH, U.K.*

[b]*Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, University Road, Leicester, LE1 7RH, U.K.*

[c]*Health Economics, Health Sciences - Methodology Research Group, School of Community Based Medicine, The University of Manchester, Jean McFarlane Building, Manchester, M13 9PL, U.K.*

[d]*NIHR Manchester Musculoskeletal Biomedical Research Unit, Arthritis Research UK Epidemiology Unit, School of Translational Medicine, University of Manchester, Manchester, M13 9PT, U.K.*

*Correspondence to: Sylwia Bujkiewicz, Department of Health Sciences, University of Leicester, University Road, Leicester, LE1 7RH, U.K.

†E-mail: sb309@le.ac.uk

through MRMA is that of 'borrowing of strength' across studies as well as across outcomes, which can potentially lead to reduced uncertainty around the resulting effectiveness estimates. In a Bayesian framework, a wider range of sources of evidence can be integrated. For example, additional data (external data from observational studies, clinical trials, or systematic reviews) or experts' opinions can be incorporated in the form of prior distributions, which can further inform a multivariate meta-analysis model. The use of external data can potentially lead to a further reduction of uncertainty around the estimates. The main advantage of Bayesian analysis, however, is that it enables an analyst to combine evidence from multiple sources, which is an important factor in evidence-based medicine [6–8].

The MRMA models typically have a hierarchical structure in which the correlated multiple outcomes estimate underlying true effects (within-study model) and the multiple true effects are correlated and follow the same common distribution (between-study model). The elements of the within-study covariance matrices are assumed to be known; however, in practice, often only the variances are available. The correlations between the estimates (due to sampling variability), however, are rarely reported. The within-study correlations could be obtained if individual patient data (IPD) for all of the studies in MRMA were available, but this is hardly ever the case, and therefore, estimation of the within-study correlation is one of the major challenges in multivariate meta-analysis. Previous studies, which addressed this issue, included, for example, those suggesting the use of the average correlation obtained from the subset of those studies, which listed IPD [2], implementation of an alternative formulation of bivariate meta-analysis for studies with unknown within-study correlation, which combined covariances from both the within-study and the between-study model in a single term [9], or recently, a method of approximating the within-study covariances, based on the bivariate delta method, was developed [10]. In contrast to the within-study covariance, the between-study covariance matrix is estimated from the model using study level data [11]. However, in a Bayesian framework, the between-study covariance matrix or its elements require prior distributions to be placed on them (as do all the parameters: the between-study variances and correlations as well as the pooled effects).

Some of the previously developed Bayesian MRMA models were restricted to the bivariate case [2, 12, 13]. Nam *et al.* [4] developed more general MRMA models; however, the examples used were also restricted to two outcomes. Arends *et al.* [14] proposed extension to trivariate meta-analysis, but the focus was mostly on a frequentist approach, and the use of Bayesian methods was only briefly discussed. In this article, we introduce a novel approach to MRMA in a Bayesian framework, which can be implemented for an extended number of outcomes and is presented here for a trivariate random-effects meta-analysis (TRMA) of mixed (continuous and binary) outcomes. This model allows us to incorporate data from a variety of sources, by not only including in the TRMA studies that are a mixture of those reporting all or only some of the outcomes but also by incorporating external data in the form of prior distributions. The use of informative prior distributions also extends previously developed methods, which used noninformative prior distributions, such as uniform distributions placed on correlations and inverse gamma distributions placed on variances, or Wishart prior distributions for precision matrices in models by Nam *et al.* [4]. By contrast, we use external sources of data to construct informative prior distributions for both the within-study and between-study correlations. In our model, we assume that none of the studies in TRMA report the within-study correlations. We employ external IPD to construct prior distributions for the within-study correlations. To obtain the distributions of the correlations between mixed outcomes from the IPD, we used a double bootstrap method [15]. We also describe methods of parameterization of the between-study model, which with an appropriate choice of prior distributions, allow the explicit incorporation of external information into the model through the Bayesian framework. Independent prior distributions are typically placed on the parameters of the multivariate model [2], but perhaps more appealingly, they can be placed on the correlations and other parameters in such a way as to account for their interdependence.

We apply this meta-analytical framework to an example from rheumatoid arthritis (RA). We carried out a meta-analysis of tumor necrosis factor-alpha inhibitors used sequentially in patients treated with RA. In studies assessing long-term effectiveness and/or cost-effectiveness of treatments in RA, the Health Assessment Questionnaire (HAQ) score, a self-reported measure of physical function, is of main interest as it is commonly used to estimate quality of life of patients following the treatment [16]. In a systematic review, Lloyd *et al.* [17] found that very few studies reported the HAQ score, but there were other instruments used to assess patients' responses to treatment. In the multivariate evidence synthesis models, we take into account data from studies reporting the HAQ score, as well as studies reporting other measures of response to treatment with the aim of integrating all available evidence and potentially decreasing uncertainty around the HAQ. We have devised a Bayesian framework in which external

IPD is used to construct prior distributions for the within-study correlations between the HAQ and the alternative outcomes whereas external summary data (ESD) from a systematic review allows us to derive prior distributions for the between-study correlations.

The outline of this article is as follows. In Section 2, we introduce the motivating example and describe the data and the logic behind the meta-analysis model. Section 3 contains the details of trivariate meta-analysis model, including methods of constructing the prior distributions for the correlations. In Section 4, we briefly describe the implementation of these models in WinBUGS [18]. We include the results of applying the model to the RA example in Section 5, which is followed by the discussion in Section 6. We describe further technical details, including the description of the multivariate model for any number of outcomes, in the Appendices.

2. Motivating example

2.1. Systematic review: 'Lloyd data'

A systematic review and meta-analysis was carried out [17] to investigate the effectiveness of tumor necrosis factor-alpha inhibitors such as etanercept, infliximab, and adalimumab used as second line treatment in patients with RA. Standard instruments for measuring response to treatment in RA were considered: the HAQ and Disease Activity Score (DAS-28) measures, and the American College of Rheumatology (ACR) response criteria. The results of the meta-analyses of all three outcomes individually showed that the biologic interventions are effective when used sequentially. Data collected in this systematic review are used to investigate how multivariate meta-analysis can be applied to incorporate multiple outcomes, such as 20% response according to the ACR criteria (ACR20; a binary outcome used on log odds scale in MRMA) and changes from baseline of DAS-28 and HAQ scores (continuous outcomes), in evidence synthesis, which aims to estimate the change from baseline of the HAQ score. The estimate of the HAQ is of a particular interest to clinicians and decision-makers in health care as it

Table I. Studies in 'Lloyd data' reporting outcomes: 20% response according to the American College of Rheumatology criteria, Disease Activity Score, and Health Assessment Questionnaire.

Study	ACR20 r/n	DAS-28 Mean* (se)	HAQ Mean* (se)
Bennet 2005	—	-1.7 $(0.25)^{\ddagger}$	-0.31 $(0.13)^{\ddagger}$
Bingham 2009	85/188	-1.6 $(0.1)^{\ddagger}$	$-0.35^{\dagger}$ $(0.05)^{\ddagger}$
Bombardieri 2007	486/810	-1.9 (0.05)	-0.48 (0.02)
Buch 2005	18/25	—	—
Buch 2007	55/72	-1.47 $(0.18)^{\ddagger}$	—
Cohen 2005	—	$-1.87^{\dagger}$ $(0.24)^{\ddagger}$	—
Di Poi 2007	—	$-2.1^{\dagger}$ $(0.29)^{\ddagger}$	—
Finckh 2007	—	-0.98 $(0.18)^{\ddagger}$	—
Haroui 2004	14/22	—	$-0.45^{\dagger}$ $(0.14)^{\ddagger}$
Hjardem 2007	—	-1 $(0.11)^{\ddagger}$	—
Hyrich 2008	—	—	-0.12 $(0.03)^{\ddagger}$
Iannone 2009	—	—	$0.15^{\dagger}$ $(0.13)^{\ddagger}$
Karlsson 2008	172/337	—	—
Laas (InTol) 2008	—	$-1.17^{\dagger}$ $(0.66)^{\ddagger}$	—
Laas (InEff) 2008	—	$-1.26^{\dagger}$ $(0.35)^{\ddagger}$	—
Navarro-Sarabia 2009	—	-1.1 $(0.18)^{\ddagger}$	-0.21 $(0.07)^{\ddagger}$
Nikas 2006	18/24	$-2.4^{\dagger}$ $(0.16)^{\ddagger}$	—
Van der Bijl 2008	19/41	-1.5 (0.25)	-0.21 (0.08)
Van Vollenhoven 2003	12/18	—	—
Wick (EA) 2005	7/9	-1.9 $(0.22)^{\ddagger}$	—
Wick (IA) 2005	19/27	-1.3 $(0.28)^{\ddagger}$	—

ACR20, 20% response according to the American College of Rheumatology criteria; DAS-28, Disease Activity Score; HAQ, Health Assessment Questionnaire; se, standard error.
*mean change from baseline; †mean values were obtained from other measures; ‡standard errors were obtained from other measures or imputed as described in Appendix 2 of [17].

is often used to estimate quality of life of patients following treatments of RA. Table I gives details of the three outcomes, which were reported in each of the studies within the systematic review. As indicated in the footnote of the table, standard errors (and occasionally the mean values) were often obtained from other measures or imputed. These unreported values could be treated as unknown in a Bayesian analysis. However, for the simplicity and exposition of this model, we chose to treat them as constant. The details of how these values were obtained can be found in Appendix 2 of the original meta-analysis of these data [17]. We will refer to these data as the Lloyd data throughout this paper.

2.2. External individual patient data

None of the studies included in the Lloyd data reported correlations between outcomes, whereas such study-level correlations are required to fully specify a model of correlated outcomes. These correlations cannot be obtained directly from summary data, such as the Lloyd data, and none of the studies listed the IPD; thus, external data were required to estimate the correlations.

The IPD was obtained from the British Rheumatoid Outcome Study Group (BROSG) trial, which was designed to assess the benefit of aggressive disease-modifying antirheumatic drug treatments in patients with established RA, conducted by Symmons *et al.* [19]. It was a randomized trial, which recruited 466 patients with stable RA. The trial assessed clinical outcomes (i.e., HAQ, DAS-28, and ACR20) in two cohorts of patients managed using either a regime focussed on symptomatic control of pain and stiffness in the shared care setting or a more aggressive regime focussing on control of symptoms and joint inflammation in the hospital setting.

The BROSG trial found no difference between the aggressive versus symptomatic treatment arms. Therefore, this data set may be used as a single cohort of 466 patients with established RA whose condition deteriorated modestly over a 3 year period. The data from the BROSG trial were used to construct prior distributions for the within-study correlations between outcomes in the multivariate models. Given the complexity of the models in this paper, only those 293 patients for whom all three clinical outcomes (HAQ, DAS-28, and ACR20) were reported were included in the analysis (to avoid further data imputation).

2.3. External summary data

In contrast to the within-study correlations, we can estimate the between-study correlations from the model using study level summary data. In a Bayesian framework, this could entail placing noninformative prior distributions on these correlations. However, one of the advantages of Bayesian approach we want to exploit here is the possibility of incorporating external information in the analysis. We can achieve this by using external information to construct informative prior distributions. To illustrate this, we conducted an *ad hoc* meta-analysis of ESD to obtain estimates of the between-study correlations, which can be used as prior distributions in our models (as described in Sections 3.4 and 3.5). The ESD included studies of the same type of treatment as in the Lloyd data but used as the first-line treatment. Supporting Information[‡] include further details with the full list of studies included in the ESD.

2.4. Logic of the meta-analysis model and notation

In our motivating example, we aim to model the summary data of the correlated outcomes from the Lloyd data using a multivariate meta-analysis in a Bayesian form. To do so, we need to place prior distributions on the within-study and the between-study correlations (which are not known in the Lloyd data). We use the IPD, described in Section 2.2, to construct the prior distributions for the within-study correlations and the ESD, described in Section 2.3, to construct the prior distributions for the between-study correlations. Figure 1 illustrates this data structure and the role of each element within it. We use the external data to construct the prior distributions for the within-study and between-study correlations only. The remaining parameters of the model, such as the pooled effects and the between-study standard deviations, are given noninformative prior distributions [6]. Note that the external data set used in this example was not very large. However, in more general circumstances, the relevance and rigor of the external evidence can be taken into account. For example, the variance of the prior distribution can be adjusted to construct a less informative distribution [6, 20]. In addition, when there are multiple external data sources, we can carry out a random effects meta-analysis. A number of authors have advocated

---

[‡]*Supporting information may be found in the online version of this article.*
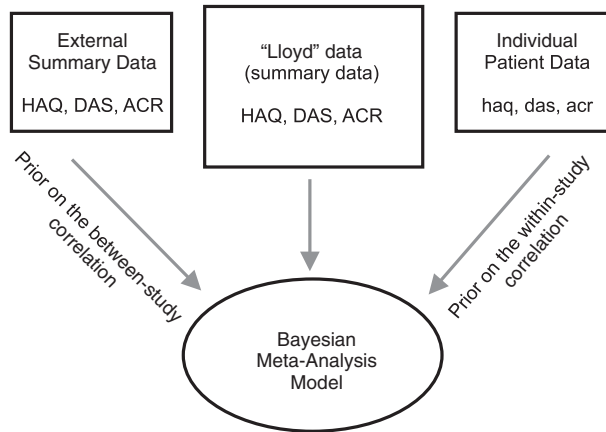
Figure 1. Structure of the data and the role of the data elements in the model.

using posterior predictive distribution from such external meta-analysis as a source of external evidence in the form of a prior distribution [6, 21].

3. Trivariate random-effects meta-analysis

For the purpose of simplicity and direct link to the Lloyd data, the model presented here includes only three outcomes. The full multivariate model is described in Appendix A. Suppose that we have summary data available on at least one of three outcomes (Y_1, Y_2, and Y_3). To be able to combine data from studies reporting these outcomes, we can carry out a meta-analysis of the three outcomes simultaneously. If we assume that all three outcomes are normally distributed (i.e., they are continuous outcomes or binomial outcomes on the log scale as for example log odds of response), we can model them simultaneously in the TRMA model (by extending bivariate models discussed by van Houwelingen *et al.* [22] and Riley *et al.* [23]) as follows:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \end{pmatrix}, \boldsymbol{\Sigma_i} \right), \boldsymbol{\Sigma_i} = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} & \sigma_{1i}\sigma_{3i}\rho_{wi}^{13} \\ \sigma_{2i}\sigma_{1i}\rho_{wi}^{12} & \sigma_{2i}^2 & \sigma_{2i}\sigma_{3i}\rho_{wi}^{23} \\ \sigma_{3i}\sigma_{1i}\rho_{wi}^{13} & \sigma_{3i}\sigma_{2i}\rho_{wi}^{23} & \sigma_{3i}^2 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \mathbf{T} \right), \mathbf{T} = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b^{12} & \tau_1\tau_3\rho_b^{13} \\ \tau_2\tau_1\rho_b^{12} & \tau_2^2 & \tau_2\tau_3\rho_b^{23} \\ \tau_3\tau_1\rho_b^{13} & \tau_3\tau_2\rho_b^{23} & \tau_3^2 \end{pmatrix} \quad (2)$$

In the previous model, we assume outcomes Y_{1i}, Y_{2i}, and Y_{3i} to be estimates of correlated effects μ_{1i}, μ_{2i}, and μ_{3i} with corresponding within-study covariance matrices $\boldsymbol{\Sigma_i}$ of the estimates. These study-level effects follow a trivariate normal distribution with means $(\beta_1, \beta_2, \beta_3)$ and covariance $\mathbf{T}$ in this hierarchical framework. We will refer to (1) as the within-study model and (2) as the between-study model. Prior distributions need to be specified for missing data (which we describe in Section 3.1), the within-study correlations (constructed in Section 3.2), and the elements of the between-study covariance matrix. The latter we propose to construct for an alternative parameterization of the between-study model as described in Section 3.3 with the choice of prior distributions discussed in Section 3.4 and finally constructed in Section 3.5.

Note that the bivariate random-effects meta-analysis (BRMA) is a special case of TRMA in (1)–(2), which can be obtained simply by reducing the dimension of TRMA, as described in Appendix B. If all the within-study correlations in TRMA, that is, the within-study correlations ρ_{wi}^{12}, ρ_{wi}^{13}, and ρ_{wi}^{23}, and the between-study correlations ρ_b^{12}, ρ_b^{13}, and ρ_b^{23} are equal to zero, then the model (1)–(2) reduces to three univariate random-effects meta-analysis (URMA) models for Y_{1i}, Y_{2i}, and Y_{3i}:

$$\begin{aligned} Y_{ji} &\sim N\left(\mu_{ji}, \sigma_{ji}^2\right), \\ \mu_{ji} &\sim N\left(\beta_j, \tau_j^2\right), \ j = 1, 2, 3 \end{aligned} \quad (3)$$

which in a Bayesian framework require specification of prior distributions such as the normal prior distribution for the mean effect $\beta_j \sim N(0, 1000)$ and the half normal for the between-study standard deviation $\tau_j \sim N(0.0, 1000)I(0,)$ (where $N(-, -)I(-,)$ denotes the half normal distribution [6]). Examples of alternative prior distributions for URMA can be found elsewhere [24].

3.1. Missing data

In our model, we assume that not all the studies report all the outcomes Y_1, Y_2, and Y_3; that is, there are missing data for at least one of the outcomes for some of the studies. The advantage of conducting this meta-analysis in a Bayesian framework is that we give all the outcomes (whether known or missing) a distribution, which leads to these missing values being estimated directly from the model through the Markov Chain Monte Carlo (MCMC) simulation. Defining the trivariate model, as in (1), already provides distributions for missing values of Y_{1i}, Y_{2i}, and Y_{3i}. However, corresponding missing standard deviations σ_{1i}, σ_{2i}, and σ_{3i} still need to be estimated. By assuming exchangeability of the variances, we can assume the corresponding population variances (rather than the variances of the mean) to come from the same distribution, for example,

$$
\begin{aligned}
\text{var}_{1(2,3)i} &\sim N(0, h_{1(2,3)})I(0,), \\
h_{1(2,3)} &\sim \Gamma(1.0, 0.01)
\end{aligned}
\tag{4}
$$

and $\sigma_{1i}^2 = \text{var}_{1i}/N_i$, $\sigma_{2i}^2 = \text{var}_{2i}/N_i$, and $\sigma_{3i}^2 = \text{var}_{3i}/N_i$. By 'borrowing of information' from the studies reporting the Y_{1i} (Y_{2i}, Y_{3i}) and their standard deviations, the variances (and hence standard deviations) for studies with missing Y_{1i} (Y_{2i}, Y_{3i}) are predicted by the MCMC simulation, conditional upon both the data and the posterior estimates of the model parameters.

3.2. Prior distributions for the within-study correlations

Clinical studies very rarely report the correlations between clinical outcomes, that is, the within-study correlations ρ_{wi}^{12}, ρ_{wi}^{13}, and ρ_{wi}^{23}; therefore, in the model (1), we assume them to be unknown. In the Bayesian framework, however, we must give the correlations prior distributions. These prior distributions can be either noninformative (such as a uniform distribution over range -1 to 1) or more informative prior distributions potentially based on external information, for example, IPD containing observations on all outcomes from another, but relevant, study such as the IPD described in Section 2.2.

Suppose we have data available from an external study with individual participants for whom data on all three outcomes have been collected. We want to use IPD to construct prior distributions for the within-study correlations ρ_{wi}^{12}, ρ_{wi}^{13}, and ρ_{wi}^{23}. If all the outcomes in IPD were continuous, we could assume that they follow a common trivariate normal distribution and modeling them this way would allow us to evaluate the within-study correlation directly from the covariance matrix. Often, however, outcomes are a mixture of continuous ones and others, for example, binary ones (as is the case of ACR20, in our motivating example, which in the summary data is taken as log odds of ACR20 response, but in IPD, it is a binary outcome). Then, we cannot assume that the three variables follow a common trivariate normal distribution (unlike in the case of aggregate data in (1), we are unable to estimate log odds of response for individual patients).

There are two ways of overcoming this problem of constructing the prior distributions for the within-study correlations in TRMA. One is to use an approximation where the correlation between the mean of the continuous variable and the log odds of the binary one equals the correlation between the continuous observation and the binary observation. Another approach is to carry out a double bootstrap analysis. We have chosen the latter as it may be applied to other types of mixed outcomes. Double bootstrap methods have been developed to, for example, estimate confidence intervals [15, 25]. We have carried out a simulation, described in Appendix C, to show that this method can be applied to estimate summary statistics and the correlations between them.

We applied the double bootstrap method in the following way. From the IPD (containing observations on all three outcomes), we sampled $N_{b1} = 500$ bootstrap samples of size equal to that of the IPD, allowing for repetitions (first level bootstrap samples). From each of the bootstrap samples, we sampled another $N_{b2} = 500$ bootstrap samples (second level bootstrap samples). For each of the second level bootstrap samples, we calculated the mean values of the continuous variables and log odds of the binary ones. For the sets of summary statistics corresponding to each of the first level bootstrap samples, we

then calculated the correlations between them. We used the simulated series of 500 sets of the three correlations (in the form of the 500×3 matrix) as a set of empirical prior distributions for the within-study correlations. In our main TRMA model, we sampled rows of this matrix (containing three correlations) to preserve the between-outcome correlation structure to ensure the nonsingularity of the within-study covariance matrices. In our example, the between-study variability exceeds the variability within the included studies; therefore, the estimates of the within-study correlations have little impact on the final results of TRMA [13], and hence, the same prior distributions for all within-study correlations in each of the studies could be used as in Nam et al [4] where the correlations do not vary with the study suffix i. However, in a more general scenario, this may not be appropriate. Therefore, for each of the studies in meta-analysis, we sample the three within-study correlations from different copies of the empirical prior distribution, thus allowing ρ_{wi} to have a different value for each study i, in every MCMC iteration.

Note that, as mentioned previously, if all three variables were continuous, it would be possible to estimate the within-study correlation directly (without the need for the bootstrap simulation). This is described for the bivariate version of the model in Appendix B.

3.3. Parameterization of the between-study model

For the between-study model (2), we need to specify prior distributions for the between-study variances τ_1^2, τ_2^2, and τ_3^2, and the between-study correlations ρ_b^{12}, ρ_b^{13}, and ρ_b^{23}. One approach, often adopted, is to give the inverse of the covariance matrix in (2), that is, the precision matrix $\mathbf{T}^{-1}$, a Wishart prior distribution [26]. An alternative approach is to parameterize the between-study model (2) in the product normal formulation, that is, in the form of a product of a series of univariate conditional normal distributions [27]. We chose this latter approach as it is more intuitive and allows us to use an explicit formulation of prior distributions for ρ_b^{12}, ρ_b^{13}, and ρ_b^{23} on the basis of external evidence. This approach is easier to implement, however, it requires making an additional assumption of independence of Y_2 and Y_3 conditional on Y_1. This assumption means that elements $(2, 3)$ and $(3, 2)$ of the precision matrix are zero as the partial correlation coefficient between Y_2 and Y_3 is zero [28]. We express this formulation as follows

$$\begin{cases} \mu_{1i} \sim N\left(\eta_1, \psi_1^2\right) \\ \mu_{2i} \mid \mu_{1i} \sim N\left(\eta_{2i}, \psi_2^2\right) \\ \eta_{2i} = \lambda_{20} + \lambda_{21}\left(\mu_{1i} - \overline{\mu_{1i}}\right) \\ \mu_{3i} \mid \mu_{1i} \sim N\left(\eta_{3i}, \psi_3^2\right) \\ \eta_{3i} = \lambda_{30} + \lambda_{31}\left(\mu_{1i} - \overline{\mu_{1i}}\right) \end{cases} \tag{5}$$

which can then be used to estimate the parameters of the between-study model (2):

$$\begin{cases} \beta_1 = \eta_1 \\ \tau_1^2 = \psi_1^2 \\ \beta_2 = \lambda_{20} \\ \tau_2^2 = \psi_2^2 + \lambda_{21}^2 \psi_1^2 \\ \rho_b^{12} = \dfrac{\lambda_{21}\psi_1^2}{\sqrt{\psi_1^2\left(\psi_2^2 + \lambda_{21}^2 \psi_1^2\right)}} \\ \beta_3 = \lambda_{30} \\ \tau_3^2 = \psi_3^2 + \lambda_{31}^2 \psi_1^2 \\ \rho_b^{13} = \dfrac{\lambda_{31}\psi_1^2}{\sqrt{\psi_1^2\left(\psi_3^2 + \lambda_{31}^2 \psi_1^2\right)}} \end{cases} \tag{6}$$

Note that the variable μ_{1i} in (5) has been centered to avoid high autocorrelation in the MCMC simulation.

3.4. Choice of the prior distributions for the between-study correlations

The formulae in (6) show the interdependencies between the parameters (i.e., the correlations, regression coefficients, and the standard deviations). Because they are inter-related, placing prior distributions on such parameters requires caution to ensure that they are plausible and realistic. For example, placing noninformative prior distributions on the standard deviations ψ_1, ψ_2, and ψ_3 and the regression coefficients
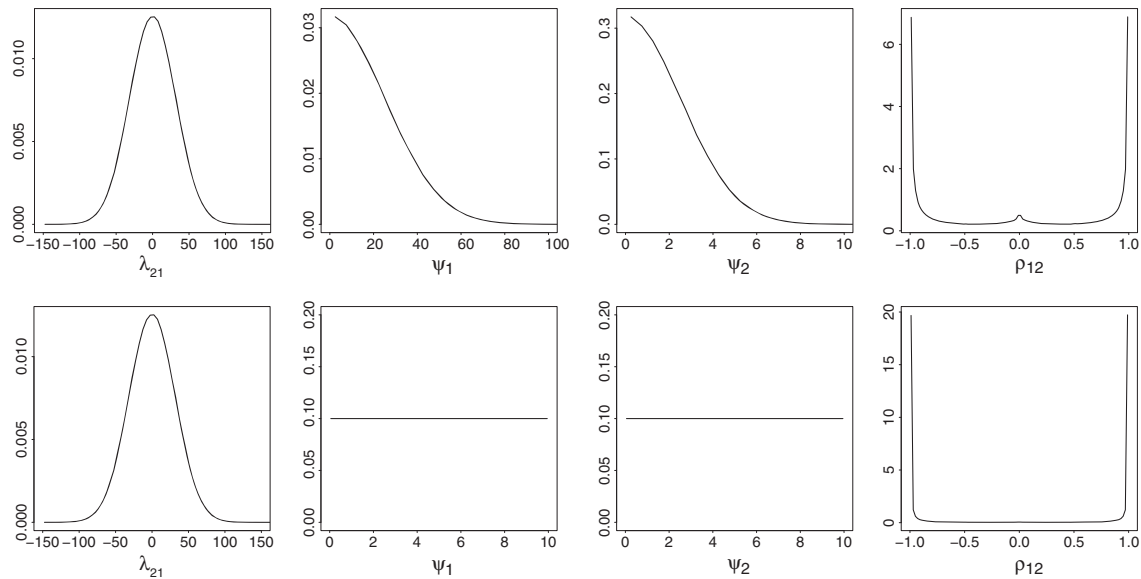
Figure 2. Examples of constructing prior distribution for the between-study correlation using independent noninformative prior distributions for the standard deviations and the regression coefficient, which can lead to implausible prior distribution for the correlation.

λ_{20}, λ_{21}, λ_{30}, and λ_{31} in (5) may appear to be an appropriate approach; however, it can result in inadequately constructed prior distributions for the between-study correlations, ρ_b^{12}, ρ_b^{13} and ρ_b^{23}, as shown in Figure 2 for the correlation between Y_1 and Y_2, ρ_b^{12}. For example, using a normal prior distribution for the regression coefficient λ_{21} and an independent half normal distributions for the between-study standard deviations ψ_1 and ψ_2 can lead to a W-shaped prior distribution for the correlation (top row in Figure 2), whereas using a normal prior distribution for the regression coefficient and independent uniform distributions for the between-study standard deviations can lead to a U-shaped prior distribution for the correlation (bottom row in Figure 2). This choice of the hyper parameter for the prior distribution leads to sampling from extreme values of the correlation, also leading to poor convergence.

We found that in this case, it is more suitable to give the correlations an appropriate prior distribution first. Defining them together with the prior distributions for the between-study standard deviations ψ_1, ψ_2, and ψ_3 enables evaluation of the regression coefficients λ_{21} and λ_{31} from (6):

$$\lambda_{21} = \frac{\psi_2}{\psi_1} \frac{\rho_b^{12}}{\sqrt{1 - \left(\rho_b^{12}\right)^2}} \quad \text{and} \quad \lambda_{31} = \frac{\psi_3}{\psi_1} \frac{\rho_b^{13}}{\sqrt{1 - \left(\rho_b^{13}\right)^2}} \tag{7}$$

Figure 3 shows examples of such parameterizations of the prior distributions. Using an informative prior distribution for the between-study correlation (such as a normal prior distribution for the Fisher transformation of a correlation) and 'vague'; for example, half normal distributions for the standard deviations ψ_1 and ψ_2 gives a more plausible implied prior distribution for the regression coefficient λ_{21} (top row in Figure 3). Another possible approach is to use the vague uniform prior distribution for the correlation as well as the between-study standard deviations, which also gives a plausible prior distribution for the regression coefficient (bottom row in Figure 3). Placing prior distributions first on ρ_b^{12} and ρ_b^{13} and on the between-study standard deviations ψ_1, ψ_2, and ψ_3 and then estimating the implied prior distributions for the coefficients λ_{21} and λ_{31} largely improves the convergence of all the parameters, that is, the between-study correlations ρ_b^{12} and ρ_b^{13} and the coefficients λ_{21} and λ_{31}. We give the coefficients λ_{20} and λ_{30} noninformative prior distributions as they do not affect the correlations.

3.5. Construction of the prior distributions for the between-study correlations

We can use external data to obtain the prior distributions for the between-study correlations. We can obtain such data, for example, from a systematic review of studies (i.e., involving patients with the same condition and under comparable treatment) reporting all outcomes such as the ESD introduced in Section 2.3. A trivariate meta-analysis (1)-(2) is carried out on the three outcomes in ESD simultaneously in the same way as the main TRMA described in this section. Here, however, we choose to
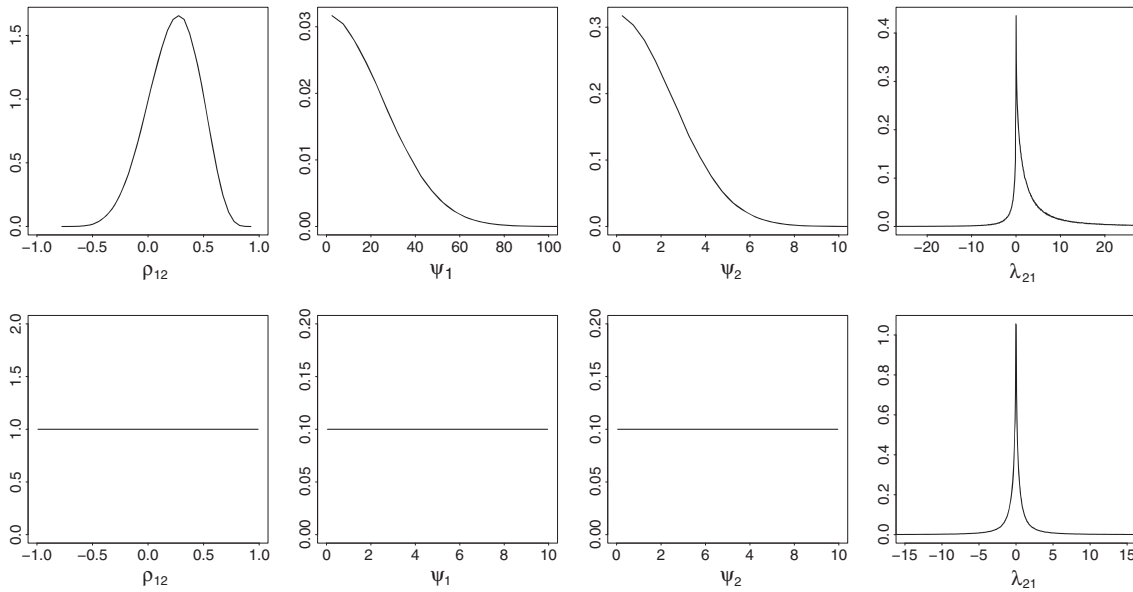
Figure 3. Examples of constructing prior distributions for the between-study model parameters using interdependent prior distributions for the parameters.

include only studies reporting all three outcomes; hence, there will be no missing data. We use vague prior distributions for the within-study and between-study correlations and the between-study standard deviations (obtaining informative prior distributions for these parameters would require another external data meta-analysis leading to an infinite process). We can use directly the posterior between-study correlations from this ESD analysis, ρ_{ESD}^{12} and ρ_{ESD}^{13}, as the prior distributions for ρ_b^{12} and ρ_b^{13} in (7). In the MCMC simulation of the main TRMA model, we sampled pairs of ρ_b^{12} and ρ_b^{13} (thinned out to 500 pairs) to preserve the correlation structure between the conditionally dependent variables.

Having defined all prior distributions, we can obtain relevant estimates (posterior β_1, β_2, and β_3 and the corresponding variances and correlations) from MCMC of the trivariate hierarchical model (defined in (1) and (2)).

4. Implementation in WinBUGS

We implemented all models in WinBUGS [18] where the estimates were obtained using MCMC simulation using 100,000 iterations (including 50,000 burn-in). We checked convergence by visually assessing the history, chains, and autocorrelation using graphical tools in WinBUGS and using the Geweke method in the BOA package [29]. We present all posterior estimates as means with the 95% highest probability density intervals (HPDIs) except for the estimates displayed in the forest plots, which are means with the 95% credible intervals (CrI).

5. Results

We have applied the TRMA model to the Lloyd data to simultaneously model data on the HAQ, DAS-28, and ACR20 (with the IPD and ESD used to construct the prior distributions for the within-study and the between-study correlations between the outcomes, as described in Sections 3.2 and 3.5). In the Lloyd data, the HAQ and DAS-28 estimates represent the mean change from baseline of the outcomes, and ACR20 is a proportion of responders, which is transformed to log odds of response to be able to follow the normal distribution in TRMA. To investigate the impact of including more data in the analysis on uncertainty around the HAQ estimate, we explored results of the meta-analyses on three levels: using URMA of HAQ (as in (3)), bivariate random-effects meta-analysis (BRMA) combining the HAQ and the DAS-28 (as in (B1)-(B2) in Appendix B), and finally, TRMA by extending the data by ACR20. Table II lists the parameters for the prior within-study and between-study correlations used in both BRMA and TRMA. Table III shows the results obtained from all the models.

Table II. Prior correlations.

	Within-study correlations			Between-study correlations	
	ρ_w^{DH}	ρ_w^{AH}	ρ_w^{DA}	ρ_b^{DH}	ρ_b^{AH}
BRMA	0.24 [0.13, 0.35]	—	—	0.86 [0.46, 0.999]	—
TRMA	0.24 [0.10, 0.38]	−0.13 [−0.29, 0.0103]	−0.20 [−0.31, −0.08]	0.78 [0.27, 0.998]	−0.14 [−0.80, 0.56]

BRMA, bivariate random-effects meta-analysis; TRMA, trivariate random-effects meta-analysis.

Table III. Results of the univariate meta-analyses of Health Assessment Questionnaire (HAQ), Disease Activity Score (DAS-28), and 20% response according to the American College of Rheumatology criteria (ACR20) separately, bivariate meta-analysis of HAQ and DAS-28, and trivariate of HAQ, DAS-28, and ACR20.

	Posterior mean (Standard error) [95% HPDI]				
	Univariate analyses			Bivariate	Trivariate
	HAQ	DAS-28	ACR20	HAQ & DAS-28	HAQ, DAS-28, & ACR20
HAQ	−0.25 (0.09) [−0.43, −0.09]	—	—	−0.28 (0.07) [−0.41, −0.14]	−0.28 (0.08) [−0.43, −0.12]
DAS	—	−1.57 (0.13) [−1.84, −1.31]	—	−1.51 (0.08) [−1.67, −1.35]	−1.52 (0.09) [−1.71, −1.34]
ACR	—	—	0.62 (0.05) [0.53, 0.71]	—	0.61 (0.05) [0.52, 0.71]
τ_H	0.21 (0.09) [0.08, 0.38]	—	—	0.21 (0.07) [0.10, 0.35]	0.22 (0.09) [0.10, 0.39]
τ_D	—	0.44 (0.11) [0.25, 0.67]	—	0.44 (0.11) [0.24, 0.67]	0.44 (0.11) [0.25, 0.67]
τ_A	—	—	0.52 (0.19) [0.20, 0.90]	—	0.53 (0.19) [0.21, 0.91]
ρ_b^{DH}	—	—	—	0.89 (0.12) [0.65, 0.994]	0.83 (0.18) [0.45, 0.99]
ρ_b^{AH}	—	—	—	—	−0.14 (0.30) [−0.63, 0.49]

HPDI, highest probability density interval.

5.1. Results for bivariate random-effects meta-analysis of Health Assessment Questionnaire and Disease Activity Score

We carried out BRMA of HAQ and DAS-28 with the prior distributions for the within-study correlations estimated from the IPD (methods described in Appendix B) and for the between-study correlation estimated from the ESD (as in Section 3.5). The BRMA allowed us to include 10 cohorts (from the eight studies reporting DAS-28 but not HAQ in Table I) in addition to those eight that could be used in URMA of HAQ (studies reporting HAQ in Table I). The posterior correlation between DAS-28 and HAQ obtained from the IPD (used as a prior distribution for the within-study correlation in each study, as discussed in Section 2.2) was relatively weak, with mean $\rho_{wi}^{das,haq} = 0.24$ (95% HPDI [0.13, 0.35]). The posterior correlation between DAS-28 and HAQ obtained from BRMA of the ESD (used here as a prior distribution for the between-study correlation) had a higher mean; $\rho_b^{DH} = 0.86$ (95% HPDI [0.46, 0.999]).

The HAQ estimate shifted toward a more extreme value when using BRMA (change in HAQ from baseline shifts from -0.25 in URMA to -0.28 in BRMA) with uncertainty reduced by 25% of the width of the 95% HPDI compared with the interval from URMA. The estimate of DAS-28, however, moved to a less extreme value when using BRMA (from -1.57 to -1.51), with reduced uncertainty. Adding more data to the m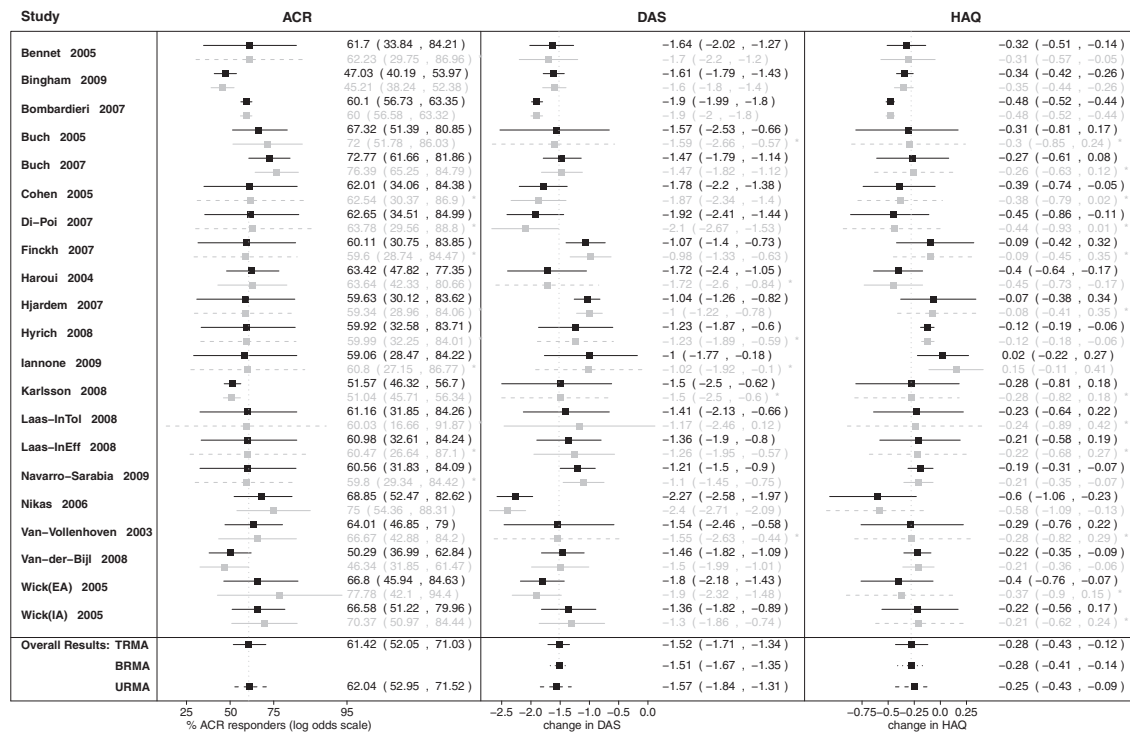eta-analysis by using BRMA reduced uncertainty but not heterogeneity; the between-study variances τ_H and τ_D remain almost the same, with only reduced uncertainty for τ_H. However, the direct comparison of the variances with those from URMA is confounded by the different prior distributions being used: half normal prior distributions are used for τ_H and τ_D in URMAs, whereas in BRMA, half normal prior distributions are used for ψ_H and ψ_D (ψ_1 and ψ_2 in (5) for TRMA); hence, the implied probabilities for τ_H and τ_D are different.

Figure 4 shows three forest plots representing estimates of the HAQ from URMA (left) and BRMA (middle), and DAS-28 from BRMA (right). As in all forest plots (Figures 4 and 5), black solid lines correspond to the 'shrunken' estimates μ_{Hi} and pooled estimates β_H, the grey solid lines show the estimates obtained from the systematic review (data used in this meta-analysis), and the grey dashed lines (estimates also marked with a $*$ on their right) correspond to the predicted estimates for the studies that did not report the HAQ but reported the DAS-28 (or reverse in plots for the DAS-28). Dashed and dotted black lines below the pooled estimates represent the pooled estimates obtained from meta-analyses of reduced numbers of outcomes, for example, estimates from URMA below the estimates obtained from BRMA for comparison.

In the forest plot of the estimates from URMA, the estimates are shrunken toward the mean, which is especially noticeable for those studies with higher uncertainty around the known estimates (i.e., Bennet, Haroui, and Iannone). For example, in the case of the Haroui study, the HAQ estimate of -0.45 (95% confidence interval (CI) $[-0.73, -0.17]$) is now shifted toward the overall mean
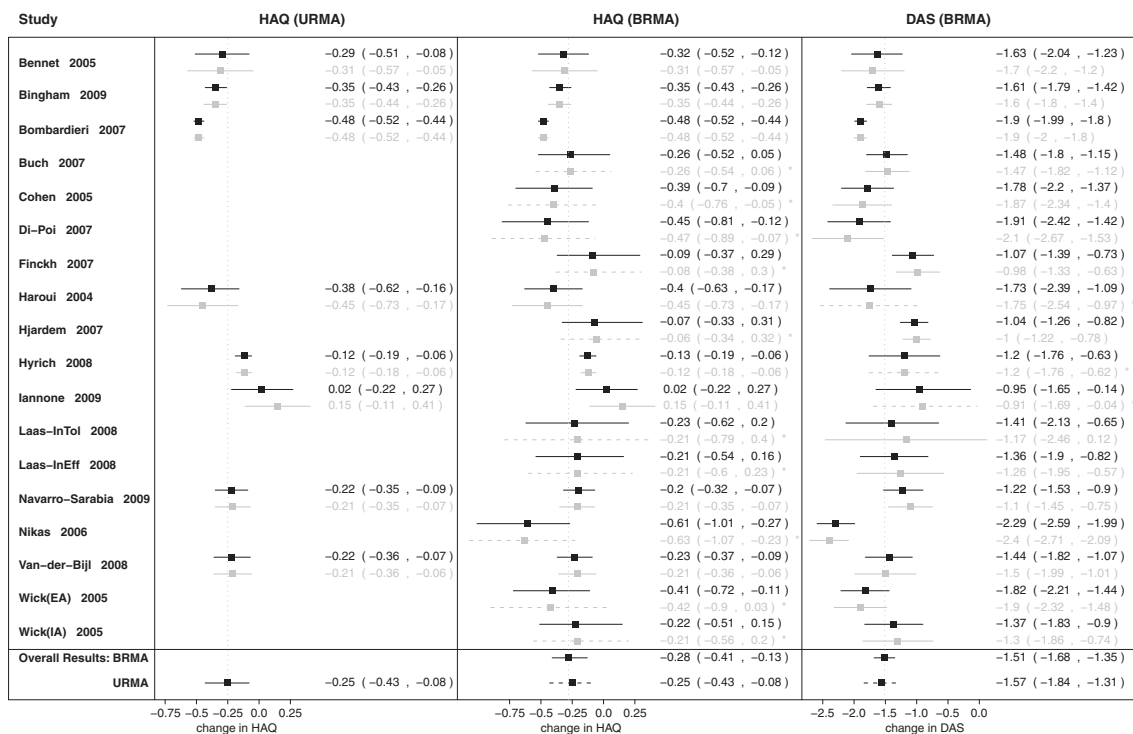
Figure 4. Forest plots for Health Assessment Questionnaire (HAQ): from univariate random-effects meta-analysis (URMA; left) and from bivariate random-effects meta-analysis (BRMA) of HAQ and Disease Activity Score (DAS-28; middle) and for DAS-28 also from BRMA (right). Graph shows estimates from the systematic review with 95% confidence intervals (grey solid lines), predicted missing estimates from BRMA with 95% credible intervals (CrIs; grey dashed lines), 'shrunken' estimates with 95% CrIs (black solid lines), and the pooled estimates with 95% CrIs (black solid lines for pooled effect from each of the meta-analyses and black dashed lines representing results from URMA for comparison).

Study	ACR	DAS	HAQ
Bennet 2005	61.7 (33.84, 84.21) / 62.23 (29.75, 86.96)	−1.64 (−2.02, −1.27) / −1.7 (−2.2, −1.2)	−0.32 (−0.51, −0.14) / −0.31 (−0.57, −0.05)
Bingham 2009	47.03 (40.19, 53.97) / 45.21 (38.24, 52.38)	−1.61 (−1.79, −1.43) / −1.6 (−1.8, −1.4)	−0.34 (−0.42, −0.26) / −0.35 (−0.42, −0.26)
Bombardieri 2007	60.1 (56.73, 63.35) / 60 (56.58, 63.32)	−1.9 (−1.99, −1.8) / −1.9 (−2, −1.8)	−0.48 (−0.52, −0.44) / −0.48 (−0.52, −0.44)
Buch 2005	67.32 (51.39, 80.85) / 72 (51.78, 86.03)	−1.57 (−2.53, −0.66) / −1.59 (−2.66, −0.57)	−0.31 (−0.81, 0.17) / −0.3 (−0.85, 0.24)
Buch 2007	72.77 (61.66, 81.86) / 76.39 (65.25, 84.79)	−1.47 (−1.79, −1.14) / −1.47 (−1.82, −1.12)	−0.27 (−0.61, 0.08) / −0.26 (−0.63, 0.12)
Cohen 2005	62.01 (34.06, 84.38) / 62.54 (30.37, 86.9)	−1.78 (−2.2, −1.38) / −1.87 (−2.34, −1.4)	−0.39 (−0.74, −0.05) / −0.38 (−0.79, 0.02)
Di−Poi 2007	62.65 (34.51, 84.99) / 63.78 (29.56, 88.8)	−1.92 (−2.41, −1.44) / −2.1 (−2.67, −1.53)	−0.45 (−0.86, −0.11) / −0.44 (−0.93, 0.01)
Finckh 2007	60.11 (30.75, 83.85) / 59.6 (28.74, 84.47)	−1.07 (−1.4, −0.73) / −0.98 (−1.33, −0.63)	−0.09 (−0.42, 0.32) / −0.09 (−0.45, 0.35)
Haroui 2004	63.42 (47.82, 77.35) / 63.64 (42.33, 80.66)	−1.72 (−2.4, −1.05) / −1.72 (−2.6, −0.84)	−0.4 (−0.64, −0.17) / −0.45 (−0.73, −0.17)
Hjardem 2007	59.63 (30.12, 83.62) / 59.34 (28.96, 84.06)	−1.04 (−1.26, −0.82) / −1 (−1.22, −0.78)	−0.07 (−0.38, 0.34) / −0.08 (−0.41, 0.35)
Hyrich 2008	59.92 (32.58, 83.71) / 59.99 (32.25, 84.01)	−1.23 (−1.87, −0.6) / −1.23 (−1.89, −0.59)	−0.12 (−0.19, −0.06) / −0.12 (−0.19, −0.06)
Iannone 2009	59.06 (28.47, 84.22) / 60.9 (27.15, 86.77)	−1 (−1.77, −0.18) / −1.02 (−1.92, −0.1)	0.02 (−0.22, 0.27) / 0.15 (−0.11, 0.41)
Karlsson 2008	51.57 (46.32, 56.7) / 51.04 (45.71, 56.34)	−1.5 (−2.5, −0.62) / −1.5 (−2.5, −0.6)	−0.28 (−0.81, 0.18) / −0.28 (−0.82, 0.2)
Laas−InTol 2008	61.16 (31.85, 84.26) / 60.03 (16.66, 91.87)	−1.41 (−2.13, −0.66) / −1.17 (−2.46, 0.12)	−0.23 (−0.64, 0.22) / −0.24 (−0.89, 0.42)
Laas−InEff 2008	60.98 (32.61, 84.24) / 60.47 (26.64, 87.11)	−1.36 (−1.9, −0.8) / −1.26 (−1.95, −0.57)	−0.21 (−0.58, 0.19) / −0.22 (−0.68, 0.27)
Navarro−Sarabia 2009	60.56 (31.83, 84.09) / 59.8 (29.34, 84.42)	−1.21 (−1.5, −0.9) / −1.1 (−1.45, −0.75)	−0.19 (−0.31, −0.07) / −0.21 (−0.35, −0.07)
Nikas 2006	68.85 (52.47, 82.62) / 75 (54.36, 88.31)	−2.27 (−2.58, −1.97) / −2.4 (−2.71, −2.09)	−0.6 (−1.06, −0.23) / −0.58 (−1.09, −0.13)
Van−Vollenhoven 2003	64.01 (46.85, 79) / 66.67 (42.88, 84.2)	−1.54 (−2.46, −0.58) / −1.55 (−2.63, −0.44)	−0.29 (−0.76, 0.22) / −0.28 (−0.82, 0.29)
Van−der−Bijl 2008	50.29 (36.99, 62.84) / 46.34 (31.85, 61.47)	−1.46 (−1.82, −1.09) / −1.5 (−1.99, −1.01)	−0.22 (−0.35, −0.09) / −0.21 (−0.36, −0.06)
Wick(EA) 2005	66.8 (45.94, 84.63) / 77.78 (42.1, 94.4)	−1.8 (−2.18, −1.43) / −1.9 (−2.32, −1.48)	−0.4 (−0.76, −0.07) / −0.37 (−0.9, 0.15)
Wick(IA) 2005	66.58 (51.22, 79.96) / 70.37 (50.97, 84.44)	−1.36 (−1.82, −0.89) / −1.3 (−1.86, −0.74)	−0.22 (−0.5, 0.17) / −0.21 (−0.62, 0.24)
Overall Results: TRMA	61.42 (52.05, 71.03)	−1.52 (−1.71, −1.34)	−0.28 (−0.43, −0.12)
BRMA		−1.51 (−1.67, −1.35)	−0.28 (−0.41, −0.14)
URMA	62.04 (52.95, 71.52)	−1.57 (−1.84, −1.31)	−0.25 (−0.43, −0.09)

25 50 75 95
% ACR responders (log odds scale)

−2.5 −2.0 −1.5 −1.0 −0.5 0.0
change in DAS

−0.75 −0.5 −0.25 0.0 0.25
change in HAQ

Figure 5. Forest plots for estimates of (from left to right) 20% response according to the American College of Rheumatology (ACR) criteria, Disease Activity Score (DAS-28), and Health Assessment Questionnaire (HAQ) from trivariate random-effects meta-analysis (TRMA) of HAQ, DAS-28, and ACR20. Graph shows estimates from the systematic review with 95% confidence intervals (grey solid lines), predicted missing estimates from TRMA with 95% credible intervals (CrIs; grey dashed lines), 'shrunken' estimates with 95% CrIs (black solid lines), and the pooled estimates with 95% CrIs (black solid lines for pooled effect from each of the TRMAs and black dotted (dashed) lines representing results from bivariate random-effects meta-analysis (univariate random-effects meta-analysis) for comparison).

(which for URMA is −0.26) with shrunken estimate equal −0.38 (95% CrI [−0.63, −0.15]). Borrowing of strength across studies leads to both the shift toward the mean and decrease in uncertainty reducing the width of the credible interval by 14% for the Haroui estimate. Note that the uncertainty corresponding to the pooled estimate is higher compared with the estimate obtained using the frequentist univariate meta-analysis [17]; −0.25 (95% CI [−0.4, −0.11]). This is due to the between-study variance having a probability distribution (fixed in the frequentist approach) adding to the overall uncertainty in the model.

In BRMA, in addition to borrowing of strength across studies, there is also borrowing of strength between outcomes. In the middle forest plot in Figure 4 (representing estimates of HAQ from BRMA), the predicted estimates contribute to the pooled estimate even though there is considerable uncertainty associated with them. This additional borrowing of information across outcomes leads to the aforementioned reduction in uncertainty around the pooled estimate of the HAQ.

As can be seen in Figure 4 (middle and right-hand-side forest plots of estimates from BRMA), the predicted estimates of the HAQ (DAS-28) follow the heterogeneity pattern of the corresponding known estimates from the DAS-28 (HAQ) because of the informative (and of high mean) prior between-study correlation $\rho_b^{DH} = 0.86$ (95% HPDI [0.46, 0.999]). Studies that reported the DAS-28, but not the HAQ, had on average relatively high positive estimates, which led to the high positive predicted estimates for the HAQ. Especially extreme values predicted for the HAQ were those for studies by Cohen, Di-Poi, Nikas, and one cohort of the Wick study, which had extreme values for the DAS-28. Also, two studies out of those three reporting the HAQ, but not the DAS-28, had more extreme estimates, showing little or no effect (only small improvement in HAQ in the Hyrich study and no improvement in Iannone), which led to the more extreme predicted estimates for the DAS-28 and, in consequence, reduced the pooled effect measured by the change from baseline of the DAS-28.

5.2. Results for trivariate random-effects meta-analysis of Health Assessment Questionnaire, Disease Activity Score, and 20% response according to the American College of Rheumatology criteria

We were able to include three further studies (which reported the ACR20, but not the HAQ or DAS-28 scores) by extending BRMA to TRMA. By incorporating an additional outcome, ACR20, we not only extended the data by those three studies but also incorporated the data on ACR20 from eight studies already in BRMA, which reported the HAQ and/or DAS-28 as well as ACR20 (the details can be found in Table I). Adding ACR20 in TRMA did not, however, lead to any change in the point estimate of HAQ or its uncertainty; they remained almost the same as in BRMA. The between-study heterogeneity increased slightly (τ_H increased from 0.21 to 0.22) after inclusion of the three studies reporting the ACR20. This is likely due to the relatively high between-study heterogeneity of studies reporting the ACR20 and the prior correlation between the ACR20 and the HAQ being of rather low mean and relatively noninformative.

Figure 5 shows forest plots for all three outcomes: the ACR20, DAS-28, and HAQ. The predicted estimates (grey dashed lines) for the HAQ and DAS-28 follow the heterogeneity pattern of the corresponding known estimates of the DAS-28 and HAQ, respectively, as these are highly correlated outcomes, with the posterior between-study correlation $\rho_b^{DH} = 0.83$ (95% HPDI [0.45,0.99]). However, the predicted estimates of the HAQ and DAS-28, for studies where neither of these outcomes were reported, are close to the overall mean with considerable associated uncertainty (as are the predicted estimates for ACR20 for studies reporting HAQ and/or DAS-28 but not ACR20). This is likely due to the lack of correlation between the ACR20 and the HAQ. This resulting level of uncertainty around the predicted estimates and the high between-study heterogeneity of studies reporting the ACR20 is the most likely explanation for the lack of a further reduction in uncertainty around HAQ when extending the analysis to TRMA. An additional forest plot showing all estimates of HAQ (obtained from all the models: URMA, BRMA, and TRMA) plotted together for comparison can be found in the Supporting Information.

6. Discussion

We have developed a Bayesian multivariate meta-analytic framework for incorporating data from multiple sources of evidence: studies reporting multiple mixed endpoints and additional external data, which are used to inform the parameters of the model in the form of prior distributions. The developed model gives a flexible way for combining data on mixed outcomes beyond the bivariate case. In this framework, a product normal formulation is adopted to describe the between-study variability of correlated outcomes. The advantage of this approach is the possibility of an explicit incorporation of external information to inform the prior distribution for the between-study correlation. However, as we have demonstrated, careful specification of the univariate prior distributions is essential so that the resultant prior distributions that are induced for the correlations are both plausible and realistic. We also show how individual patient data, consisting of mixed outcomes, can be used to construct empirical prior distributions for the within-study correlations. Our model comprised three parts as the prior distributions for the correlations resulted from the analyses of the two external data sets. It would be possible to carry out all components of the analysis within a single comprehensive Bayesian analysis. However, the flow of information would need to be restricted when putting the parts of the model together to prevent the main analysis from feeding back to, and influencing, the part of the model in which the empirical prior distributions were constructed [18]. In our model, we used directly the posterior distributions of the relevant correlations, obtained from the analysis of external data, as the empirical prior distributions for the correlations. By sampling directly from the posterior distributions, we avoided the potential problem of having to make distributional assumptions to summarize the prior external evidence.

It has been previously discussed [30] that applying a multivariate approach to meta-analysis makes little difference to the results compared with those from the univariate approach. However, our findings show that this is clearly not always the case, and the multivariate approach can be of practical importance. We found that employing this Bayesian approach to evidence synthesis of outcomes in RA by combining the HAQ endpoint with the DAS-28 leads to a significant reduction in the uncertainty around the HAQ from −0.25 (95% HPDI [−0.43,−0.09]) obtained from URMA to −0.28 (95% HPDI [−0.41,−0.14]) obtained from BRMA. We have obtained similar results when using TRMA of the HAQ, DAS-28, and ACR20. The inclusion of the third outcome, ACR20, did not contribute to the HAQ estimate or reduction of uncertainty around it.

We cannot predict the extent of the gain in the precision of the pooled estimate, and the inclusion of additional outcomes through a multivariate approach does not always lead to reduced uncertainty around the estimate of the outcome of interest. In fact, in our example, extending the bivariate meta-analysis to the three-dimensional case did not contribute to a further reduction in uncertainty around the clinical outcomes. The additional number of studies included in the meta-analysis, by extending the analysis to multiple outcomes, can also lead to additional heterogeneity and therefore potentially *increase* the level of uncertainty. Sturtz *et al.* [31], have discussed a similar effect in the context of mixed treatment comparison meta-analysis where extending a network (by inclusion of additional interventions and studies) can sometimes lead to increased uncertainty around the estimates because of increased heterogeneity. However, there are other potential advantages of using a multivariate approach. It has been recently suggested by Kirkham *et al.* [32] that under some circumstances, a multivariate approach to meta-analysis can lead to a more appropriate estimate of the clinical outcome in the presence of outcome reporting bias. This may apply to the case of RA (discussed previously) where, because of the availability of a number of instruments for measuring the disease activity or response to treatment, authors may choose to report only those outcomes whose estimates are positive and/or significant. Although, in our model, we make no assumption regarding outcome reporting bias, this multivariate approach may reduce the effect of such bias if it existed.

One of the limitations of this particular example in RA is that the IPD comes from a study with treatments, which are very different from those used in the Lloyd data. It is possible that not only is the treatment difference an issue, but also that patients did not respond particularly well to the treatment and there was relatively little variability between patients leading to a low correlation. However, despite this limitation, we chose to use these prior distributions as they serve the purpose of illustrating how IPD can be used to construct a set of empirical prior distributions for the whole correlation structure of the within-study correlations. However, it is not always possible to obtain relevant individual patient data. Nevertheless, the Bayesian approach has another advantage as it allows elicitation techniques to be used to incorporate experts' opinions in the form of prior distributions in a model [33]. For example, methods for inclusion of expert opinion on bias in the synthesis of studies have been developed [20].

We can apply the modeling framework presented in this article to a range of settings in evidence-based medicine. For example, alternative outcomes such as surrogate endpoints are increasingly considered in decision-making in health care, especially in the early stages of the drug development process when extended follow-up time is required to obtain the main clinical outcome of interest. In addition to the multivariate framework as a way of combining data on multiple correlated outcomes, the Bayesian approach gives another level of integrating diverse sources of evidence, as it allows for incorporation of external information in the form of prior distribution. In our view, this approach is an important step toward the coherent synthesis of multiple sources of evidence, which is vital in policy decisions in health care [7].

Appendix A. Multivariate random-effects meta-analysis

Suppose that we have summary data available on at least one of N outcomes $(Y_1, Y_2, \ldots, Y_N)$. If we assume that all outcomes are normally distributed (i.e., they are continuous outcomes or transformations of other outcomes), we can model them simultaneously in the MRMA model (by extending TRMA model described in Section 3) as follows:

$$
\begin{pmatrix} Y_{1i} \\ Y_{2i} \\ \vdots \\ Y_{Ni} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{Ni} \end{pmatrix}, \Sigma_i \right), \quad \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} & \cdots & \sigma_{1i}\sigma_{Ni}\rho_{wi}^{1N} \\ \sigma_{2i}\sigma_{1i}\rho_{wi}^{12} & \sigma_{2i}^2 & \cdots & \sigma_{2i}\sigma_{Ni}\rho_{wi}^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{Ni}\sigma_{1i}\rho_{wi}^{1N} & \sigma_{Ni}\sigma_{2i}\rho_{wi}^{2N} & \cdots & \sigma_{Ni}^2 \end{pmatrix} \tag{A1}
$$

$$
\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{Ni} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix}, T \right), \quad T = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b^{12} & \cdots & \tau_1\tau_N\rho_b^{1N} \\ \tau_2\tau_1\rho_b^{12} & \tau_2^2 & \cdots & \tau_2\tau_N\rho_b^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_N\tau_1\rho_b^{1N} & \tau_N\tau_2\rho_b^{2N} & \cdots & \tau_N^2 \end{pmatrix} \tag{A2}
$$

In the this model, outcomes $Y_{1i}, Y_{2i}, \ldots, Y_{Ni}$ are assumed to be estimates of correlated effects μ_{1i}, $\mu_{2i}, \ldots, \mu_{Ni}$ with corresponding within-study covariance matrices Σ_i of the estimates. These study-level effects follow a multivariate normal distribution with means $(\beta_1, \beta_2, \ldots, \beta_N)$ and covariance $\mathbf{T}$ in this hierarchical framework. Equations (A1) and (A2) describe the within-study and the between-study models, respectively.

A.1. The between-study model in the product normal formulation

Similarly as for TRMA, by assuming independence of $Y_2, \ldots, Y_N$ conditional on Y_1, we can parameterize the between-study model (A2) in the form of product normal formulation as follows:

$$
\begin{cases}
\mu_{1i} \sim N\left(\eta_1, \psi_1^2\right) \\
\mu_{2i} \mid \mu_{1i} \sim N\left(\eta_{2i}, \psi_2^2\right) \\
\eta_{2i} = \lambda_{20} + \lambda_{21}\left(\mu_{1i} - \overline{\mu_{1i}}\right) \\
\vdots \\
\mu_{Ni} \mid \mu_{1i} \sim N\left(\eta_{Ni}, \psi_N^2\right) \\
\eta_{Ni} = \lambda_{N0} + \lambda_{N1}\left(\mu_{1i} - \overline{\mu_{1i}}\right)
\end{cases}
\tag{A3}
$$

which can then be used to estimate the parameters of the between-study model (A2)

$$
\begin{cases}
\beta_1 = \eta_1 \\
\tau_1^2 = \psi_1^2 \\
\beta_2 = \lambda_{20} \\
\tau_2^2 = \psi_2^2 + \lambda_{21}^2 \psi_1^2 \\
\rho_b^{12} = \dfrac{\lambda_{21}\psi_1^2}{\sqrt{\psi_1^2\left(\psi_2^2 + \lambda_{21}^2\psi_1^2\right)}} \\
\vdots \\
\beta_N = \lambda_{N0} \\
\tau_N^2 = \psi_N^2 + \lambda_{N1}^2 \psi_1^2 \\
\rho_b^{1N} = \dfrac{\lambda_{N1}\psi_1^2}{\sqrt{\psi_1^2\left(\psi_N^2 + \lambda_{N1}^2\psi_1^2\right)}}
\end{cases}
\tag{A4}
$$

Rearranging the formulas in (A4) leads to the following relationships between the regression coefficients, standard deviations, and the between-study correlations:

$$
\lambda_{21} = \frac{\psi_2}{\psi_1}\frac{\rho_b^{12}}{\sqrt{1 - \left(\rho_b^{12}\right)^2}}, \quad \lambda_{31} = \frac{\psi_3}{\psi_1}\frac{\rho_b^{13}}{\sqrt{1 - \left(\rho_b^{13}\right)^2}}, \quad \ldots \text{ and } \lambda_{N1} = \frac{\psi_N}{\psi_1}\frac{\rho_b^{1N}}{\sqrt{1 - \left(\rho_b^{1N}\right)^2}}
\tag{A5}
$$

which can be used to define inter-related prior distributions in a similar manner as described in Section 3.4 for TRMA.

Appendix B. Bivariate random-effects meta-analysis

In the algebra in this section, we adopt the following notation. For the summary data, we denote the outcomes in the upper case: The variables Y_1 and Y_2 represent the mean values of the continuous outcomes. While when referring to the individual observations in Section B.1, we use the lower case: Variables y_1 and y_2 are observations for an individual.

We obtain the model for BRMA by reducing the dimension of TRMA in the following form:

$$
\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \Sigma_i\right), \quad \Sigma_i = \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} \\ \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} & \sigma_{2i}^2 \end{pmatrix}
\tag{B1}
$$

$$\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \mathbf{T} \right), \quad \mathbf{T} = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_b^{12} \\ \tau_1\tau_2\rho_b^{12} & \tau_2^2 \end{pmatrix} \tag{B2}$$

In this model, similarly as in TRMA, we assume outcomes Y_1 and Y_2 to be estimates of correlated effects μ_{1i} and μ_{2i} with corresponding covariance matrices $\boldsymbol{\Sigma}_i$. These study-level effects follow a bivariate normal distribution with means (β_1, β_2) and covariance $\mathbf{T}$. We predict the missing outcomes and standard deviations in the same manner as for TRMA (as described in Section 4.1). We can obtain the prior distributions for the within-study and the between-study correlations in the same manner as for TRMA. However, when both outcomes Y_1 and Y_2 are continuous, we can obtain the within-study correlation from an analytical model described below.

B.1. The within-study correlation

When both outcomes are continuous, we can obtain prior distribution for the within-study correlation by solving an analytical bivariate model for IPD, which will give a posterior distribution for the correlation between the patients' outcomes. We can then use this posterior distribution directly as a prior distribution for the within-study correlation.

Suppose we have data available from an external study with $j = 1, \ldots, J$ participants, for whom data on both outcomes have been collected. Because the outcomes are correlated, we can estimate both outcomes from a bivariate model, assuming that the two outcomes simultaneously follow a bivariate normal distribution (we use the lower case for y_1 and y_2 to distinguish the change from baseline for individuals in IPD from the mean values in aggregate data in meta-analysis, which have been denoted in the upper case):

$$\begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \Omega \right), \quad \Omega = \begin{pmatrix} \omega_1^2 & \omega_1\omega_2\rho_{\text{IPD}}^{12} \\ \omega_1\omega_2\rho_{\text{IPD}}^{12} & \omega_2^2 \end{pmatrix} \tag{B3}$$

where θ_1 and θ_2 are mean values for y_1 and y_2 and Ω is the covariance matrix with the study-level correlation ρ_{IPD}^{12}, whose estimate we will use to define prior distributions for the within-study correlations ρ_{wi}^{12} in (B1).

We can parameterize the model (B3) in a product normal formulation as

$$\begin{cases} y_{1j} \sim N\left(\gamma_1, \xi_1^2\right) \\ y_{2j} \mid y_{1j} \sim N\left(\gamma_{2j}, \xi_2^2\right) \\ \gamma_{2j} = \alpha_0 + \alpha_1 y_{1j} \end{cases} \tag{B4}$$

and prior distributions for this model are: $\gamma_1, \alpha_0, \alpha_1 \sim N(0, 0.001)$, $\xi_1, \xi_2 \sim U(0, 10)$, which are meant to be relatively noninformative. Estimates from this parameterization can then be used to obtain the parameters of the model (B3) as follows:

$$\begin{cases} \theta_1 = \gamma_1 \\ \omega_1^2 = \xi_1^2 \\ \theta_2 = \alpha_0 + \alpha_1\gamma_1 \\ \omega_2^2 = \xi_2^2 + \alpha_1^2\xi_1^2 \\ \rho_{\text{IPD}}^{12} = \dfrac{\alpha_1\xi_1^2}{\sqrt{\xi_1^2\left(\xi_2^2 + \alpha_1^2\xi_1^2\right)}} \end{cases} \tag{B5}$$

We can use the posterior correlation ρ_{IPD}^{12} obtained from (B5) directly (in the same model and code) as a prior distribution for the within-study correlations ρ_{wi}^{12} in (B1) in a coherent way with no need to make further assumptions about the family of the distribution. Alternatively, we can obtain the prior correlation by calculating the Fisher transformation $FT = \frac{1}{2}\log\left(\frac{1+\rho_{\text{IPD}}^{12}}{1-\rho_{\text{IPD}}^{12}}\right)$. We can assume the Fisher transformed correlation to be normally distributed; $\rho^{FT} \sim N(FT, FT_{SE}^2)$, and then, we obtain the prior distribution for the between-study correlation by back-transforming ρ^{FT} into $\rho_{wi}^{12} = [\exp(2\rho^{FT}) - 1]/[\exp(2\rho^{FT}) + 1]$.

Appendix C. Simulation to test the double bootstrap method for summary statistics

We have carried out a simulation to show that a double bootstrap method [15, 25] can be applied to estimate correlation between summary statistics with uncertainty from individual subject data containing correlated observations. For the simulation, we chose to use the bivariate case of normally distributed data as in the normal case we know what the correlations between the summary statistics are: The correlation between the normally distributed individual observations equals the correlation between the means.

We simulated data, of size N, containing two correlated variables:

$$\begin{cases} y_1 \sim N\left(\mu_1, \xi_1^2\right) \\ y_2 \mid y_1 \sim N\left(\mu_2, \xi_2^2\right) \\ \mu_2 = \alpha_0 + \alpha_1 y_1 \end{cases} \tag{C1}$$

For these data, we calculated the correlation r and corresponding confidence interval by transforming the correlation (using the Fisher transformation):

$$Fcorr = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) \tag{C2}$$

We obtained the standard error using the formula: $Fcorr\_SE = 1/\sqrt{N-3}$, and then, the confidence interval for the transformed correlation is

$$Fcorr_{lci} = Fcorr - 1.96 * Fcorr\_SE \text{ and } Fcorr_{uci} = Fcorr + 1.96 * Fcorr\_SE \tag{C3}$$

This confidence interval is then back-transformed to obtain the confidence interval for the correlation:

$$r_{lci} = \frac{\exp(2Fcorr_{lci}) - 1}{\exp(2Fcorr_{lci}) + 1} \text{ and } r_{uci} = \frac{\exp(2Fcorr_{uci}) - 1}{\exp(2Fcorr_{uci}) + 1} \tag{C4}$$

Using the simulated data, we then performed a double bootstrap simulation in the following way. From the simulated data, we sampled N_{b1} bootstrap samples. From each of the bootstrap samples, we sampled another N_{b2} bootstrap samples (second level bootstrap samples). For each of the second level bootstrap samples, we calculated the mean values of the outcomes. We calculated the correlations between the means: $\rho_{Y_1Y_2}^{boot}$. We obtained confidence interval using the Fisher transformation in the same manner as for the correlation r previously. We then compared $\rho_{Y_1Y_2}^{boot}$ with r.

We ran this simulation for the following example: $N = 1000$, $y_1 \sim N(0, 5)$, $y_2 \sim N(2 + 4 * y_1, 10)$. We obtained the following results: $r = 0.946$, (95% CI: [0.939;0.952]) for the correlation obtained from the data directly and $\rho = 0.946$, (95% CI: [0.939;0.953]) for the correlation obtained from the double bootstrap simulation with $N_{b1/b2} = 5000$ bootstrap samples on both levels.

Acknowledgements

References

1. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**:2537–2550. DOI: 10.1002/(SICI)1097-0258 (19981130)17:22<2537::AID-SIM953>3.0.CO;2-C.
2. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; **16**:1965–1982. DOI: 10.1002/(SICI)1097-0258(19970915)16:17<1965::AID-SIM630>3.0.CO;2-M.
3. Harbord RM, Deeks JJ, Egger ME, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007; **8**:239–251. DOI: 10.1093/biostatistics/kxl004.

4. Nam I-S, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**:2309–2333. DOI: 10.1002/sim.1410.

5. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Statistics in Medicine* 2012; **30**:2481–2498. DOI: 10.1002/sim.4172.

6. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley and Sons, Ltd.: Chichester, 2004.

7. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine* 2003; **22**:3687–3709. DOI: 10.1002/sim.1586.

8. Ashby D, Smith AFM. Evidence-based medicine as Bayesian decision-making. *Statistics in Medicine* 2000; **19**: 3291–3305. DOI: 10.1002/1097-0258(20001215)19:23<3291::AID-SIM627>3.0.CO;2-T.

9. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when within-study correlations are unknown. *Biostatistics* 2008; **9**:172–186. DOI: 10.1093/biostatistics/kxm023.

10. Wei Y, Higgins JPT. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine* 2013; **32**:1191–1205. DOI: 10.1002/sim.5679.

11. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**:3. DOI: 10.1186/1471-2288-7-3.

12. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization – a miltivariate approach. *Statistics in Medicine* 2005; **24**:2241–2254. DOI: 10.1002/sim.2100.

13. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A* 2009; **172**:789–811. DOI: 10.1111/j.1467-985X.2008.00593.x.

14. Arends LR, Vokó Z, Stijnen T. You have full text access to this content combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* 2003; **22**:1335Ű-1353. DOI: 10.1002/sim.1370.

15. Beran R. Prepivoting to reduce level error in confidence sets. *Biometrika* 1987; **74**:457–68. DOI: 10.1093/biomet/74.3.457.

16. Bansback N, Ara R, Karnon J, Anis A. Economic evaluations in rheumatoid arthritis. A critical review of measures used to define health states. *Pharmacoeconomics* 2008; **26**:395–408.

17. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott D. The effectiveness of anti-TNF-alpha therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. *Rheumatology* 2010; **49**:2313–2321. DOI: 10.1093/rheumatology/keq169.

18. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS User Manual (Version 1.4), 2003. Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf [Accessed on August 16, 2012].

19. Symmons DPM, Tricker K, Harrison MJ, Roberts C, Davis M, Dawes P, Hassell A, Knight S, Mulherin D, Scott D. Patients with stable long-standing rheumatoid arthritis continue to deteriorate despite intensified treatment with traditional disease modifying anti-rheumatic drugs–results of the British Rheumatoid Outcome Study Group randomized controlled clinical trial. *Rheumatology* 2006; **45**:558–565. DOI: 10.1093/rheumatology/kei169.

20. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson S. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society A* 2009; **172**:21–47. DOI: 10.1111/j.1467-985X.2008.00547.x.

21. Ades AE, Lu G, Higgins JPT. The interpretation of random-effects meta-analysis in decision models. *Medical Decision Making* 2005; **25**:646–654. DOI: 10.1177/0272989X05282643.

22. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624. DOI: 10.1002/sim.1040.

23. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; **26**:78–97. DOI: 10.1002/sim.2524.

24. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**:277–303. DOI: 10.1177/096228020101000404.

25. Hall P, Martin MA. On bootstrap resampling and iteration. *Biometrika* 1988; **75**:661–71. DOI: 10.1093/biomet/75.4.661.

26. Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models (analytical methods for social research)*. Cambridge University Press: New York, 2007.

27. Spiegelhalter DJ. Bayesian graphical modelling: a case-study in monitoring health outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1998; **47**:115–133. DOI: 10.1111/1467-9876.00101.

28. Edwards D. *Introduction to Graphical Modelling*. Springer-Verlag: New York, 1995.

29. Smith BJ. boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* 2007; **21**(11):1–37.

30. Empirical and simulation-based comparison of univariate and multivariate meta-analysis for binary outcomes. *Draft Methods Report prepared for Agency for Healthcare Research and Quality*, 2012. Available from: http:// effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1182&pageaction =displayproduct [Accessed on August 16, 2012].

31. Sturtz S, Bender R. Unsolved issues of mixed treatment comparison meta-analysis: network size and inconsistency. *Research Synthesis Methods* 2012; **3**:300–311. DOI: 10.1002/jrsm.1057.

32. Kirkham JJ, Riley RD, Williamson PR. A miltivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine* 2012; **31**:2179–2195. DOI: 10.1002/sim.5356.

33. O'Hagan A, Buck CE, Daneshkhah A, Eiser JE, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T. *Uncertain Judgements: Eliciting Expert Probabilities*. Wiley: Chichester, 2006.