Research Article

Statistics in Medicine

Received 19 April 2010, Accepted 7 June 2011 Published online 10 August 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4327

Multivariate meta-analysis: a robust approach based on the theory of U-statistic

Yan Ma[a,b*†] and Madhu Mazumdar[b]

Meta-analysis is the methodology for combining findings from similar research studies asking the same question. When the question of interest involves multiple outcomes, multivariate meta-analysis is used to synthesize the outcomes simultaneously taking into account the correlation between the outcomes. Likelihood-based approaches, in particular restricted maximum likelihood (REML) method, are commonly utilized in this context. REML assumes a multivariate normal distribution for the random-effects model. This assumption is difficult to verify, especially for meta-analysis with small number of component studies. The use of REML also requires iterative estimation between parameters, needing moderately high computation time, especially when the dimension of outcomes is large. A multivariate method of moments (MMM) is available and is shown to perform equally well to REML. However, there is a lack of information on the performance of these two methods when the true data distribution is far from normality. In this paper, we propose a new nonparametric and non-iterative method for multivariate meta-analysis on the basis of the theory of U-statistic and compare the properties of these three procedures under both normal and skewed data through simulation studies. It is shown that the effect on estimates from REML because of non-normal data distribution is marginal and that the estimates from MMM and U-statistic-based approaches are very similar. Therefore, we conclude that for performing multivariate meta-analysis, the U-statistic estimation procedure is a viable alternative to REML and MMM. Easy implementation of all three methods are illustrated by their application to data from two published meta-analysis from the fields of hip fracture and periodontal disease. We discuss ideas for future research based on U-statistic for testing significance of between-study heterogeneity and for extending the work to meta-regression setting.

Keywords: meta-analysis; multiple correlated outcomes; random-effects model; U-statistic

1. Introduction

A systematic review is a literature review focused on a single question that tries to identify, appraise, select, and synthesize all high-quality research evidence relevant to that question. Meta-analysis is the quantitative extension of the systematic review and deals with statistical methods for examining the validity of the extracted summary statistics (effect size) from each component study, for quantifying the heterogeneity between the effect sizes, and at the end, for providing an estimate of the overall pooled effect size with optimal precision. With rising cost of medical research, many clinical trials are carried out with small sample sizes resulting in low power for detecting clinically useful effect size. This phenomenon has increased the chance of producing conflicting results from different studies. By pooling the estimated effect sizes from each of the component studies through meta-analytic methods, information from larger number of patients and increased power is expected [1]. Because of its important impact, choice of methods for performing meta-analysis has come under scrutiny and development of novel methods and softwares has become a rapidly growing field.

[a]*Research Division, Hospital for Special Surgery, New York, NY 10021, USA*
[b]*Division of Biostatistics and Epidemiology, Department of Public Health, Weill Medical College of Cornell University, New York, NY 10065, USA*
**Correspondence to: Yan Ma, Research Division, Hospital for Special Surgery, New York, NY 10021, USA.*
[†]*E-mail: yam2007@med.cornell.edu*

Univariate meta-analysis is used when a single outcome is of interest. Typically, a random-effects model [2] is described as

$$Y_i = \beta + b_i + \varepsilon_i;\ i = 1, 2, \ldots, n \tag{1}$$

where for study i, Y_i represents the effect size, β the population effect size, b_i the random between-study effect with mean 0 and variance τ^2, and ε_i the sampling error with mean 0 and variance σ_i^2. Within-study variance σ_i^2 is estimated using sample variance of effect size $\widehat{\sigma}_i^2$ from each study. However, information of between-study variance τ^2 is often not available and methods commonly used for assessing between-study heterogeneity include the DerSimonian and Laird's method of moments (MM) [2], the maximum likelihood estimation (MLE) method [3], and the restricted maximum likelihood (REML) method [4]. MM is a distribution free and non-iterative approach, whereas both MLE and REML are parametric methods and need iteration for estimating τ^2 and β.

When synthesis of M multiple outcomes $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iM})^\top$, $M \geqslant 2$, is of interest, it is often found in published literature to have been analyzed by performing univariate meta-analysis for each outcomes Y_{ij}, $j = 1, 2, \ldots, M$. However, a multivariate meta-analysis for all outcomes jointly is shown to produce estimates with higher degree of precision by 'borrowing of strength' through the between-study and within-study correlations, in particular when data are missing at random (MAR) [5]. For example, in the field of orthopedic surgery, when the effect of an antifibrinolytic agent on two outcomes of operative blood loss and blood transfusions is of interest, two univariate meta-analysis was utilized for pooling each effect size and for estimating their related precision [6]. A joint synthesis of the amounts of blood loss and blood transfusions would, however, be more meaningful as the two outcomes are clearly correlated (higher amount of blood loss needing higher number of blood transfusions).

Despite these advantages, use of multivariate meta-analysis approaches are limited by the lack of reporting of the within-study correlation in the original publications, inadequate awareness of the availability of methods among practitioners, and increasing but still limited availability of statistical softwares, to name a few [7]. Application of multivariate meta-analysis methods could benefit a broad spectrum of research fields including ecology [8–10], psychiatry [11–13], nutrition [14–16], health prevention [17–19], ethics [20–22], and gender difference in labor market [23–26] beyond many other fields listed in [7], as research questions involving multiple correlated outcomes are common.

Following the notations for univariate meta-analysis, the formulation for multivariate meta-analysis, on the basis of multivariate random-effects model, is written as

$$\begin{aligned}
&\mathbf{Y}_i = \beta + \mathbf{b}_i + \varepsilon_i;\ i = 1, 2, \ldots, n,\\
&\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iM})^\top;\ \beta = (\beta_1, \beta_2, \ldots, \beta_M)^\top,\\
&\mathbf{b}_i = (b_{i1}, b_{i2}, \ldots, b_{iM})^\top;\ \varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iM})^\top,
\end{aligned} \tag{2}$$

where for the ith study, β describes the vector of population effect size, $\mathbf{b}_i$ the vector of between-study random effect, and ε_i the vector of within-study sampling error. It is often assumed that $\mathbf{b}_i$ and ε_i are independent with zero means and $M \times M$ covariance matrices $\mathrm{Var}(\mathbf{b}_i) = \mathbf{D}$ and $\mathrm{Var}(\varepsilon_i) = \mathbf{\Omega}_i$, respectively.

Similar to the univariate situation, methods for multivariate meta-analysis that are frequently used are likelihood based including MLE and REML [27]. These methods are parametric and rely on specific distributional assumption. For computational convenience, both random effects $\mathbf{b}_i$ and sampling errors ε_i are often assumed to follow normal distribution. When data distribution cannot be verified or are established to be skewed, these assumptions become problematic. This issue has been discussed and criticized heavily in the random-effects model literature [28–31]. Moreover, as pointed out in the discussion of univariate meta-analysis, iterations between $\widehat{\beta}$ and $\widehat{\mathbf{D}}$ are required for the estimation process. This makes the computation burdensome in the situation of high-dimensional outcomes. Therefore, research for development of distribution free and non-iterative estimation procedures has been ongoing.

A procedure extending the DerSimonian and Laird's MM [2] for performing multivariate meta-analysis has recently been proposed by Jackson *et al.* [32]. This multivariate method of moments (MMM) is a non-iterative procedure without needing the assumption of normality. Although MMM was shown to perform similar to REML through simulation studies performed under bivariate normal distribution, there is a lack of information on the performance of these two methods when the data departs from

normality. In this paper, we propose a new nonparametric and non-iterative method on the basis of U-statistic for multivariate meta-analysis and assess the operating characteristics of the three procedures under both normal and non-normal data.

For introducing U-statistic, let $X_1, X_2, \ldots, X_n$ be a random sample from an unknown distribution and λ be a parameter of interest in the form of $\lambda = E[h(X_1, X_2, \ldots, X_m)]$, where the function h is symmetric in its m $(m < n)$ arguments. It is easy to verify that

$$U_n = \binom{n}{m}^{-1} \sum_{(i_1, i_2, \ldots, i_m) \in C_m^n} h\left(X_{i_1}, X_{i_2}, \ldots, X_{i_m}\right) \tag{3}$$

is an unbiased estimator of λ, where C_m^n denotes the set of $\binom{n}{m}$ combinations of m-distinct elements $(i_1, i_2, \ldots, i_m)$ from $\{1, 2, \ldots, n\}$. The statistic U_n in Equation (3) is called a one-sample m-argument U-statistic. Because the seminal work by Hoeffding [33], U-statistics have been used to estimate a wide range of statistics such as order statistics (e.g., Mann–Whitney–Wilcoxon statistic [34]), area under the receiver operating characteristics curve [35], and functions of higher-order moments (e.g., concordance correlation coefficient [36] , Kappa coefficient [37]).

We will introduce the U-statistic-based estimates for performing multivariate meta-analysis using the format of U_n defined previously (Section 2.3). Our motivation for using U-statistic stems from the fact that it would be a nonparametric approach. Additionally, the asymptotic behavior of the related statistics and their estimates would be easy to derive as they would be based on the theorems already available for U-statistic. Because the between-study variance matrix for the random-effects multivariate meta-analysis model involves second-order moments, U-statistic formulation would be specially beneficial as it can be easily applied to estimate the variance matrix components and to develop their joint asymptotic distribution for related inference.

The outline of the paper is as follows. We briefly describe the existing methods for multivariate meta-analysis and derive the new U-statistic-based estimation procedure in Section 2. We report statistical simulation-based examination of the operating characteristics of the proposed approach and its comparison with REML and MMM in Section 3. We describe the results from the application of all three methods to data from published meta-analysis from the fields of periodontal disease and hip fracture in Section 4. Section 5 includes our conclusions and ideas for future research. We report mathematical derivations in Appendix A.

2. Methods for multivariate meta-analysis

We start this section by first reviewing the multivariate random-effects model as described in Equation (2) and focus on the simple case of meta-analysis with bivariate outcome variable $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^\top; i = 1, 2, \ldots, n$ as these methods can be easily extended to meta-analysis with $M > 2$ outcomes. Throughout the paper, $\mathbf{y}_i = (y_{i1}, y_{i2})^\top$ denotes the sample version of $\mathbf{Y}_i$; $i = 1, 2, \ldots, n$. With $M = 2$ in the random-effects model, the marginal variability in $\mathbf{Y}_i$ accounts for both between-study variation ($\mathbf{D}$) and within-study variation ($\mathbf{\Omega}_i$). In particular, the covariance matrix for a bivariate meta-analysis boils down to

$$\begin{aligned} \mathrm{Var}(\mathbf{Y}_i) &= \mathbf{D} + \mathbf{\Omega}_i, \\ \mathbf{D} &= \begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}, \ \mathbf{\Omega}_i = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i12} \\ \sigma_{i12} & \sigma_{i2}^2 \end{bmatrix}, \\ \tau_{12} &= \tau_1 \tau_2 \rho_b, \ \sigma_{i12} = \sigma_{i1} \sigma_{i2} \rho_{iw}, \end{aligned} \tag{4}$$

where τ_1^2, τ_2^2, and ρ_b describe the between-study variation and correlation, whereas σ_{i1}^2, σ_{i2}^2, and ρ_{iw} capture the within-study variation and correlation. Similar to the univariate meta-analysis setting, the within-study variances σ_{i1}^2 and σ_{i2}^2 are estimated using sample variances $\widehat{\sigma}_{i1}^2$ and $\widehat{\sigma}_{i2}^2$. Although the within-study correlation ρ_{iw} is generally unknown, several approaches for addressing this issue have been discussed in [7]. For simplicity, in this paper, the within-study correlation is assumed equal across studies $(\rho_{iw} = \rho_w, i = 1, 2, \ldots, n)$ and a sensitivity analysis for a wide range of ρ_w is conducted. Next, we review two existing methods for multivariate meta-analysis: REML approach [27] and the multivariate extension of the DerSimonian and Laird's MM [32].

2.1. Restricted maximum likelihood method

The REML approach has been widely used and is incorporated in most statistical softwares. By assuming the within-study variance matrix $\mathbf{\Omega}_i$ to be known, the remaining parameters of interest to be estimated include β, τ_1^2, τ_2^2, and ρ_b. Under REML, $\mathbf{b}_i$ and ε_i are usually assumed to follow bivariate normal distributions, $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$, $\varepsilon_i \sim N(\mathbf{0}, \mathbf{\Omega}_i)$. The outcome variable $\mathbf{Y}_i$, as a result, follows a bivariate normal distribution with mean β and variance $\mathbf{D} + \mathbf{\Omega}_i; i = 1, 2, \ldots, n$.

Normal random effects and sampling errors are discussed as that is most commonly assumed. No closed-form derivation for REML estimates exist, and therefore iterative procedures (e.g., Newton–Raphson algorithm, expectation-maximization algorithm) have been developed for estimating β and variance $\mathbf{D}$. Because the REML-based inference for meta-analysis has been extensively discussed in the literature [38], we will not elaborate upon the technical issues in this paper. Briefly, the REML estimate of β can be derived as a function of $\widehat{\mathbf{D}}$. This is equivalent to the restricted iterative generalized least square estimate

$$\widehat{\beta} = \left(\sum_{i=1}^{n} \left(\widehat{\mathbf{D}} + \mathbf{\Omega}_i\right)^{-1}\right)^{-1} \left(\sum_{i=1}^{n} \left(\widehat{\mathbf{D}} + \mathbf{\Omega}_i\right)^{-1} \mathbf{y}_i\right) \tag{5}$$

when outcomes are normally distributed [5].Asymptotically, the aforementioned estimate $\widehat{\beta}$ follows a normal distribution with mean β and variance

$$\mathrm{Var}\left(\widehat{\beta}\right) = \left(\sum_{i=1}^{n} \left(\widehat{\mathbf{D}} + \Omega_i\right)^{-1}\right)^{-1}. \tag{6}$$

Limitations of the REML estimation procedure has been discussed extensively in the random-effects model literature [28–31]. Briefly, the assumption of normality for $\mathbf{b}_i$ and ε_i is often not true in practice and is difficult to verify for moderate number of component studies in a meta-analysis. When departures from normality occur, the parameter estimation and inference from REML may be seriously biased. REML also involves iterative estimation between β and $\mathbf{D}$ resulting in computational inefficiency.

2.2. Multivariate extension of the DerSimonian and Laird's method of moments

Jackson *et al.* [32] extended the Q-statistic proposed by DerSimonian and Laird in [2] for estimating the between-study variance matrix to a multivariate meta-analysis setting by using the MM approach. The MMM method is a non-iterative approach, and the estimate of β and its variance are obtained by replacing $\mathbf{D}$ in Equations (5) and (6) with the MMM estimate of $\mathbf{D}$. This approach is incorporated in the software STATA (StataCorp LP, College Station, TX, USA) [39]. Although, it has been demonstrated in Reference [32], through a simulation study, that the MMM method performs similar to the REML estimation procedure under bivariate normal distribution, further elucidation of its advantage for being a distribution-free method will be useful.

Next, we propose a U-statistic-based nonparametric and non-iterative method for multivariate meta-analysis.

2.3. U-statistic-based distribution free and non-iterative method for meta-analysis

In this section, we propose a U-statistic-based method for bivariate meta-analysis with either complete outcomes or outcomes with some components missing under the assumption of MAR.

2.3.1. U-statistic-based approach for complete data. To estimate $\mathbf{D}$ from Equation (4), we define the weights $w_{1ii'} = (\sigma_{i1}^2 + \sigma_{i'1}^2)^{-1}$, $w_{2ii'} = (\sigma_{i2}^2 + \sigma_{i'2}^2)^{-1}$, $w_{3ii'} = (\sigma_{i12} + \sigma_{i'12})^{-1}$, and $\mathbf{w}_{ii'} = \left(w_{1ii'}, w_{2ii'}, w_{3ii'}\right)^{\top}$ for each $\left(i, i'\right) \in C_2^n = \left\{\left(i, i'\right); 1 \leq i < i' \leq n\right\}$.

Let $\eta = (\eta_1, \eta_2, \eta_3)^\top$, $\eta_1 = \tau_1^2$, $\eta_2 = \tau_2^2$, $\eta_3 = \tau_{12}$, and

$$\mathbf{h}_{ii'} = \begin{pmatrix} h_{1ii'} \\ h_{2ii'} \\ h_{3ii'} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \frac{w_{1ii'}\left(\left(y_{i1}-y_{i'1}\right)^2-\sigma_{i1}^2-\sigma_{i'1}^2\right)}{\sum_{(i,i')\in C_2^n} w_{1ii'}} \\ \frac{w_{2ii'}\left(\left(y_{i2}-y_{i'2}\right)^2-\sigma_{i2}^2-\sigma_{i'2}^2\right)}{\sum_{(i,i')\in C_2^n} w_{2ii'}} \\ \frac{w_{3ii'}\left(\left(y_{i1}-y_{i'1}\right)\left(y_{i2}-y_{i'2}\right)-\sigma_{i12}-\sigma_{i'12}\right)}{\sum_{(i,i')\in C_2^n} w_{3ii'}} \end{pmatrix}, \left(i, i'\right) \in C_2^n. \tag{7}$$

Then, it can be easily verified that

$$\widehat{\eta} = (\widehat{\eta}_1, \widehat{\eta}_2, \widehat{\eta}_3)^\top = \sum_{(i,i')\in C_2^n} \mathbf{h}_{ii'} \tag{8}$$

is a one-sample, vector-valued U-statistic and $E(\widehat{\eta}) = \eta$ [33]. By applying the theory of multivariate U-statistic [33], $\widehat{\eta}$ is a consistent and asymptotically normal estimate of η with a variance matrix Σ_η. A consistent estimate $\widehat{\Sigma}_\eta$ of Σ_η is provided in Appendix A.

Note that the U-statistic estimate of η in Equation (8) does not guarantee that the estimated between-study variance matrix $\widehat{\mathbf{D}}$ is positive semidefinite. A popular approach of transforming $\widehat{\mathbf{D}}$ to a positive semidefinite matrix is to replace any of its negative eigenvalues with 0 so that $\widehat{\mathbf{D}}$ is truncated to $\widehat{\mathbf{D}}^+$

$$\widehat{\mathbf{D}}^+ = \sum_{i=1}^{2} \max(0, \delta_i)\xi_i\xi_i^\top, \tag{9}$$

where δ_i is the ith eigenvalue of the 2×2 matrix $\widehat{\mathbf{D}}$ and ξ_i is its associated normalized eigenvector. It has been proved that $\widehat{\mathbf{D}}^+$ is the closest matrix to $\widehat{\mathbf{D}}$ among all positive semidefinite matrices [40]. Hence, the U-statistic-based estimate $\widehat{\beta}$ and its variance Var$\left(\widehat{\beta}\right)$ are obtained by substituting $\widehat{\mathbf{D}}^+$ for $\widehat{\mathbf{D}}$ in Equations (5) and (6) .

Of note, an unbiased estimate of the between-study variance τ^2 in the univariate meta-analysis context in Equation (1) can be readily derived from Equation (8):

$$\hat{\eta} = \frac{\sum_{(i,i')\in C_2^n} w_{ii'}\left(\left(y_i - y_{i'}\right)^2 - \sigma_i^2 - \sigma_{i'}^2\right)}{\sum_{(i,i')\in C_2^n} w_{ii'}}, \tag{10}$$

where $w_{ii'} = (\sigma_i^2 + \sigma_{i'}^2)^{-1}$. The estimate is truncated at 0 to make sure that it is non-negative by using $\widehat{\tau}^2 = \max(0, \widehat{\eta})$. Also, $\widehat{\beta}$ and Var$\left(\widehat{\beta}\right)$ for a univariate meta-analysis are obtained by replacing $\widehat{\mathbf{D}}$ and Ω_i with $\widehat{\tau}^2$ and $\widehat{\sigma}_i^2$ in Equations (5) and (6), respectively.

2.3.2. U-statistic-based approach for missing data. Missing one outcome for a moderate percentage of studies is quite common in bivariate meta-analysis. Nonsignificant or negative findings are often less likely to be published causing one of the outcomes to be missing. This phenomenon is called publication bias. Although unreported outcomes in general may be MAR [5], in meta-analytic setting where missing outcomes are primarily because of publication bias, MAR [41] might not be the underlying missing data mechanism. However, it is almost impossible to determine the underlying missing data mechanism in a meta-analytic setting because of the limited information available through the summarized effect sizes from each study. In the context of bivariate meta-analysis and under the assumption of MAR, Riley *et al.* [5] demonstrated that compared with two independent univariate meta-analysis, one is able to provide more precise estimations by utilizing the 'borrowing of strength' through the between-study and within-study correlations. Therefore, we decided to extend the U-statistic approach for this setting as well.

To generalize the proposed approach in Section 2.3.1 under the assumption of MAR, we define a vector of binary missing data indicator variables,

$$r_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ and } \sigma_{ij}^2 \text{ are observed} \\ 0 & \text{if } Y_{ij} \text{ and } \sigma_{ij}^2 \text{ are unobserved} \end{cases}, \quad \mathbf{r}_i = (r_{i1}, r_{i2})^\top, \; j = 1, 2; \; i = 1, 2, ..n, \qquad (11)$$

and new weights, $\widetilde{w}_{1ii'} = r_{i1}r_{i'1}\left(\sigma_{i1}^2 + \sigma_{i'1}^2\right)^{-1}$, $\widetilde{w}_{2ii'} = r_{i2}r_{i'2}\left(\sigma_{i2}^2 + \sigma_{i'2}^2\right)^{-1}$, $\widetilde{w}_{3ii'} = r_{i1}r_{i2}r_{i'1}$ $r_{i'2}\left(\sigma_{i12} + \sigma_{i'12}\right)^{-1}$, and $\widetilde{\mathbf{w}}_{ii'} = \left(\widetilde{w}_{1ii'}, \widetilde{w}_{2ii'}, \widetilde{w}_{3ii'}\right)^\top$. Then, an estimate of the between-study variance matrix $\mathbf{D}$ is readily obtained by replacing $\mathbf{w}_{ii'}$ in Section 2.3.1 with $\widetilde{\mathbf{w}}_{ii'}$ and by assigning an arbitrary value (e.g., 0) to any of the missing observations in y, σ^2, or σ_{12}. To compute $\widehat{\beta}$ and Var $\left(\widehat{\beta}\right)$, the method for handling missing data in Reference [32] is followed, which recommends assigning very large (e.g., 10^{12}) within-study variances and setting y and within-study correlation to be 0 in Equations (5) and (6) whenever an observation is missing. Next, we compare the U-statistic-based approach with REML and MMM methods using statistical simulation.

3. Simulation study

3.1. Simulation method

3.1.1. Simulation parameter specification. To assess and compare the finite sample properties of the U, MMM, and REML procedures, we conducted a simulation study for a sample size of $n = 10$ per meta-analysis. We carried out all data simulations and analyses using the software R and STATA. We consider a wide range of values of τ_1^2, τ_2^2, ρ_b, and ρ_w under complete and missing data scenarios. We generated data in three steps:

Step 1. Within-study variances σ_{i1}^2 and σ_{i2}^2 were independently generated from an exponential distribution exp(λ), $\lambda = 0.5$. All within-study variances were truncated to the range of $[0.5, 10]$ resulting in σ_i^2 with a mean of 2.5, median of 1.8, and first and third quartiles of 1.1 and 3.2, respectively.

Step 2. We used the between-study heterogeneity statistic, I^2 (I_1^2 for the first outcome and I_2^2 for the second outcome) defined by Higgins and Thompson [42] as described in the following text to represent the proportion of total variation in study estimates that is due to between-study variance:

$$I_j^2 = \frac{\tau_j^2}{s_j^2 + \tau_j^2}, \; s_j^2 = \lim_{n\to\infty} \frac{(n-1)\sum_{i=1}^n \sigma_{ij}^{-2}}{\left(\sum_{i=1}^n \sigma_{ij}^{-2}\right)^2 - \sum_{i=1}^n \sigma_{ij}^{-4}}; \; j = 1, 2 \qquad (12)$$

Three values of I^2 were considered to reflect low $\left(I^2 = 0.25\right)$, moderate $\left(I^2 = 0.5\right)$, and high $\left(I^2 = 0.85\right)$ between-study heterogeneity among estimates for each outcome. We then computed between-study variance τ_j^2 using Equation (12) with prespecified I_j^2. This resulted in $\tau^2 = 0.494$ when $I^2 = 0.25$, $\tau^2 = 1.482$ when $I^2 = 0.5$, and $\tau^2 = 8.398$ when $I^2 = 0.85$.

Step 3. We set both β_1 and β_2 to 10 without loss of generality. We chose two values of ρ_b (ρ_w) to represent mild (0.3) and strong (0.8) between-study (within-study) correlations.

To further elucidate the dependence of properties of parameter estimation on the relative size of the between-study heterogeneity $\left(I^2\right)$, we performed an additional simulation run focusing on the settings of $I_1^2 = I_2^2 = I^2$, $\rho_b = \rho_w = \rho$, and $n = 10$. We selected seven values of I^2 in the range of $[0.3, 0.9]$ to reflect a wide range of heterogeneities over which the trend will become apparent. ρ was set to be 0.9.

3.1.2. Data generation and estimation process. In the first simulation study, we generated outcome $\mathbf{Y}_i$ from the bivariate normal distribution with marginal mean β and variance $(\mathbf{D} + \Omega_i)$ on the basis of the random-effects model in Equation (2). In the second simulation study, we generated data under the assumption that both random effects $\mathbf{b}_i$ and sampling errors ε_i follow exponential distribution. This enabled us to investigate the robustness of REML (which works under an assumed normal distribution) to the scenario of highly skewed data distribution and compare its performance with the other two procedures. We generated $\mathbf{b}_i$ from a Gaussian copula model [43], $\mathbf{\Phi}\rho_c\left(\Phi^{-1}\left(F_1\left(b_{i1}\right)\right), \Phi^{-1}\left(F_2\left(b_{i2}\right)\right)\right)$,

where F_j represents the marginal univariate cumulative distribution function of b_{ij}, $j = 1, 2$, Φ^{-1} the quantile function of standard normal and $\mathbf{\Phi}\rho_c$ the distribution function of standard bivariate normal with correlation ρ_c. This ensures that b_{i1} and b_{i2} are correlated and that each of them follow an exponential distribution: $b_{i1} \sim \exp(\lambda_1)$, $b_{i2} \sim \exp(\lambda_2)$.

Note that ρ_c is the Pearson correlation coefficient between $\Phi^{-1}(F_1(b_{i1}))$ and $\Phi^{-1}(F_2(b_{i2}))$ in the Gaussian copula model and not the Pearson between-study correlation ρ_b in general. The two correlations would be equal if a rank correlation formulation such as Spearman's ρ or Kendall's τ was utilized because the rank correlation coefficients are invariant to monotone transformations [44]. It is also known from the copula literature that the Spearman rank correlation $\rho_s = \frac{6}{\pi}\sin^{-1}(0.5\rho_c) \approx \rho_c$ [44]. We calculated the between-study correlation using both Spearman (ρ_s) and Pearson (ρ_b) correlations and found that they agreed to three decimal places in our simulation. Because both ρ_c and ρ_b are approximately equal to ρ_s and the impact of the small difference between these correlations on parameter estimation is negligible in this study, we considered $\rho_c = \rho_b\ (= \rho_s)$ throughout the simulation study.

The procedure in step 2 was followed to determine between-study variances τ_1^2 and τ_2^2. Further, we obtained λ_1 and λ_2 through $\lambda_1 = (\tau_1^2)^{-0.5}$ and $\lambda_2 = (\tau_2^2)^{-0.5}$ given the fact that $\mathrm{Var}(b_{ij}) = (\lambda_j)^{-2}$ under exponential distribution. Similarly, we utilized a Gaussian copula model to generate ε_i with a correlation ρ_w and marginal exponential distributions $\exp(c_{i1})$ and $\exp(c_{i2})$ for ε_{i1} and ε_{i2}, respectively, where $c_{ij} = \left(\sigma_{ij}^2\right)^{-0.5}$.

We centered both $\mathbf{b}_i$ and ε_i at 0. For generating bivariate meta-analysis with missing data, we used the same data created for the complete case and randomly assigned 20% of the data for the first outcome Y_{i1} and 20% for the second outcome Y_{i2} to be missing.

We then applied the U-statistic, MMM, and the REML methods to the simulated datasets in various scenarios and estimated the effect sizes and the between-study variance matrix. We implemented the REML procedure using STATA function *mvmeta* and programmed the U-statistic and the MMM in R (R Foundation for Statistical Computing, Vienna, Austria).

3.1.3. Reported summary statistics. We reported the relative bias of the averaged $\widehat{\beta}$, $\widehat{\tau}_1^2$, $\widehat{\tau}_2^2$, and $\widehat{\tau}_{12}$ over 1000 Monte Carlo replications utilizing formulas like $\frac{\widehat{\beta}-\beta}{\beta}\times 100$. We calculated the 95% confidence interval (CI) of $\widehat{\beta}_i$ using $\hat{\beta}_i \pm t_{0.025,(n-1)}\sqrt{\mathrm{Var}(\hat{\beta}_i)}; i = 1, 2$, where $t_{0.025,(n-1)}$ denotes the 0.025 percentile of a t-distribution with $(n-1)$ degrees of freedom. We also presented the proportions of the CIs that cover the true β_i's. We estimated the mean square error $\mathrm{MSE}\left(\hat{\beta}_i\right)$ using $\sum_{j=1}^{1000}\left(\hat{\beta}_{ij}-\beta_i\right)^2/1000$.

In multivariate meta-analysis procedure, the estimate of between-study variance matrix is often truncated at the boundary of its parameter space to comply with the positive semidefinite restriction expected of variance matrices. This leads to inflated between-study variance estimates [32, 45]. Therefore, we recorded the incidence of $\widehat{\tau}_1^2 = 0$, $\widehat{\tau}_2^2 = 0$, or $|\widehat{\rho}_b| = 1$ to assess the impact of this truncation on the parameter estimates $\widehat{\beta}$, $\widehat{\tau}_1^2$, $\widehat{\tau}_2^2$, and $\widehat{\tau}_{12}$. We obtained these incidences from Equation (9) for the U-statistic and the MMM approaches. We used the rule described in [32] of substituting 0 for any between-study variances below 0.00005 and 1 for any absolute between-study correlations above 0.9995 to compute these values for the REML procedure.

3.2. Simulation result

We present our simulation results in two tables with twelve scenarios and in one figure. Scenarios 1–6 consider the situation when proportions of between-study heterogeneities in Y_{i1} and Y_{i2} are similar $\left(I_1^2 = I_2^2\right)$, and scenarios 7–12 describe the situations when $I_1^2 < I_2^2$. We use assumption of $\rho_b = \rho_w = \rho$ for all twelve scenarios.

3.2.1. Bivariate meta-analysis; underlying distribution skewed; complete data. Table I shows the simulation results with data generated under exponential distribution. Across all scenarios, the estimation of the between-study variance matrix was more sensitive to the change in the magnitude of I^2 when compared with other parameters of interest. The relative bias of $\widehat{\tau}^2$ for all three methods was large ($> 65\%$) with a high chance of observing $|\widehat{\rho}_b| = 1$ (e.g., for scenario 1, 66% in REML, 43% in MMM , 44% in U) when the relative size of between-study variation was low. This resulted in a higher coverage probability for the β estimates than the nominal 95%. However, the impact of the inflated bias in $\widehat{\tau}^2$ on the pooled

Table I. Simulation results over 1000 Monte Carlo replications under exponential distribution when sample size $n = 10$.

Scenario	$I_1^2/I_2^2/\rho$	Method	R-bias $(\widehat{\beta}_1)$	R-bias $(\widehat{\beta}_2)$	MSE $(\widehat{\beta}_1)$	MSE $(\widehat{\beta}_2)$	Coverage $(\widehat{\beta}_1)$	Coverage $(\widehat{\beta}_2)$	R-bias $(\widehat{\tau}_1^2)$	R-bias $(\widehat{\tau}_2^2)$	R-bias $(\widehat{\tau}_{12})$	$\%(\widehat{\tau}_1^2 = 0)$	$\%(\widehat{\tau}_2^2 = 0)$	$\%(\lvert\widehat{\rho}_b\rvert = 1)$
1	0.25/0.25/0.3	REML	−0.03	0.021	0.236	0.217	0.976	0.981	79.315	92.328	50.18	17.9	17.7	66.1
		MMM	0.008	0.062	0.233	0.214	0.978	0.988	67.953	77.993	2.238	18.6	18.3	42.8
		U	0.119	0.175	0.238	0.217	0.98	0.988	85.225	92.466	11.45	21.4	21.3	44.2
2	0.25/0.25/0.8	REML	0.002	0.025	0.238	0.223	0.975	0.982	89.336	97.329	81.36	16.7	17.5	65.2
		MMM	0.041	0.061	0.234	0.219	0.974	0.983	66.023	71.22	39.458	17	16.1	46.5
		U	0.183	0.206	0.239	0.224	0.974	0.983	83.698	87.247	52.85	19.6	19.6	43.3
3	0.5/0.5/0.3	REML	−0.174	−0.091	0.348	0.311	0.948	0.964	11.227	17.654	−7.948	5.5	5.5	58.2
		MMM	−0.156	−0.087	0.343	0.307	0.948	0.967	7.806	13.982	−21.122	5.5	5.6	40.9
		U	−0.023	0.034	0.347	0.309	0.949	0.967	11.128	16.461	−20.359	6.9	6.0	39.5
4	0.5/0.5/0.8	REML	−0.154	−0.124	0.348	0.323	0.939	0.954	14.577	18.038	9.258	5.8	5.7	55.8
		MMM	−0.152	−0.13	0.342	0.316	0.938	0.957	6.853	10.481	−3.325	5.3	5.2	40.5
		U	−0.013	0.033	0.347	0.318	0.938	0.959	10.55	13.392	−1.274	6.0	5.6	40.4
5	0.85/0.85/0.3	REML	−0.506	−0.375	1.047	0.934	0.904	0.925	−6.345	−2.654	−20.49	0.1	0.2	13.1
		MMM	−0.597	−0.471	1.043	0.927	0.905	0.928	−5.37	−1.777	−20.344	0	0	7.5
		U	−0.482	−0.365	1.039	0.925	0.903	0.929	−5.841	−1.916	−20.273	0.2	0	8
6	0.85/0.85/0.8	REML	−0.528	−0.458	1.049	0.968	0.898	0.915	−6.336	−4.010	−9.794	0.1	0.1	15.1
		MMM	−0.666	−0.605	1.043	0.959	0.9	0.912	−6.108	−3.478	−9.704	0.1	0	10.2
		U	−0.51	−0.45	1.035	0.954	0.9	0.912	−6.131	−3.415	−9.623	0.1	0	7.8
7	0.25/0.5/0.3	REML	−0.024	−0.092	0.236	0.312	0.98	0.961	81.031	17.723	14.822	10.3	9.9	65.7
		MMM	0.002	−0.079	0.233	0.306	0.977	0.965	69.855	13.639	−15.846	11.8	9.2	43.2
		U	0.112	0.041	0.237	0.308	0.978	0.968	87.266	16.173	−12.628	11.4	10.3	46.6

Table I. *continued*

8	0.25/0.5/0.8	REML	−0.016	−0.114	0.237	0.321	0.975	0.957	93.037	20.577	38.297	8.6	8.2	67
		MMM	0.013	−0.098	0.234	0.316	0.973	0.957	68.115	10.574	8.808	8.6	7	45.7
		U	0.151	0.056	0.239	0.319	0.973	0.959	85.888	13.65	14.191	9.5	8.3	44.1
9	0.25/0.85/0.3	REML	−0.048	−0.388	0.237	0.931	0.975	0.933	81.647	−1.962	−5.137	2.9	0.5	59.5
		MMM	−0.025	−0.452	0.234	0.923	0.977	0.925	70.583	−1.686	−22.32	3.6	0.8	40.2
		U	0.081	−0.338	0.238	0.919	0.977	0.931	87.76	−1.8	−23.487	3.1	0.8	37.9
10	0.25/0.85/0.8	REML	−0.03	−0.382	0.233	0.943	0.974	0.924	100.46	−4.1	12.668	2.4	0.2	63.1
		MMM	−0.057	−0.456	0.234	0.941	0.975	0.914	74.763	−3.336	−8.613	1.8	0.2	44.3
		U	0.072	−0.307	0.24	0.94	0.977	0.915	92.889	−3.201	−8.259	1.8	0.2	40.5
11	0.5/0.85/0.3	REML	−0.198	−0.383	0.349	0.93	0.951	0.931	11.065	−2.285	−16.732	1.5	0.4	40.7
		MMM	−0.193	−0.463	0.344	0.924	0.944	0.927	7.664	−1.694	−21.66	1.4	0.4	26
		U	−0.067	−0.35	0.346	0.921	0.944	0.93	10.852	−1.818	−22.303	2	0.4	28.5
12	0.5/0.85/0.8	REML	−0.153	−0.419	0.342	0.953	0.948	0.918	16.677	−2.157	−2.505	1.5	0.4	49.1
		MMM	−0.208	−0.517	0.339	0.949	0.946	0.912	8.818	−3.395	−9.75	1	0.3	33.2
		U	−0.049	−0.355	0.342	0.944	0.943	0.917	12.697	−3.288	−9.713	1.1	0.2	34.3

I_1^2 denotes the I^2 statistic for the first outcome. ρ denotes both within-study and between-study correlation ($\rho = \rho_w = \rho_b$). REML, MMM, and U denote the restricted maximum likelihood method, the multivariate method of moments, and the U-statistic method, respectively. $R\text{-bias}\left(\widehat{\beta}_1\right)$ denotes the relative bias of the first effect estimate $\widehat{\beta}_1$ $\left(\left(\widehat{\beta}_1 - \beta_1\right)/\beta_1 * 100\right)$. $\text{MSE}\left(\widehat{\beta}_1\right)$ denotes the mean square error of $\widehat{\beta}_1$. $\text{Coverage}\left(\widehat{\beta}_1\right)$ denotes coverage of 95%CI for $\widehat{\beta}_1$. $R\text{-bias}\left(\widehat{\tau}_1^2\right)$ denotes the relative bias of the between-study variance estimate $\widehat{\tau}_1^2$. $\%\left(\widehat{\tau}_1^2 = 0\right)$ and $\%\left(|\widehat{\rho}_b| = 1\right)$ denote percentage of $\left(\widehat{\tau}_1^2 = 0\right)$ and $\left(|\widehat{\rho}_b| = 1\right)$, respectively. The true values are set to $\beta_1 = \beta_2 = 10$, $\tau_1^2\left(\tau_2^2\right) = 0.494$ when $I^2 = 0.25$, 1.482 when $I^2 = 0.5$, and 8.398 when $I^2 = 0.85$.

effect size estimates was low as the between-study variation was only a very small portion of the total variation. As the relative size of between-study variation increased, the bias in $\widehat{\tau}^2$ descended quickly for all methods and at the same time the between-study variances became increasingly influential in the estimation of β. For example, compared with the solutions from scenario 2 $\left(I_1^2 = I_2^2 = 0.25\right)$, the averaged absolute relative bias and MSE of $\widehat{\beta}_{1,U}$ in scenario 6 $\left(I_1^2 = I_2^2 = 0.85\right)$ increased from 0.2% to 0.5% and 0.2 to 1 respectively, though the absolute relative bias of $\widehat{\tau}^2_{1,U}$ dropped from 84% to 6%. In most scenarios, $\widehat{\tau}^2_{\text{REML}}$ had the largest bias, whereas $\widehat{\tau}^2_{\text{MMM}}$ had the smallest bias. The upward bias in between-study variance estimates is a common finding in multivariate meta-analysis [32, 45], and it occurs because of the truncation of between-study variance matrix. As described in Reference [45], it would be more appropriate to consider this phenomenon as a 'natural consequence of the sensible positive semidefinite restriction' of the variance matrix rather than 'poor estimation'. We will not elaborate on this issue further, as more discussion about this estimation problem can be found in [45]. In addition, it was found in the estimation of β that the U-statistic estimate tended to have smaller relative bias and MSE when compared with REML and MMM as I^2 increased (e.g., scenario 6).

3.2.2. Bivariate meta-analysis; underlying distribution symmetric; complete data. Table II shows the simulation results with data generated from normal distribution. Compared with Table I, under bivariate normal distribution, the inflation in the estimation of between-study variances reduced significantly for all three methods, especially when I^2 was small; the relative bias and MSE of β from the three methods were less distinguishable.

3.2.3. Trend in relative bias and mean squared error over between-study heterogeneity. Figures 1a–1c depicted the trend of the relative bias of $\widehat{\beta}_1$, the ratio of MSE of $\widehat{\beta}_1$ between any two methods, and the relative bias of $\widehat{\tau}^2_1$ over the seven values of I^2 when the data are generated from under exponential distribution. The relative bias of $\widehat{\beta}_{1,U}$ tended to be smaller than REML and MMM when $I^2 > 0.5$ with REML falling in between U and MMM over I^2. Both U and MMM methods tended to have smaller MSEs of $\widehat{\beta}_1$ than REML over I^2. In particular, U had slightly smaller MSEs of $\widehat{\beta}_1$ than MMM when $I^2 > 0.7$, whereas MMM was shown to be better when $I^2 < 0.5$. The MSE of REML tended to become smaller and closer to that of MMM and U when I^2 was extremely large. The inflated bias in $\widehat{\tau}^2_1$ was more prominent for REML and dropped quickly for all three methods as I^2 increased.

Note that we excluded the cases where the relative size of between-study variation is extremely small $\left(I^2 \leqslant 0.2\right)$. In such situations, the total variation is dominated by the within-study variation, and fixed-effects bivariate meta-analysis as opposed to random effects is recommended in practice [41].

3.2.4. Additional simulation studies. In the simulation study with missing data, comparing the results with those in the complete case under both normal and exponential distributions, the relative bias of $\widehat{\tau}^2$ increased with reduced information about between-study heterogeneity because of missing data, but it did not have a significant impact on the estimation of β.

When the number of component studies were increased to $n = 50$, the three procedures produced more similar results with reduced relative bias and smaller MSE compared with the small sample case ($n = 10$) under both exponential and normal distributions.

In addition to the bivariate meta-analysis, we also conducted a simulation study to compare the performance of the U-statistic, the DerSimonian and Laird's MM [2], and the REML estimation procedures in univariate meta-analysis. We kept the true values of parameters in steps 1–3 when simulating data in a univariate data setting. The three procedures were very similar in every respect under normal distribution and followed similar trends to those shown in Figures 1a–1c for bivariate meta-analysis under exponential distribution.

Additional information for the simulation results not described here will be made available by the first author upon request.

3.2.5. Computational efficiency. In the simulation study, the U-statistic method being a non-iterative approach gained considerable computational efficiency compared with the REML. This was very similar to the advantage of MMM. For each scenario with 1000 replications, it took REML 20 min to produce

Table II. Simulation results over 1000 Monte Carlo replications under normal distribution when sample size $n = 10$.

Scenario	$I_1^2/I_2^2/\rho$	Method	R-bias $(\widehat{\beta}_1)$	R-bias $(\widehat{\beta}_2)$	MSE $(\widehat{\beta}_1)$	MSE $(\widehat{\beta}_2)$	Coverage $(\widehat{\beta}_1)$	Coverage $(\widehat{\beta}_2)$	R-bias $(\widehat{\tau}_1^2)$	R-bias $(\widehat{\tau}_2^2)$	R-bias $(\widehat{\tau}_{12})$	$\%(\widehat{\tau}_1^2 = 0)$	$\%(\widehat{\tau}_2^2 = 0)$	$\%(\lvert\widehat{\rho}_b\rvert = 1)$
1	0.25/0.25/0.3	REML	−0.131	−0.054	0.236	0.218	0.962	0.974	49.226	60.582	47.245	10.6	9.6	67.1
		MMM	−0.112	−0.042	0.236	0.218	0.963	0.977	47.692	59.205	24.444	8.2	8.4	45
		U	−0.107	−0.037	0.237	0.218	0.958	0.976	62.245	71.701	28.349	10.5	10.2	44.1
2	0.25/0.25/0.8	REML	−0.113	−0.077	0.232	0.221	0.968	0.975	47.497	55.386	48.819	10	10	68
		MMM	−0.088	−0.054	0.232	0.22	0.965	0.967	41.01	48.497	30.635	7.3	7.3	47.1
		U	−0.091	−0.053	0.232	0.221	0.964	0.97	52.544	59.06	38.101	9.8	9.5	46
3	0.5/0.5/0.3	REML	−0.165	−0.048	0.353	0.318	0.949	0.959	5.978	11.937	4.434	2.6	2.8	49.7
		MMM	−0.145	−0.042	0.352	0.317	0.952	0.963	4.933	10.585	−2.14	2.3	2.1	29.4
		U	−0.132	−0.03	0.353	0.318	0.953	0.965	8.233	13.906	−1.645	3	3.2	29.6
4	0.5/0.5/0.8	REML	−0.141	−0.078	0.348	0.327	0.946	0.957	5.962	9.982	6.967	2.5	3.3	50.3
		MMM	−0.12	−0.068	0.349	0.327	0.941	0.948	3.313	7.078	1.195	2	1.8	31.7
		U	−0.108	−0.052	0.349	0.327	0.942	0.948	5.843	9.627	2.584	3	3.2	31.1
5	0.85/0.85/0.3	REML	−0.276	−0.121	1.082	0.949	0.944	0.955	−3.165	−1.025	−3.992	0	0	4.2
		MMM	−0.278	−0.128	1.084	0.95	0.942	0.952	−3.224	−1.604	−4.329	0.1	0	1.4
		U	−0.271	−0.123	1.083	0.949	0.947	0.955	−2.888	−0.814	−3.582	0.1	0	2.8
6	0.85/0.85/0.8	REML	−0.27	−0.178	1.071	0.99	0.939	0.954	−3.137	−1.693	−2.656	0	0	5.6
		MMM	−0.26	−0.175	1.073	0.992	0.935	0.95	−3.329	−2.203	−3.113	0	0	3.3
		U	−0.259	−0.175	1.07	0.99	0.941	0.952	−2.894	−1.489	−2.525	0	0	2.8–
7	0.25/0.5/0.3	REML	−0.137	−0.066	0.235	0.319	0.963	0.962	50.664	12.156	22.939	6	4.9	63.3
		MMM	−0.112	−0.055	0.236	0.318	0.962	0.963	49.047	10.698	6.918	5.3	4.3	42.1
		U	−0.107	−0.046	0.237	0.318	0.961	0.967	63.622	13.997	7.521	7.2	5.9	40.1

Table II. *continued*

Scenario	$I_1^2/I_2^2/\rho$	Method	R-bias $(\widehat{\beta}_1)$	R-bias $(\widehat{\beta}_2)$	MSE $(\widehat{\beta}_1)$	MSE $(\widehat{\beta}_2)$	Coverage $(\widehat{\beta}_1)$	Coverage $(\widehat{\beta}_2)$	R-bias $(\widehat{\tau}_1^2)$	R-bias $(\widehat{\tau}_2^2)$	R-bias $(\widehat{\tau}_{12})$	$\%(\widehat{\tau}_1^2=0)$	$\%(\widehat{\tau}_2^2=0)$	$\%(\lvert\widehat{\rho}_b\rvert=1)$
8	0.25/0.5/0.8	REML	−0.115	−0.096	0.232	0.327	0.967	0.956	52.598	11.664	25.769	4.8	4.2	65.2
		MMM	−0.092	−0.083	0.233	0.326	0.967	0.951	43.093	7.339	10.303	3.9	2.8	45.8
		U	−0.088	−0.077	0.233	0.326	0.963	0.952	54.72	9.947	13.433	4.7	4.	42.4
9	0.25/0.85/0.3	REML	−0.138	−0.163	0.235	0.95	0.963	0.953	51.534	−0.46	9.709	1.9	0.2	51.4
		MMM	−0.128	−0.156	0.235	0.949	0.958	0.953	47.598	−1.497	−1.026	1.4	0	32
		U	−0.116	−0.146	0.237	0.948	0.955	0.954	61.864	−0.668	−1.974	1.3	0.1	36.9
10	0.25/0.85/0.8	REML	−0.147	−0.295	0.23	0.974	0.966	0.955	62.578	0.991	13.647	1.2	0	59.7
		MMM	−0.13	−0.262	0.23	0.972	0.965	0.949	48.502	−2.068	−1.331	1.8	0	38
		U	−0.114	−0.248	0.231	0.972	0.963	0.952	60.771	−1.312	−0.855	1.2	0	36.9
11	0.5/0.85/0.3	REML	−0.18	−0.146	0.354	0.948	0.946	0.956	5.808	−0.729	−0.213	0.7	0.1	29.1
		MMM	−0.172	−0.147	0.353	0.949	0.95	0.952	4.163	−1.533	−3.116	0.9	0	16.9
		U	−0.159	−0.136	0.356	0.948	0.95	0.955	7.265	−0.719	−3.189	0.8	0	18.8
12	0.5/0.85/0.8	REML	−0.188	−0.273	0.343	0.979	0.949	0.954	9.933	−0.291	2.685	0.5	0	39.6
		MMM	−0.175	−0.249	0.342	0.978	0.947	0.95	4.747	−2.107	−2.64	0.9	0	22.8
		U	−0.159	−0.235	0.343	0.977	0.946	0.952	7.544	−1.366	−2.232	1.2	0	26.8

I_1^2 denotes the I^2 statistic for the first outcome. ρ denotes both within-study and between-study correlation ($\rho = \rho_w = \rho_b$). REML, MMM, and U denote the restricted maximum likelihood method, the multivariate method of moments, and the U-statistic method, respectively. R-bias$\left(\widehat{\beta}_1\right)$ denotes the relative bias of the first effect estimate $\widehat{\beta}_1$ $\left(\left(\widehat{\beta}_1 - \beta_1\right)/\beta_1 * 100\right)$. MSE$\left(\widehat{\beta}_1\right)$ denotes the mean square error of $\widehat{\beta}_1$. Coverage$\left(\widehat{\beta}_1\right)$ denotes coverage of 95%CI for $\widehat{\beta}_1$. R-bias$\left(\widehat{\tau}_1^2\right)$ denotes the relative bias of the between-study variance estimate $\widehat{\tau}_1^2$. $\%\left(\widehat{\tau}_1^2 = 0\right)$ and $\%\left(\lvert\widehat{\rho}_b\rvert = 1\right)$ denote percentage of $\left(\widehat{\tau}_1^2 = 0\right)$ and $\left(\lvert\widehat{\rho}_b\rvert = 1\right)$, respectively. The true values are set to $\beta_1 = \beta_2 = 10$, $\tau_1^2\,(\tau_2^2) = 0.494$ when $I^2 = 0.25$, 1.482 when $I^2 = 0.5$, and 8.398 when $I^2 = 0.85$.

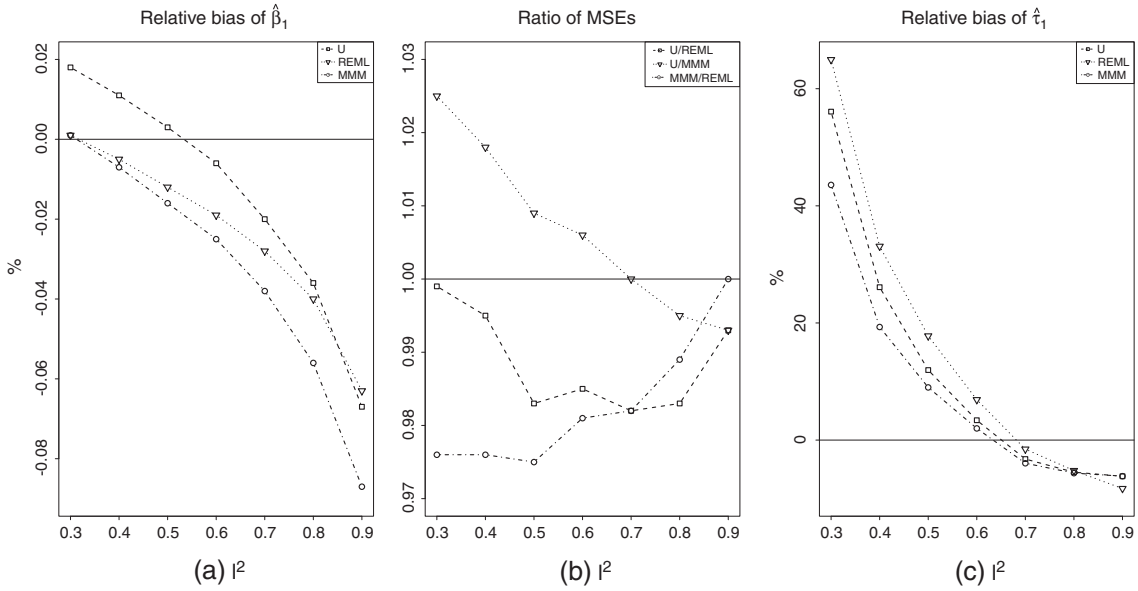

Figure 1. Bivariate meta-analysis ($n = 10$): trends of relative bias and ratio of mean square error of $\hat{\beta}_1$ and relative bias of $\hat{\tau}_1^2$ over heterogeneity index I^2 under exponential distribution.

results. In contrast, both U-statistic and MMM finished computations in few seconds. Although the computational speed is not crucial for performing computation for an individual multivariate meta-analysis with small number of outcomes, it is still an advantage for high-dimensional outcome setting.

4. Applications

We apply the proposed method to two published meta-analysis, one with complete data and the other one with missing data, and compare the results with those from the REML and the method of moments procedures.

4.1. Example 1. A meta-analysis of surgical versus nonsurgical procedure for treating periodontal disease

A meta-analysis of five randomized controlled trials was conducted for comparing a surgical versus a nonsurgical procedure for treating periodontal disease with two endpoints of probing depth (Y_1) and attachment level (Y_2) [46]. Summarized data in terms of mean differences of each endpoint between the two treatments (surgical minus nonsurgical) and the within-study variance matrix with known correlation are available in published literature [7].

The three bivariate meta-analysis (BRMA) methods were applied. All indicate that the surgical procedure improved probing depth, whereas the nonsurgical procedure improved attachment level and reveal a strong between-study correlation across outcomes (Table III). These methods produced similar estimates for $\widehat{\beta}_1$ (0.35), $\widehat{\beta}_2$ (−0.34), and $\widehat{\tau}_1^2$ (0.01), but quite different $\widehat{\tau}_2^2$.

Compared with the solutions in the BRMA, the pooled estimates of the three methods showed greater similarity and indicated larger effect sizes of surgical versus nonsurgical procedures in both outcomes among the univariate meta-analysis solutions. This illustrated the impact of 'borrowing of strength' across outcomes on the estimation of β in BRMA [5].

4.2. Example 2. A meta-analysis of Gamma nail versus sliding hip screw for extracapsular hip fractures in adults

A systematic review was conducted for assessing comparative effect of cephalocondylic intramedullary nails versus extramedullary fixation implants for treating extracapsular hip fractures in adults [47]. Numerous endpoints included operative details, fractures fixation complications, and postoperative complications. Study results were integrated through multiple univariate meta-analysis. We performed a

Table III. Bivariate and univariate meta-analysis results of example 1: surgical versus nonsurgical procedure for treating periodontal disease.

	Probing depth $(I_1^2 = 0.69)$			Attachment level $(I_2^2 = 0.96)$			
Bivariate method	$\widehat{\beta}_1$(SE)	95% CI	$\widehat{\tau}_1^2$	$\widehat{\beta}_2$(SE)	95%CI	$\widehat{\tau}_2^2$	$\widehat{\rho}_b$
U	0.354(0.059)	[0.188, 0.521]	0.012	−0.342(0.104)	[−0.631, −0.052]	0.048	0.615
REML	0.353(0.059)	[0.19, 0.517]	0.012	−0.339(0.088)	[−0.583, −0.095]	0.033	0.609
MMM	0.348(0.056)	[0.193, 0.503]	0.01	−0.34(0.113)	[−0.655, −0.026]	0.057	0.748
Univariate method	$\widehat{\beta}_1$(SE)	95% CI	$\widehat{\tau}_1^2$	$\widehat{\beta}_2$(SE)	95% CI	$\widehat{\tau}_2^2$	
U	0.361(0.06)	[0.194, 0.528]	0.012	−0.346(0.104)	[−0.635, −0.057]	0.048	
REML	0.361(0.059)	[0.196, 0.525]	0.012	−0.346(0.088)	[−0.591, −0.1]	0.033	
MM	0.359(0.056)	[0.203, 0.515]	0.01	−0.346(0.113)	[−0.66, −0.032]	0.057	

U, U-statistic; REML, restricted maximum likelihood; MMM, multivariate method of moments; MM, univariate method of moments.

Table IV. Data of example 2: Gamma nail versus sliding hip screw for extracapsular hip fractures in adult.

	Sample size		Length of surgery (minutes)				Operative blood loss (ml)			
Study	G	S	Mean (G)	SD (G)	Mean (S)	SD (S)	Mean (G)	SD (G)	Mean (S)	SD (S)
1	203	197	55.4	20	61.3	22.2	244.4	384.9	260.4	325.5
2	60	60	47.1	20.8	53.4	8.3	152.3	130.7	160.3	110.8
3	73	73	65	29	51	22	240	190	280	280
4	53	49	59	23.9	47	13.3	258.7	145.4	259.2	137.5
5	104	106	46	11	44	15	NA	NA	NA	NA
6	31	36	56.7	17	54.3	16.4	NA	NA	NA	NA
7	93	93	NA	NA	NA	NA	814	548	1043	508

G, Gamma nail; S, SHS; NA, not available; SD, standard deviation.

bivariate meta-analysis to compare Gamma nail (an intramedullary nail) and sliding hip screw (an extramedullary implant) with two outcome measures of length of surgery (in minutes) and amount of operative blood loss (in milliliters). Advantages of Gamma nail over sliding hip screw are hypothesized to be reduced blood loss and shorter operative time [47].

The data presented in Table IV contains seven studies, among which four studies reported both outcomes, six reported length of surgery only, and five reported operative blood loss only. In this meta-analysis, the effect sizes Y_1 (length of surgery) and Y_2 (operative blood loss) were defined using Hedges' g [48]. For illustration purpose, we analyzed data under two assumptions: (i) $\rho_w > 0$ in those four studies, providing both outcomes as the amount of operative blood loss is likely to be positively associated with length of surgery and (ii) missing data are considered MAR in those three studies providing only one outcome. We applied the three methods to the data in BRMA with moderate ($\rho_w = 0.5$) and strong ($\rho_w = 0.8$) within-study correlations, respectively. We considered univariate meta-analysis (URMA) a special case of BRMA by assuming $\rho_w = \rho_b = 0$.

Table V shows the estimates of $(\beta_1, \beta_2, \tau_1^2, \tau_2^2, \rho_b)$ and the 95%CI of $\widehat{\beta}$. Results of all three methods imply that Gamma nail was associated with longer length of surgery and less amount of blood loss, but these findings were not statistically significant.

In BRMA, when ρ_w was assumed 0.8, the three methods produced similar estimates of β_1 and β_2, though a slightly different estimate of the between-study variance matrix appeared in the REML. When ρ_w was changed to 0.5, the solutions from U-statistic and the MMM remained almost same except for smaller between-study correlations. As nearly 85% of the total variation in length of surgery was from the between-study variation, the REML obtained a smaller $\widehat{\tau}_1^2$, resulting in a relatively larger $\widehat{\beta}_1$ with a smaller standard error compared with the other two methods when $\rho_w = 0.5$. This behavior reflected the findings in the simulation studies when I^2 was extremely large. The two extreme values of $\widehat{\rho}_b$ obtained in the REML also agreed with the simulation studies where we found the REML estimate of between-study correlation was more likely to lie on the boundary of the parameter space than the estimates from the other two methods.

Table V. Bivariate and univariate meta-analysis results of example 2: Gamma nail versus sliding hip screw for extracapsular hip fractures in adults.

		Length of surgery $(I_1^2 = 0.85)$			Operative blood loss $(I_2^2 = 0.3)$			
Correlation ρ_w	Method	$\widehat{\beta}_1$(SE)	95% CI	$\widehat{\tau}_1^2$	$\widehat{\beta}_2$(SE)	95%CI	$\widehat{\tau}_2^2$	$\widehat{\rho}_b$
0.8 (BRMA)	U	0.135(0.167)	[−0.274, 0.545]	0.142	−0.158(0.074)	[−0.34, 0.023]	0.007	−1
	REML	0.138(0.164)	[−0.263, 0.539]	0.136	−0.162(0.078)	[−0.353, 0.028]	0.01	−1
	MMM	0.135(0.168)	[−0.276, 0.546]	0.143	−0.159(0.076)	[−0.346, 0.028]	0.008	−0.927
0.5 (BRMA)	U	0.136(0.168)	[−0.276, 0.548]	0.142	−0.154(0.075)	[−0.337, 0.029]	0.007	−0.765
	REML	0.148(0.166)	[−0.258, 0.554]	0.139	−0.16(0.076)	[−0.346, 0.025]	0.008	−1
	MMM	0.137(0.169)	[−0.277, 0.55]	0.143	−0.155(0.077)	[−0.345, 0.034]	0.008	−0.718
0 (URMA)	U	0.117(0.169)	[−0.318, 0.553]	0.142	−0.142(0.076)	[−0.354, 0.069]	0.007	(0)
	REML	0.117(0.169)	[−0.318, 0.552]	0.141	−0.144(0.081)	[−0.369, 0.082]	0.01	(0)
	MM	0.117(0.17)	[−0.32, 0.546]	0.143	−0.143(0.079)	[−0.362, 0.076]	0.008	(0)

BRMA, bivariate meta-analysis; URMA, univariate meta-analysis; U, U-statistic; REML, restricted maximum likelihood; MMM, multivariate method of moments; MM, univariate method of moments.

In URMA, the three methods performed similarly in every respect. Compared with BRMA, smaller pooled estimates were obtained in URMA. Explanation of this difference between the solutions under BRMA and URMA because of the effect of ignoring the within-study correlation is well explained in [7]. Compared with the URMA that ignores the within-study correlation, in the BRMA, the estimation of length of surgery outcome is more likely to be influenced by the within-study correlation as the relative size of the within-study variation for operative blood loss outcome was large $\left(I_2^2 = 0.3\right)$. Additionally, Riley *et al.* [5] compared URMA and BRMA using REML and showed that in BRMA, by modeling β_1 and β_2 simultaneously, $\widehat{\beta}_1\left(\widehat{\beta}_2\right)$ 'borrows strength' from the second (first) outcome, despite the fact that the first (second) outcome is missing in few studies. This leads to more precise estimation with smaller standard errors and narrower CIs under BRMA. The statement is also confirmed in the current study in all three methods with narrower CIs in BRMA than URMA.

5. Discussion

A new approach to multivariate meta-analysis on the basis of the theory of U-statistic is proposed as an alternative to the commonly used parametric REML method and the recently developed MMM method. This method does not depend on parametric distributional assumption for both random effect $\mathbf{b}_i$ and sampling error ε_i in Equation (2) and thus provides robust inference irrespective of the data distribution. In addition, the proposed approach is computationally less intensive because of its non-iterative parameter estimation process and stands to gain from this feature when information on a large number of outcomes are being combined in a multivariate meta-analysis. Computational program in R is available upon request from the first author.

Through extensive simulation studies with data generated under normal and skewed distributions, it is shown that the effect on estimates from REML because of non-normal data distribution is marginal and that estimates from MMM and U-statistic-based approaches are very similar. Therefore, we conclude that for performing multivariate meta-analysis, the U-statistic estimation procedure is a viable alternative to REML and MMM.

Throughout the paper, by convention, the within-study variances were assumed known and replaced by their sample estimates. Thus, imprecision in within-study variance estimates may affect the estimation of pooled effect size especially when the size of within-study variation is relatively large. In addition to pooled effect size $\widehat{\beta}$ and its associated uncertainty, Higgins *et al.* [49] suggested reporting a prediction interval even in frequentist meta-analysis. The idea was to reflect the whole distribution of underlying effects in random-effects models. In contrast to the usual 95%CI, a prediction interval is defined as $\widehat{\beta} \pm t_{0.025,(n-2)}\left(\widehat{\tau}^2 + \mathrm{SE}^2\left(\widehat{\beta}\right)\right)^{1/2}$, where $\widehat{\tau}^2 + \mathrm{SE}^2(\widehat{\beta})$ is expected to reflect the variability resulting from the future studies that are currently not included in the meta-analysis but are similar to the studies included. As $\widehat{\tau}^2$ influences both terms in the variance, a prediction interval is more sensitive to the estimation of τ^2 than a CI, especially to the inflated bias in $\widehat{\tau}^2$ appearing in all three methods when the relative size of between-study variation is small. Currently, this concept is underutilized and should be promoted. In addition, the use of a t-distribution in the prediction interval under skewed distributions may not be appropriate and needs further research.

Future research utilizing U-statistic could be developed for testing significance of between-study heterogeneity and for performing meta-regression. A test of between-study homogeneity using the proposed estimates in this paper can be formulated as

$$\mathrm{H}_0 : L\eta = \mathbf{0} \quad \text{versus} \quad \mathrm{H}_a : L\eta \neq \mathbf{0}, \tag{13}$$

where L is a $p \times 3$ full rank matrix of known constants. For example, when $L = (1, 0, 0)$, the hypothesis in Equation (13) is equivalent to testing the between-study homogeneity in the first outcome. The Wald statistic, $Q_n^2 = n\widehat{\eta}^\top L^\top \left(\widehat{\Sigma}_\eta\right)^{-1} L\widehat{\eta}$, which has an asymptotic central χ^2 distribution with p degrees of freedom under the null H_0, can be used for significance testing. However, similar to other tests of homogeneity, the test built on the U-statistic estimates in Equation (13) might also be underpowered when the number of studies in a meta-analysis is small. A distribution-free approach for meta-regression could also be developed on the basis of the U-statistics proposed in this paper. Meta-regression based on mixed effects model (parametric approach) is usually considered where study-specific covariates explain part of the heterogeneity between studies and random effects account for the remainder. The U-statistic

approach can be easily generalized for incorporating covariates into the estimation of heterogeneity. Research on meta-regression always needs careful considerations as additional biases are introduced by involving covariates from different studies, which may cause spurious results [41].

APPENDIX A.

By the theorem of U-statistic, $\widehat{\eta}$ in Equation (8) is consistent and asymptotically normal:

$$\begin{aligned} \widehat{\eta} \to_p \eta, \quad \sqrt{n}\left(\widehat{\eta}-\eta\right) \to_d N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right), \\ \boldsymbol{\Sigma}_\eta = 4Var\left(E\left(\mathbf{h}_{ii'} \mid \mathbf{Y}_i\right)\right), \end{aligned} \tag{14}$$

where $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively.

To find a consistent estimate of $\boldsymbol{\Sigma}_\eta$, first note that

$$\begin{aligned} E\left(h_{1ii'} \mid \mathbf{y}_i\right) &= \frac{1}{2}\frac{w_{1ii'}\left(y_{i1}^2 - 2\beta_1 y_{i1} + \beta_1^2 + \tau_1^2 - \sigma_{i1}^2\right)}{\sum_{(i,i')\in C_2^n} w_{1ii'}}, \\ E\left(h_{2ii'} \mid \mathbf{y}_i\right) &= \frac{1}{2}\frac{w_{2ii'}\left(y_{i2}^2 - 2\beta_2 y_{i2} + \beta_2^2 + \tau_2^2 - \sigma_{i2}^2\right)}{\sum_{(i,i')\in C_2^n} w_{2ii'}}, \\ E\left(h_{3ii'} \mid \mathbf{y}_i\right) &= \frac{1}{2}\frac{w_{3ii'}\left(y_{i1}y_{i2} - \beta_1 y_{i2} - \beta_2 y_{i1} + \beta_1\beta_2 + \tau_{12} - \sigma_{i12}\right)}{\sum_{(i,i')\in C_2^n} w_{3ii'}}, \\ E\left(\mathbf{h}_{ii'} \mid \mathbf{y}_i\right) &= \left(E\left(h_{1ii'} \mid \mathbf{y}_i\right), E\left(h_{2ii'} \mid \mathbf{y}_i\right), E\left(h_{3ii'} \mid \mathbf{y}_i\right)\right)^\top. \end{aligned} \tag{15}$$

Thus, a consistent estimate of $\boldsymbol{\Sigma}_\eta$ is given by the following:

$$\widehat{\boldsymbol{\Sigma}}_\eta = \frac{4}{n-1}\sum_{i=1}^{n}\left[\left(\widehat{E}\left(\mathbf{h}_{ii'} \mid \mathbf{y}_i\right) - \widehat{\eta}\right)\left(\widehat{E}\left(\mathbf{h}_{ii'} \mid \mathbf{y}_i\right) - \widehat{\eta}\right)^\top\right],$$

where $\widehat{E}\left(\mathbf{h}_{ii'} \mid \mathbf{y}_i\right)$ denotes $E\left(\mathbf{h}_{ii'} \mid \mathbf{y}_i\right)$ given in Equation (15) with estimated $(\widehat{\eta}_1, \widehat{\eta}_2, \widehat{\eta}_3)$ in Equation (8) and $\left(\widehat{\beta}_1, \widehat{\beta}_2\right)$ in Equation (5) substituted in place of $\left(\tau_1^2, \tau_2^2, \tau_{12}\right)$ and (β_1, β_2).

Acknowledgements

The following grants supported this study: Clinical Translational Science Center (UL1-RR024996), Center for Education and Research in Therapeutics (AHRQ 1 U18 HS016075), and Collaborative Program in Nutrition and Prevention (R25CA105012). We sincerely thank Drs. Dan Jackson and Richard Riley for sharing data and computational programs with us. We also acknowledge an associate editor, two anonymous reviewers, three of our colleagues Drs. Matthias Kormaksson, Partha P. Debroy, and Xin M. Tu for their constructive critiques that help improve this manuscript immensely.

References

1. Peto R. Why do we need systematic overviews of randomized trials? *Statistics in Medicine* 1987; **6**:233–240.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
3. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
4. Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *Journal of Educational Statistics* 1985; **10**(2):75–98.
5. 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.
6. Gill JB, Chin Y, Levin A, Feng D. The use of antifibrinolytic agents in spine surgery: a meta-analysis. *The Journal of Bone and Joint Surgery* 2008; **90**:2399–2407.
7. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A* 2009; **172**:789–811.
8. Lajeunesse MJ. Meta-analysis and the comparative phylogenetic method. *The American Naturalist* 2009; **174**(3):369–381.
9. Hendriks AJ, MassDiepeveen JLM, Heugens EHW, van Straalen NM. Meta-analysis of intrinsic rates of increase and carrying capacity of populations affected by toxic and other stressors. *Environmental Toxicology and Chemistry* 2005; **24**:2267–2277.

10. Gripenberg S, Mayhew PJ, Parnell M, Roslin T. A meta-analysis of preference-performance relationships in phytophagous insects. *Ecology Letters* 2010; **13**(3):383–393.
11. Mugavin ME. A meta-synthesis of filicide classification systems: psychosocial and psychodynamic issues in women who kill their children. *Journal of Forensic Nursing* 2005; **1**(2):65–72.
12. Conner KR, Pinquart M, Gamble SA. Meta-analysis of depression and substance use among individuals with alcohol use disorders. *Journal of Substance Abuse Treatment* 2009; **37**(2):127–137.
13. van der Feltz-Cornelis CM, Van Os TW, Van Marwijk HW, Leentjens AF. Effect of psychiatric consultation models in primary care. A systematic review and meta-analysis of randomized clinical trials. *Journal of Psychosomatic Research* 2010; **68**(6):521–533.
14. Stratton RJ, Bircher G, Fouque D, Stenvinkel P, de Mutsert R, Engfer M, Elia M. Multinutrient oral supplements and tube feeding in maintenance dialysis: a systematic review and meta-analysis. *American Journal of Kidney Diseases* 2005; **46**(3):387–405.
15. Weng HQ, He XL, Li FX, Liu T, Wang PZ. Enteral versus parenteral nutrition after gastrointestinal surgery: a meta-analysis of randomized controlled trials. *Zhonghua Wai Ke Za Zhi [Chinese Journal of Surgery]* 2009; **47**(18):1368–1373.
16. Shrimpton R, Huffman SL, Zehner ER, Darnton-Hill I, Dalmiya N. Multiple micronutrient supplementation during pregnancy in developing-country settings: policy and program implications of the results of a meta-analysis. *Food & Nutrition Bulletin* 2009; **30**(4):S556–S573.
17. Portnoy DB, Scott-Sheldon LA, Johnson BT, Carey MP. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. *Preventive Medicine* 2008; **47**(1):3–16.
18. Darling EK, McDonald H. A meta-analysis of the efficacy of ocular prophylactic agents used for the prevention of gonococcal and chlamydial ophthalmia neonatorum. *The Journal of Midwifery & Women's Health* 2010; **55**(4):319–327.
19. Roth DE, Richard SA, Black RE. Zinc supplementation for the prevention of acute lower respiratory infection in children in developing countries: meta-analysis and meta-regression of randomized trials. *International Journal of Epidemiology* 2010; **39**(3):795–808.
20. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 2009; **4**(5):e5738.
21. Ketefian S. The relationship of education and moral reasoning to ethical practice: a meta-analysis of quantitative studies. *Scholarly Inquiry for Nursing Practice* 2001; **15**(1):3–18.
22. Dierckx de Casterlé B, Izumi S, Godfrey NS, Denhaerynck K. Nurses' responses to ethical dilemmas in nursing practice: meta-analysis. *Journal of Advanced Nursing* 2008; **63**(6):540–549.
23. Su R, Rounds J, Armstrong PI. Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin* 2009; **135**(6):859–884.
24. Kuoppala J, Lamminpää A, Liira J, Vainio H. Leadership, job well-being, and health effects–a systematic review and a meta-analysis. *Journal of Occupational & Environmental Medicine* 2008; **50**(8):904–915.
25. Hershcovis MS, Turner N, Barling J, Arnold KA, Dupré KE, Inness M, LeBlanc MM, Sivanathan N. Predicting workplace aggression: a meta-analysis. *Journal of Applied Psychology* 2007; **92**(1):228–238.
26. Eagly AH, Karau SJ, Makhijani MG. Gender and the effectiveness of leaders: a meta-analysis. *Psychological Bulletin* 1995; **117**(1):125–145.
27. 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.
28. Mehrotra DV. Non-iterative robust estimators of variance components in with-in subject designs. *Statistics in Medicine* 1997; **16**:1465–1479.
29. Matsuyama Y, Sakamoto J, Ohashi Y. A Bayesian hierarchical survival model for the institutional effects in a multi-centre cancer clinical trial. *Statistics in Medicine* 1998; **17**:1893–1908.
30. Marshall EC, Spiegelhalter DJ. Comparing institutional performance using Markov chain Monte Carlo methods. In *In Statistical Analysis of Medical Data: New Developments*, Everitt BS, Dunn G (eds). Arnold: London, 1998; 229–249.
31. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Statistics in Medicine* 2001; **20**:453–472.
32. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analysis. *Statistics in Medicine* 2010; **29**:1282–1297.
33. Hoeffding W. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* 1948; **19**:293–325.
34. Lehmann EL. Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematical Statistics* 1951; **22**:165–179.
35. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**:837–845.
36. King TS, Chinchilli VM, Carrasco JL. A repeated measures concordance correlation coefficient. *Statistics in Medicine* 2007; **26**:3095–3113.
37. Ma Y, Tang W, Feng C, Tu XM. Inference for kappas for longitudinal study data: applications to sexual health research. *Biometrics* 2008; **64**:781–789.
38. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
39. White IR. Multivariate random effects meta-analysis. *The Stata Journal* 2008; **9**:40–56.
40. Demidenko E. *Mixed Models: Theory and Applications*. Wiley: New York, 2004.
41. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. Wiley: New York, 2000.
42. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**:1539–1558.

43. Frees EW, Valdez EA. Understanding relationships using copulas. *North American Actuarial Journal* 1998; **2**:1–25.
44. Fredricks GA, Nelsen RB. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference* 2007; **137**:2143–2150.
45. 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).
46. Antczak-Bouckoms A, Joshipura K, Burdick E, Tulloch JF. Meta-analysis of surgical versus non-surgical method of treatment for periodontal disease. *Journal of Clinical Periodontology* 1993; **20**:259–268.
47. Parker MJ, Handoll HH. Gamma and other cephalocondylic intramedullary nails versus extramedullary implants for extracapsular hip fractures in adults. *The Cochrane Database of Systematic Reviews* 2008; 3:CD000093. Wiley.
48. Hedges LV. Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 1981; **6**:107–128.
49. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A* 2009; **172**:137–159.