STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:670–686
Published online 11 May 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2913

Impact of approximating or ignoring within-study covariances in multivariate meta-analyses

K. Jack Ishak[1,2,*,†], Robert W. Platt[1,3], Lawrence Joseph[1,4] and James A. Hanley[1,5]

[1]*Department of Epidemiology and Biostatistics, McGill University, Montreal, Que., Canada*
[2]*Caro Research Institute, Montreal, Que., Canada*
[3]*The Montreal Children's Hospital Research Institute, McGill University, Montreal, Que., Canada*
[4]*Division of Clinical Epidemiology, Department of Medicine, Montreal General Hospital, Montreal, Que., Canada*
[5]*Division of Clinical Epidemiology, Department of Medicine, Royal Victoria Hospital, Montreal, Que., Canada*

SUMMARY

Multivariate meta-analyses are used to derive summary estimates of treatment effects for two or more outcomes from a joint model. In addition to treatment effects, these models also quantify the correlations between outcomes across studies. To be fully specified, the model requires an estimate of the covariance or correlations between outcomes observed in each study. These are rarely available in published reports, so that analysts must either approximate these or ignore correlations between effect estimates from the same studies. We examined the impact of errors in approximating within-study covariances on the parameters of multivariate models in a simulation study. We found that treatment effect and heterogeneity estimates were not strongly affected by inaccurate approximations, but estimates of the correlation between outcomes were sometimes highly biased. The potential for error is greatest when the covariance between outcomes within- and between-studies are of comparable scale. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: multivariate meta-analysis; multiple outcomes; unknown covariance; correlation

1. INTRODUCTION

Meta-analyses typically examine the effect of a treatment on a set of outcomes that are relevant in the context of the disease and treatment being studied. The standard practice is to perform separate meta-analyses to obtain summary estimates of the effect of treatment on each outcome.

*Correspondence to: K. Jack Ishak, 185 Dorval Ave., Suite 500, Dorval, Quebec, Canada H9S 5J9.
†E-mail: jishak@caroresearch.com

Contract/grant sponsor: Natural Sciences and Engineering Research Council of Canada
Contract/grant sponsor: Fonds de la Recherche en Santé du Québec
Contract/grant sponsor: Canadian Institutes of Health Research
Contract/grant sponsor: Fonds Québécois de la Recherche sur la Nature et les Technologies

Received 18 May 2006
Accepted 22 March 2007

This approach ignores a possible correlation between the outcomes and does not allow inferences about the *overall* effect of the treatment based on all outcomes, or variations in effect across outcomes. Alternatively, a joint meta-analysis is possible with multivariate (multiple-outcome) models that account for the dependence between observed estimates and quantify the correlation between treatment effects across studies. Thus, multivariate meta-analyses provide added insight as the models include parameters describing how the outcomes are jointly distributed.

Several applications of these methods have appeared [1–9], illustrating the potential flexibility and strength of the multivariate approach. These include meta-analyses of treatment effects on two or more related endpoints [1–4], where the objective may be, for example, to estimate the relative sizes of the effects or quantify the correlation between the outcomes. Others [5, 6] have meta-analyzed occurrence rates of a single endpoint in treated and control groups at different times; in this context, multivariate models can be used to measure treatment effects at different times without making assumptions about proportionality of effects, and to examine the correlations between the probabilities of the endpoint occurrence over time [6]. Multivariate methods have also been used to evaluate the relationship between baseline risks and treatment effect estimates for a given outcome across studies [7–9], sometimes called control-rate regression. Standard methods (such as meta-regression) can produce biased results in this situation since they are affected by correlated measurement errors between estimates from the same study; a hierarchical multivariate model can be used to isolate the structural relationship between baseline risks and treatment effects.

A commonly used multivariate meta-analytic model is a multivariate extension of the DerSimonian and Laird approach [10]. The model includes random effects for outcomes within studies, which is thought to be more appropriate for meta-analyses [11–13]. In the multivariate case, the joint distribution of the random-effects yields estimates of the correlations between outcomes across studies. The joint likelihood of parameters from each study is approximated by a multivariate normal distribution with covariance matrices obtained from each study and assumed known, conditional on a vector of random effects also arising from a multivariate normal distribution. As with the variances of estimates in univariate meta-analyses, the within-study covariance matrices must be assumed known (i.e. without error), to ensure identifiability of the other parameters of the model. The model can be applied to a set of continuous effect estimates (e.g. differences in means) or log-odds-ratio estimates for dichotomous endpoints or a combination of both.

While the variances of observed estimates are typically reported or can be derived from information in the reviewed publications, the within-study covariances or correlations are rarely available. Indeed, estimates of these quantities would be available only if authors performed a joint analysis of the individual patient data. This is rarely done, however, unless specific questions about the relationship between the outcomes were of interest in the studies. This was recognized as a potential limitation in the earliest applications of multivariate models [1, 2] and recommendations were made to change reporting practices to allow correct application of the multivariate method in the future; however, this advice has had little impact to date. Within-study covariances were reported in only one [3] of the nine multivariate meta-analyses mentioned above. In all other applications, the covariances were specified by either approximating them from external information about the correlation between the outcomes [1, 2], assuming independence (i.e. setting covariances to 0) [6, 8], using approximation techniques [5] or incorporating a common within-study correlation nuisance parameter in the model [4]. In the latter case, concern was raised about potential identifiability problems, especially when only a small portion of the included studies report both outcomes being examined.

DOI: 10.1002/sim

The impact of approximating or ignoring within-study correlations has raised concern about the potential use of multivariate meta-analyses [14] but has not been examined in much detail. Berkey *et al*. [2] compared results from meta-analyses where within-study covariances were specified from an approximation of the correlation between outcomes (assumed constant across studies) to those from analyses that assume independence in fixed-effects models. They found moderate changes in pooled effect estimates and slight to moderate changes in the corresponding standard errors (SEs). We sought to perform a more formal assessment for multivariate meta-analysis models with random effects in a simulation study. We compared the accuracy and precision of point estimates and coverage probabilities of interval estimates obtained from analyses where within-study covariances are observed to those obtained when these are ignored (i.e. outcomes assumed to be independent) or approximated using external information about the correlations between the outcomes. We examined the impact of errors made in approximating the covariances on pooled effect estimates and the between-study covariance matrix parameters, with particular attention to estimates of the correlation between outcomes across studies, as this parameter may be of interest in multivariate meta-analyses.

The following section presents the general multivariate meta-analysis model that is commonly employed. We then describe the simulation design and the analyses that were performed on the simulated data. This is followed by a discussion of the results.

2. BACKGROUND: MULTIVARIATE META-ANALYSIS

2.1. Multivariate data in meta-analyses

Consider a meta-analysis of N studies examining the effect of a treatment on two endpoints. The treatment effects may be expressed as differences in means (for continuous endpoints) or risks (for dichotomous endpoints), or log transformed odds, risk, or hazard ratio estimates. We represent the estimates recorded from the ith study by the vector $y_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}$, for $i = 1, 2, \ldots, N$. It is possible that one of the components of y_i may be missing, however, since not all studies will necessarily report both outcomes. As in univariate meta-analyses, it is necessary to record the variance of the estimates, denoted S^2_{11i} and S^2_{22i}. In the multivariate context, however, the models also require the observed correlations or covariance (S_{12i}) between outcomes from each study. In short, the covariance matrix of the estimates

$$S_i = \begin{pmatrix} S^2_{11i} & S_{12i} \\ S_{12i} & S^2_{22i} \end{pmatrix}$$

is needed to fully specify the multivariate model.

2.2. Multivariate model

The random-effects multivariate model assumes that the true underlying effect of the ith study, $\theta_i = \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix}$, arises from a bivariate normal (BVN) distribution:

$$\theta_i \sim \text{BVN}\left\{\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, D = \begin{pmatrix} D^2_{11} & D_{12} \\ D_{12} & D^2_{22} \end{pmatrix}\right\}$$

DOI: 10.1002/sim

where θ is the vector of true treatment effect parameters, and D is the covariance matrix of the random effects; its diagonal elements measure the degree of heterogeneity in the true treatment effects across studies, while D_{12} represents the covariance between the effects. The correlation between effects is derived from the components of D, as $\rho_D = D_{12}/\sqrt{D_{11}^2 \cdot D_{22}^2}$.

The likelihood function of the ith vector of observed estimates is based on a BVN distribution:

$$y_i|\theta_i \sim \text{BVN}\{\theta_i, S_i\}$$

where S_i is assumed exactly known to ensure identifiability of the model. Some studies may report only one of the outcomes. If unobserved values are not imputed, the likelihood term reduces to that corresponding to a univariate normal distribution. It is important to consider, however, the possibility that outcomes may be selectively omitted from publication [1]; that is, that those observations are not missing at random—a form of publication bias in the multivariate context that could affect inferences about both the treatment effects and their correlations.

Study and outcome-specific covariates can easily be incorporated by setting $\theta_i = X_i\beta$, where X_i is a $2 \times p$ matrix of covariates and β is a $p \times 1$ vector of parameters. By default, X_i would include one or more indicators to identify outcomes in θ_i and derive summary estimates for each outcome or differences in estimates between outcomes. For instance, setting

$$X_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and not allowing an intercept in the model would produce a separate summary estimate of effect for each outcome. Otherwise, setting

$$X_i = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

would produce a summary estimate for the second endpoint (the intercept estimate) and the difference between summary estimates of the first and second endpoints.

2.3. Estimation of parameters

Frequentist inference is typically done *via* maximum likelihood, restricted or residual maximum likelihood (REML), or generalized least squares (GLS); in the latter case, distributional assumptions about the data are only required for inference, not estimation. Analytical or closed-form solutions exist only in the case where all outcomes are reported in each study and all S_i are equal [15]. This is unlikely, however, since the S_i would almost always vary due to differing sample sizes. In general, estimation is carried out with iterative procedures like the expectation–maximization (EM), Newton–Raphson, or Fisher scoring algorithms. These are implemented in standard software for mixed models, like PROC MIXED in SAS/STAT software or the *lme* function in *R* (S or S-Plus); Berkey *et al*. describe an EM algorithm [3] for GLS and likelihood approaches. Multivariate meta-analysis can also be performed using Bayesian methods [4], in which case the models can be conveniently implemented in BUGS [16].

DOI: 10.1002/sim

3. SIMULATION STUDY: IMPACT OF ERRORS IN APPROXIMATING WITHIN-STUDY COVARIANCES (S_{12i})

Since within-study covariance estimates are rarely available, proxy values are often substituted to fully specify the model—e.g. by using an observed estimate of the correlation between outcomes reported in one of the studies in the meta-analysis, or one that is not included in the review. Since estimation procedures used for multivariate meta-analysis models rely on iterative procedures (rather than closed-form analytical solutions), it is difficult to estimate the impact of erroneous approximations of within-study covariances algebraically. Thus, we undertook a series of simulations to examine the accuracy and precision of estimates of model parameters from meta-analyses when approximations of varying accuracy are employed.

3.1. Identifying key parameters for simulations

To inform the choice of the parameters to manipulate in simulations, we examined the estimation formulae involved in commonly used algorithms. All estimation procedures involve the within-study covariance only through the total variance matrix $V_i = D + S_i$. More specifically, the S_i always appear as part of the inverse of the total covariance matrices (V_i^{-1}) where

$$V_i^{-1} = \frac{1}{(D_{11}^2 + S_{11i}^2)(D_{22}^2 + S_{22i}^2) - (D_{12} + S_{12i})^2} \begin{pmatrix} D_{22}^2 + S_{22i}^2 & -(D_{12} + S_{12i}) \\ -(D_{12} + S_{12i}) & D_{11}^2 + S_{11i}^2 \end{pmatrix}$$

Components of S are assumed known without error; variances are set to their observed values, while the covariance term is approximated. Therefore, inaccurate approximations of the S_{12i} will distort estimates of all components of V_i^{-1}, and since V_i^{-1} appears in estimators or score equations of β and D, both estimates of the treatment effects and the heterogeneity covariance matrix components may be affected by errors in specifying within-study covariances.

The *accuracy* of the approximations of the S_{12i} is clearly an important determinant of the potential extent of errors in estimates from the model. It seems plausible, however, that poor approximations may be more influential in certain situations. We hypothesize that the potential for bias is greater when S_{12i} is of similar or larger scale than D_{12}, since, otherwise, the covariance term in V_i will be dominated by D_{12}. For instance, suppose the true within-study correlation is 0.50, but we erroneously approximate it to be 0.25; then, in a study where $S_{11i}^2 = 10$ and $S_{22i}^2 = 5$, the covariance would be estimated to be 1.76, instead of 3.54. The impact of this error will be more important when D_{12} is in the same order or smaller than S_{12i} (e.g. $D_{12} < 5$) compared to a situation where D_{12} is of greater magnitude (e.g. $D_{12} > 20$).

We must consider, however, that the scale of the within-study covariance matrix is inversely proportional to the size of the study: small studies will likely produce less precise estimates (i.e. large values in S_i), while variances (and so covariances) from large studies will tend to be comparatively small. Thus, the condition $S_{12i} > D_{12}$ is more likely to be met in meta-analyses with small studies. Alternatively, the same may happen when treatment effects are fairly homogeneous across studies (i.e. components of D are relatively small).

In summary, the relative scale of S_{12i} compared to D_{12} is a key factor of the potential impact of errors in approximations of within-study covariances. Thus, factors that can affect this difference

DOI: 10.1002/sim

were used as the parameters of our simulations. More specifically, we manipulated the relative size of within- and between-study correlations, the sample sizes of studies included in the meta-analysis, and the degree of between-study heterogeneity. These are described further below.

3.2. Simulation strategy

We simulated data for meta-analyses of the effect of a treatment on two correlated continuous outcomes from a hierarchical structure where we first *created* a study by assigning a set of true parameter values and then generating the responses of each patient in the study. Simulating patient-level data was necessary to allow calculation of within-study covariances; this would not be possible, for instance, by simulating observed effect estimates directly as in other simulations [3].

We assume N studies are included in the meta-analysis and n_i subjects are included in the (parallel) treated and control arms of the ith study (total sample size $= 2n_i$). We denote the *true* response of patients to treatment (e.g. reduction in blood pressure or lipid levels) by $\delta^t = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$. Patients in the control groups are assumed to receive placebo; for simplicity, we also assume that the disease remains stable, so that no change would be expected in patients' condition, i.e. $\delta^c = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Thus, δ^t also represents the true treatment effect.

Studies are likely to vary with respect to methodological design, measurement methods, study populations, and various other factors that might affect the true values of parameters. To emulate this, we added random effects to generate the true underlying responses of the ith study, δ_i^t and δ_i^c. We assume the variability of treatment effects across studies is given by the 2×2 covariance matrix $D = \text{var}(\delta_i^t - \delta_i^c)$, so that the covariance of responses in each arm is $0.5D$. Within a given study, we expect variability in the responses of different subjects, possibly in a correlated way. We assume the covariance matrix of the two responses for each subject is given by $0.5S$.

We generate data for studies from the following steps:

(1) Define a scenario by fixing values for N, n_i, δ^t, D, and S. The covariance matrices are specified by setting the variances (diagonal elements) and the correlation between random effects (ρ_D) and within-study responses for the two outcomes (ρ_S). We assume the same correlation in treated and control arms; that is, treatment does not alter the association between outcomes (although this may well happen in practice). The covariance component of each matrix (off-diagonal element) can then be derived from the correlations and variances.

(2) Generate the underlying true responses, δ_i^t and δ_i^c, for studies $i = 1, 2, \ldots, N$, by drawing $\delta_i^t \sim \text{BVN}(\delta^t, 0.5D)$ and $\delta_i^c \sim \text{BVN}(\delta^c, 0.5D)$. This allows the possibility of a correlation between responses from a given study, which may occur, for instance, if factors that affect one of the outcomes affect the other in a similar (or opposite) way.

(3) Simulate the responses of subjects in each study. For the jth subject of the ith study, we obtain $d_{ij}^t | \delta_i^t \sim \text{BVN}(\delta_i^t, 0.5S)$ and $d_{ij}^c | \delta_i^c \sim \text{BVN}(\delta_i^c, 0.5S)$. We denote by d_i^t and d_i^c the $n_i \times 2$ matrices of responses for all patients in the ith study generated from the BVN draws.

(4) Calculate estimates of effect and corresponding observed covariance matrix for each study. This is given by

$$y_i = \begin{pmatrix} \overline{d}_{1i}^t - \overline{d}_{1i}^c \\ \overline{d}_{2i}^t - \overline{d}_{2i}^c \end{pmatrix} \quad \text{and} \quad S_i = \frac{1}{n_i}(\text{cov}(d_i^t) + \text{cov}(d_i^c))$$

DOI: 10.1002/sim

where the covariance matrices of the observed data, d_i^t and d_i^c, are calculated empirically, using standard formulae.

Each replication of steps 2–4 creates a new set of studies for meta-analysis within the scenario defined in step 1. The programs used to generate the data are available from the authors.

3.3. Simulation scenarios

Scenarios were created by manipulating the within- and between-study correlations, the sample sizes of studies and degree of heterogeneity in effects. We manipulated the values of these parameters to create situations where within- and between-study covariances were of comparable size, since we expected the potential impact of errors in approximations would be most apparent in this case.

Table I summarizes the parameters of the simulations. In all scenarios we considered meta-analyses of $N = 25$ studies, and assumed the true response to treatment is a reduction of 10 units for the first outcome and a reduction of 5 units for the second, i.e. $\delta^t = (-10 \ \ -5)$, while no change is expected among the controls, i.e. $\delta^c = (0 \ \ 0)$.

We first examine a set of scenarios we consider to be reflective of most practical situations: we assume the degree of heterogeneity between studies is not so large as to completely dominate within-study variances, while at the same time the studies are not so small as to produce highly imprecise estimates. We then consider more extreme cases where we reduce sample sizes and minimize between-study heterogeneity. We start with studies including between 50 and 250 subjects per arm to create studies of moderate size. (See Table I for actual sample sizes, chosen randomly, used in simulations.) We set the between-study heterogeneity covariance matrix of treatment effects to be $D_{11}^2 = 5.0$, $D_{22}^2 = 2.5$ and $\rho_D = 0.25$, so that $D_{12} = 0.25\sqrt{2.5 \times 1.25} = 0.88$. Between-patient variances were set to $S_{11}^2 = 10.0$ and $S_{22}^2 = 5.0$, twice the magnitude of between-study variances, since one would expect less variability in aggregated measures (e.g. mean group *versus* subject-specific difference). We note that the variance of observed estimates from the ith study will be $S_{i11}^2 \approx 10.0/n_i$ and $S_{i22}^2 \approx 5.0/n_i$.

Table II summarizes the scenarios that we considered in this study in more detail. In a first pair of scenarios, we varied the relative magnitude of within- and between-study correlations since larger within-study correlation would lead to larger within-study covariances, and as a result, the impact of errors would potentially be greater. We considered the following cases:

Scenario 1a: Assume equal correlations: $\rho_S = \rho_D = 0.25$.

Scenario 1b: Assume stronger within-study correlation: $\rho_S = 2\rho_D = 0.5$.

We adopted scenario 1b as the basis the next set of scenarios where we varied the sample size of the studies. Smaller studies will produce more variable estimates—i.e. larger S_{i11}^2, S_{i22}^2—and, consequently, larger within-study covariances that would be closer in magnitude to between-study covariances. We considered the following cases:

Scenario 2a: Reduce sample sizes by a factor of $1/5$, so that the size of each arm ranges from 10 to 50.

Scenario 2b: Reduce sample sizes by a factor of $1/10$, so that the size of each arm ranges from 5 to 25.

Next, we examined the potential impact of the relative size of within- and between-study variances (based on scenario 1b). Here again, by reducing between-study variances (or heterogeneity), we reduce the gap between within- and between-study covariances making the impact of errors in

DOI: 10.1002/sim

Table I. Parameters used to define scenarios of simulations.

Parameter	Base value	Fixed/variable	Rationale
Number of studies (N)	25	Fixed	Chosen to reflect a meta-analysis where a multivariate analysis can be done reliably
Sample sizes (per arm, $n_i/2$)	230 231 160 152 248 116 198 230 158 190 55 209 60 238 88 241 138 177 108 140 54 219 208 131 99	Variable	Sample sizes were reduced in other scenarios, to create situations where within- and between-study covariances are closer in magnitude
Response among treated (δ^t)	$-10, -5$	Fixed	Held fixed since not expected to affect the impact of errors in approximations (verified by exploratory analyses not reported here)
Response among controls (δ^c)	0, 0	Fixed	Held fixed since not expected to affect the impact of errors in approximations (verified by exploratory analyses not reported here)
Between-study			
Variance (D_{11}^2, D_{22}^2)	5, 2.5	Variable	Considered scenarios with lower heterogeneity, holding within-study variances fixed, to create situations where within- and between-study covariances are closer in magnitude
Correlation (ρ_D)	0.25	Fixed	Held fixed, but varied within-study correlation to manipulate relative magnitude of within- and between-study correlations
Within-study			
Variance (S_{11}^2, S_{22}^2)	10, 5	Fixed	Held fixed, but varied within-study variances to manipulate relative magnitude of within- and between-study variances
Correlation (ρ_S)	0.25	Variable	Considered scenarios with strong within-study correlation, holding the between-study correlation fixed, to create situations where within- and between-study covariances are closer in magnitude

Only parameters expected to affect the potential impact of errors in approximations of within-study covariances were varied.

DOI: 10.1002/sim

Table II. Definition of scenarios considered in simulations.

	Sample size deflation factor	Treatment effect		Between-study variance		Within-study variance		Correlation	
		δ_1	δ_2	D_{11}^2	D_{22}^2	S_{11}^2	S_{22}^2	ρ_D	ρ_S
Scenario 1a	1.0	−10.0	−5.0	5.0	2.5	10.0	5.0	0.25	0.25
Scenario 1b	1.0	−10.0	−5.0	5.0	2.5	10.0	5.0	0.25	0.50
Scenario 2a	1/5	−10.0	−5.0	5.0	2.5	10.0	5.0	0.25	0.50
Scenario 2b	1/10	−10.0	−5.0	5.0	2.5	10.0	5.0	0.25	0.50
Scenario 3a	1.0	−10.0	−5.0	1.0	0.5	10.0	5.0	0.25	0.50
Scenario 3b	1.0	−10.0	−5.0	0.5	0.25	10.0	5.0	0.25	0.50

The number of studies was 25 in all the scenarios.

approximation potentially more influential. We considered the following cases:

Scenario 3a: Reduced between-study variances by a factor of 1/5, so that $D_{11}^2 = 1.0$ and $D_{22}^2 = 0.5$.

Scenario 3b: Reduced variances by half (1/10 of the original value), so that $D_{11}^2 = 0.5$ and $D_{22}^2 = 0.25$.

3.4. Errors in approximating within-study covariances

Within-study covariances are usually approximated from an estimate of the correlation between outcomes, denoted ρ_S^*, and calculating $S_{12i} = \rho_S^* \times S_{11i} \times S_{22i}$ for each study. In practice, ρ_S^* may be based on expert opinion or observed from a study that may or may not be included in the meta-analysis, or obtained by other means. Clearly, an important determinant of the impact of approximating unobserved covariances is the accuracy of the approximation itself. We consider the following for our simulations:

1. Setting $\rho_S^* = 0$; that is, conducting a multivariate meta-analysis where within-study correlations are ignored.
2. Setting $\rho_S^* = 1.5 \times \rho_S$; that is, the correlations are overestimated by 50 per cent.
3. Setting $\rho_S^* = 0.5 \times \rho_S$; that is, the correlations are underestimated by 50 per cent.
4. Setting $\rho_S^* = -0.25$; the direction and size of the correlation are wrong.

We compared results from meta-analyses using these four approximations to those where within-study covariance estimates were available (the reference, or best-case results).

4. ANALYSIS AND COMPARISON MEASURES

For each scenario, we simulated 5000 meta-analyses, each of which used the observed and approximated within-study covariances. Parameter estimates from replications of the five meta-analyses in each scenario were compared with respect to the bias and precision of estimates and coverage of confidence intervals.

DOI: 10.1002/sim

The models were fitted by REML with PROC MIXED using SAS/STAT software (Version 8e), as described by Van Houwelingen *et al.* [8]. Estimation was carried out by Fisher scoring instead of the standard Newton–Raphson algorithm to avoid non-positive-definite Hessian matrices, which may cause estimation to fail for some meta-analyses. Since within-study variances and covariances must be specified and fixed, the procedure also requires that starting values be provided for the components of the between-study covariance matrix. We calculated these from the observed data in each meta-analysis.

For each of the five meta-analysis models, we recorded the estimates of summary treatment effects on events A and B (i.e. $\hat{\theta}$), the between-study variances (i.e. $\hat{D}^2_{11}$, $\hat{D}^2_{22}$) and the correlation between effects across studies (i.e. $\hat{\rho}_D = \hat{D}_{12}/\sqrt{\hat{D}_{11}\hat{D}_{22}}$) with corresponding SEs and/or confidence intervals. We compared the performance of the models with respect to the relative bias in estimates (bias = (estimate − true value)/true value), precision, and coverage probabilities of the confidence intervals of the estimates of each parameter.

We summarized the median, 5th and 95th percentiles of the distribution of estimates from the 5000 replications. We favored percentiles over means and SEs, since the latter may be affected by extreme results and could be misleading if the distribution of results from the simulations are skewed. We calculated the coverage probabilities of the confidence intervals of estimates of each parameter as the proportion of runs of each scenario where the confidence interval included the true value.

5. RESULTS

5.1. Estimation of treatment effects

Table III summarizes the biases, SEs, and coverage probabilities of confidence intervals for the effect of the treatment on outcomes A. The table shows the median (5th–95th percentiles) of the relative bias, the SE, and coverage probabilities of confidence intervals of estimates from meta-analyses with observed and approximated within-study covariances across the six scenarios.

In the best-case scenario where within-study covariance estimates are available, the treatment effect on the first outcome was estimated with little or no bias: the median bias was near 0 in all cases and was generally within ±8.0 per cent across replications. The coverage probability of confidence intervals was slightly below 95 per cent. Increasing the relative magnitude of within- and between-study correlations (scenarios 1a *versus* 1b) and decreasing the size of studies (scenarios 2a and 2b) caused only slight (negligible) increases in the magnitude of the bias and SE of estimates and a slight decrease in coverage probabilities. Reducing the between-study heterogeneity (scenarios 3a and 3b), on the other hand, lead to comparatively smaller bias, and more precise estimates but the coverage of confidence intervals remained below the nominal value.

Results were essentially identical when within-study correlations were not observed and approximated. The relative magnitude of errors in approximations had no impact on the performance of meta-analyses; the bias and precision of estimates were the same even when the direction of the correlations assumed to derive approximations was incorrect (i.e. last column of Table III).

Findings were very similar for estimates of the treatment effect on outcome B (results not shown), although the magnitude of the bias tended to be slightly larger (5th and 95th percentiles were greater than ±10 per cent in scenarios 1a and b and 2a and b). While estimates were again

DOI: 10.1002/sim

Table III. Performance of estimates of the effects of treatment on the first outcome: median (5th, 95th percentiles) relative bias, standard error and coverage of confidence intervals obtained from 5000 simulated meta-analyses.

	Within-study correlation				
	Observed estimate	Approximation underestimates by 50 per cent	Underestimated approximation by 50 per cent	Assume independent ($\rho_S^* = 0$)	Negative approximation ($\rho_S^* = -0.25$)
Scenario 1a					
Per cent bias	0.02 (−7.25, 7.25)	0.02 (−7.25, 7.25)	0.02 (−7.25, 7.25)	0.02 (−7.25, 7.25)	0.02 (−7.25, 7.25)
SE	0.44 (0.34, 0.56)	0.44 (0.34, 0.56)	0.44 (0.34, 0.56)	0.44 (0.34, 0.56)	0.44 (0.34, 0.56)
Coverage	94.4	94.4	94.4	94.4	94.4
Scenario 1b					
Per cent bias	−0.15 (−7.39, 7.34)	−0.15 (−7.39, 7.34)	−0.16 (−7.39, 7.34)	−0.15 (−7.39, 7.35)	−0.16 (−7.38, 7.34)
SE	0.44 (0.34, 0.55)	0.44 (0.34, 0.55)	0.44 (0.34, 0.55)	0.44 (0.34, 0.55)	0.44 (0.34, 0.55)
Coverage	93.7	93.7	93.6	93.7	93.7
Scenario 2a					
Per cent bias	0.05 (−7.54, 7.7)	0.07 (−7.51, 7.68)	0.06 (−7.53, 7.68)	0.08 (−7.52, 7.69)	0.07 (−7.52, 7.67)
SE	0.46 (0.35, 0.57)	0.46 (0.35, 0.57)	0.46 (0.35, 0.57)	0.46 (0.35, 0.57)	0.46 (0.35, 0.57)
Coverage	93.8	93.8	93.8	93.8	93.7
Scenario 2b					
Per cent bias	−0.09 (−8.04, 7.85)	−0.11 (−8, 7.87)	−0.1 (−8.03, 7.87)	−0.11 (−8.05, 7.88)	−0.11 (−7.96, 7.88)
SE	0.47 (0.36, 0.59)	0.47 (0.36, 0.59)	0.47 (0.36, 0.59)	0.47 (0.36, 0.59)	0.47 (0.36, 0.59)
Coverage	93.3	93.3	93.3	93.3	93.4
Scenario 3a					
Per cent bias	0.01 (−3.5, 3.44)	0.02 (−3.49, 3.43)	0.01 (−3.51, 3.44)	0.01 (−3.48, 3.44)	0.02 (−3.48, 3.44)
SE	0.2 (0.16, 0.26)	0.2 (0.16, 0.26)	0.2 (0.16, 0.26)	0.2 (0.16, 0.26)	0.2 (0.16, 0.26)
Coverage	93.5	93.4	93.5	93.4	93.4
Scenario 3b					
Per cent bias	0.02 (−2.53, 2.52)	0.03 (−2.53, 2.52)	0.03 (−2.53, 2.51)	0.03 (−2.52, 2.52)	0.02 (−2.53, 2.52)
SE	0.15 (0.11, 0.19)	0.15 (0.11, 0.19)	0.15 (0.12, 0.19)	0.15 (0.11, 0.19)	0.15 (0.12, 0.19)
Coverage	93.5	93.5	93.5	93.5	93.5

The true value of the parameter is −10.

DOI: 10.1002/sim

more accurate and precise in scenarios 3a and b, errors in approximations did not affect the performance of meta-analyses in estimating the treatment effect on outcome B.

5.2. Estimation of between-study variances

Table IV summarizes findings related to estimates of between-study variability (i.e. heterogeneity) of effects for outcome A. The true value of the parameter was 5.0 in scenarios 1a and b and 2a and b, and was then reduced to 1.0 in scenario 3a, and to 0.5 in scenario 3b.

Estimates of heterogeneity were prone to relatively large biases, even when within-study covariances were observed; the 5th and 95th percentiles suggest biases close to or exceeding ± 50 per cent in all scenarios, even when the heterogeneity was assumed to be very low (scenarios 3a and b). Reducing the sample sizes of studies (scenarios 2a and b) lead to larger biases in some replications and increased the width of confidence intervals, which in turn caused the coverage of these to exceed 95 per cent.

As for treatment effects, errors in approximating within-study covariances had no impact on the performance of meta-analyses in estimating these parameters. The distribution of the bias and precision of estimates were unchanged in all cases.

Findings were very similar for estimates of the heterogeneity of the treatment effect on outcome B (results not shown).

5.3. Estimation of between-study correlation between treatment effects

Table V summarizes findings about estimation of the between-study correlation parameter, which was assumed to be 0.25 in all scenarios.

Estimates of the correlation parameter were prone to the largest biases in estimates compared to treatment effect and heterogeneity parameters. This was the case even when within-study covariances were observed. Although the median relative bias was close to 0 in all scenarios, the 5th and 95th percentiles reveal biases in estimates exceeding 150 per cent in some replications. The precision of the estimates was relatively weak as well, with confidence interval widths of 0.8, which covers almost half of the full scale of possible values of the parameter. Despite relatively large biases in point estimates, confidence intervals had coverage probabilities around 90 per cent when within-study covariance estimates were observed.

As before, meta-analyses of smaller studies (scenarios 2a and b) or with lower heterogeneity tended to have estimates with slightly larger bias and weaker precision, when within-study covariances were observed. Unlike the other parameters, however, estimates of the correlation were more prone to be affected by errors in approximations of within-study covariances. When these were overestimated by 50 per cent, the median bias was increased and the coverage of confidence intervals was slightly reduced. Underestimating within-study covariances by 50 per cent was less influential, however, with results looking similar to those from the 'Observed' case. Ignoring (i.e. assuming independence) or using a negative correlation, which represent increasingly underestimated approximations, led to increasingly more positively biased estimates and confidence intervals with lower coverage, as low as 85 per cent, for instance in scenarios 2b and 3b.

Errors in approximations were most influential in scenarios with smaller studies (2a and b) and more homogeneous effects (3a and b), and the magnitude of the effect of errors increased as these parameters were varied (i.e. 2a *versus* 2b, 3a *versus* 3b).

DOI: 10.1002/sim

Table IV. Performance of estimates of the between-study variance of the effect of the first outcome (D_1^2): median (5th, 95th percentiles) relative bias, width and coverage of confidence intervals obtained from 5000 simulated meta-analyses.

	Within-study correlation				
	Observed estimate	Approximation underestimates by 50 per cent	Underestimated approximation by 50 per cent	Assume independent ($\rho_S^*=0$)	Negative approximation ($\rho_S^*=-0.25$)
Scenario 1a					
Per cent bias	−3.34 (−43.39, 53.46)	−3.37 (−43.34, 53.35)	−3.36 (−43.37, 53.51)	−3.36 (−43.37, 53.39)	−3.37 (−43.28, 53.3)
CI width	6.54 (3.89, 10.3)	6.54 (3.89, 10.3)	6.54 (3.89, 10.3)	6.54 (3.89, 10.3)	6.54 (3.89, 10.3)
Coverage	94.8	94.9	94.8	94.8	94.9
Scenario 1b					
Per cent bias	−3.82 (−43.96, 50.07)	−3.81 (−43.84, 50.13)	−3.84 (−43.94, 50.18)	−3.83 (−43.9, 50.1)	−3.78 (−43.79, 50.21)
CI width	6.51 (3.85, 10.09)	6.51 (3.85, 10.1)	6.51 (3.85, 10.09)	6.51 (3.85, 10.1)	6.51 (3.85, 10.1)
Coverage	95.1	95.1	95.1	95.1	95.1
Scenario 2a					
Per cent bias	−2.93 (−45.12, 56.11)	−2.93 (−44.81, 56.07)	−2.78 (−44.99, 56.29)	−2.93 (−44.9, 56.2)	−2.91 (−44.84, 56.22)
CI width	7.09 (4.29, 11.02)	7.09 (4.31, 11.02)	7.1 (4.3, 11.02)	7.09 (4.31, 11.01)	7.09 (4.31, 11.01)
Coverage	95.4	95.4	95.4	95.4	95.4
Scenario 2b					
Per cent bias	−2.38 (−47.24, 58.41)	−2.73 (−47.57, 57.79)	−2.57 (−47.29, 57.58)	−2.76 (−47.44, 57.67)	−2.47 (−47.5, 58.34)
CI width	7.74 (4.8, 11.79)	7.73 (4.79, 11.82)	7.73 (4.8, 11.79)	7.73 (4.81, 11.83)	7.74 (4.79, 11.84)
Coverage	96.2	96.1	96.1	96.1	96.0
Scenario 3a					
Per cent bias	−2.74 (−45.05, 56.49)	−2.78 (−44.95, 56.55)	−2.74 (−45.03, 56.77)	−2.74 (−44.92, 56.59)	−2.73 (−45.02, 56.68)
CI width	1.42 (0.87, 2.21)	1.42 (0.87, 2.22)	1.42 (0.87, 2.21)	1.42 (0.86, 2.21)	1.42 (0.87, 2.22)
Coverage	95.6	95.5	95.6	95.5	95.4
Scenario 3b					
Per cent bias	−3.45 (−48.12, 60.93)	−3.39 (−47.91, 61.12)	−3.34 (−47.88, 61.1)	−3.64 (−48.1, 60.81)	−3.21 (−47.65, 61.27)
CI width	0.77 (0.48, 1.2)	0.77 (0.48, 1.2)	0.77 (0.48, 1.2)	0.77 (0.48, 1.2)	0.77 (0.48, 1.2)
Coverage	95.8	95.8	95.8	95.8	95.7

The true value of the parameter is 5.0 in scenarios 1a–b, 2a–b; 1.0 in scenario 3a, and 0.5 in scenario 3b.

DOI: 10.1002/sim

Table V. Performance of estimates of the between-study correlation (ρ_D): median (5th, 95th percentiles) relative bias, standard error and coverage of confidence intervals obtained from 5000 simulated meta-analyses.

	Within-study correlation				
	Observed estimate	Approximation underestimates by 50 per cent	Underestimated approximation by 50 per cent	Assume independent ($\rho_S^* = 0$)	Negative approximation ($\rho_S^* = -0.25$)
Scenario 1a					
Per cent bias	2.57 (−135.92, 120.37)	4.01 (−134.21, 121.8)	1.55 (−136.67, 119.66)	3.2 (−135.04, 121.09)	5.55 (−132.61, 123.08)
CI width	0.76 (0.57, 0.81)	0.76 (0.57, 0.81)	0.76 (0.57, 0.81)	0.76 (0.57, 0.81)	0.76 (0.57, 0.81)
Coverage	91.5	91.4	91.5	91.4	91.4
Scenario 1b					
Per cent bias	1.56 (−140.45, 120.13)	4.89 (−136.82, 123.02)	−0.05 (−142.01, 118.43)	3.15 (−138.55, 121.45)	6.53 (−135.21, 124.37)
CI width	0.76 (0.57, 0.81)	0.76 (0.56, 0.81)	0.76 (0.57, 0.81)	0.76 (0.57, 0.81)	0.76 (0.56, 0.81)
Coverage	91.1	90.9	91.2	91.0	90.7
Scenario 2a					
Per cent bias	5.02 (−147.99, 127.23)	20.27 (−131.54, 140.54)	−2.46 (−155.7, 120.12)	12.78 (−139.2, 134.27)	427.76 (−123.39, 147.44)
CI width	0.8 (0.58, 0.88)	0.79 (0.57, 0.88)	0.8 (0.59, 0.88)	0.79 (0.58, 0.88)	0.79 (0.56, 0.88)
Coverage	90.9	89.4	91.3	90.3	88.3
Scenario 2b					
Per cent bias	0 (−163.46, 133.11)	29.52 (−130.41, 159.48)	−14.15 (−179.25, 119.65)	15.15 (−145.37, 146.67)	43.93 (−114.29, 171.51)
CI width	0.85 (0.61, 0.96)	0.84 (0.58, 0.96)	0.86 (0.63, 0.97)	0.84 (0.59, 0.96)	0.83 (0.56, 0.95)
Coverage	90.8	87.4	91.0	89.4	85.2
Scenario 3a					
Per cent bias	1.39 (−152.59, 128.79)	17.34 (−135.73, 141.97)	−6.64 (−161.6, 121.49)	9.46 (−144.16, 135.78)	25.19 (−127.68, 148.46)
CI width	0.8 (0.58, 0.88)	0.79 (0.56, 0.88)	0.81 (0.59, 0.88)	0.8 (0.57, 0.88)	0.79 (0.56, 0.87)
Coverage	90.0	88.7	90.2	89.7	87.8
Scenario 3b					
Per cent bias	1.6 (−159.4, 134.2)	31.83 (−123.9, 162.4)	−14.28 (−176.74, 119.58)	17.03 (−141.73, 149.12)	46.08 (−105.84, 177.08)
CI width	0.85 (0.61, 0.97)	0.84 (0.57, 0.96)	0.86 (0.63, 0.98)	0.85 (0.59, 0.96)	0.83 (0.55, 0.96)
Coverage	90.9	87.5	91.3	89.3	85.0

The true value of the parameter is 0.25.

DOI: 10.1002/sim

6. DISCUSSION

This paper examined the impact of errors made in approximating unobserved within-study covariances in multivariate meta-analyses on estimates of treatment effects, heterogeneity parameters and the correlation between outcomes across studies. We compared results from meta-analyses based on approximations of within-study covariances to those from the reference analysis where the actual covariances are observed. These were available since the data were simulated but are generally not available in practice. The approximations that were employed in the simulations were intentionally made to be of poor accuracy to test the models under *difficult* conditions—when little is known about the correlation between outcomes.

Our analyses revealed that treatment effects were reasonably well estimated, even when within-study covariances were poorly approximated. Heterogeneity and between-study correlation parameters were subject to larger biases, however, even when within-study covariances were observed. In fact, estimates of these parameters were relatively less precise, which, in the case of between-study variances, lead to confidence intervals with coverage exceeding the nominal value. For the correlation parameter, however, the large biases in estimates caused confidence intervals to miss the true value in up to 15 per cent of replications in some scenarios. Confidence intervals of treatment effect estimates were also slightly below the expected 95 per cent threshold. We attribute this in part to the relatively small size of the data (25 studies), which is typical in meta-analyses. This was confirmed in sensitivity analyses where we reduced the number of studies (results not shown) and observed greater decline in coverage probabilities. In the case of the correlation parameters, however, the problem may also be amplified by the fact that estimates are a function of estimates of the covariance between outcomes as well as the two variances; thus, they are prone to larger errors and variability. It is possible, however, that depending on the directions, errors in these estimates cancel each other out in some cases.

We also confirmed the hypothesis that the impact of errors in estimates would be most apparent in scenarios where within- and between-study covariances were of similar scale (scenarios 2a and b, 3a and b). However, the impact was most apparent in the accuracy of between-study correlation estimates; the precision of these estimates was poor in all analyses, even when covariances were observed, but declined only slightly because of errors in approximations. Treatment effects or heterogeneity parameters were generally similar in both meta-analyses based on observed or approximated covariances. Thus, if only treatment effects are of interest, one can even ignore within-study correlations and assume independence without any significant risk of bias or loss of precision in estimates. In meta-analyses where an estimate of the correlation between outcomes is desired, however, a comparison of the relative magnitude of within- and between-study variances in the outcomes (e.g. from univariate random-effects models), sample sizes of the studies, and crude estimates of the correlation from the recorded data can be used to assess the potential for bias. Like Berkey *et al*. [2], we recommend performing sensitivity analyses around the external estimates used to approximate within-study covariances.

We did not examine cases where estimates were more variable within-studies than between, as these are likely cases where effects are fairly homogeneous across studies and random-effects models may not be necessary or adequate. We can deduce from our findings, however, that correlation estimates in such meta-analyses may be susceptible to substantial variability and possibly large biases.

The scenarios reported in this paper were based on a fixed number of studies and assumed that all studies reported both outcomes and that the between-patient variance of the outcome was

DOI: 10.1002/sim

constant across studies. We examined the potential impact of varying these factors (not reported in the manuscript). As would be expected, there was a noticeable loss in precision when fewer studies were included in the meta-analysis or when not all studies reported both outcomes. Here again, estimates of between-study correlations were most affected, with larger biases and interval estimates with poor coverage probabilities. Similarly, allowing between-patient variances to be different in studies also made the correlation parameter difficult to estimate, but did not affect summary estimates of treatment effects or heterogeneity parameters significantly.

The relative magnitude of differences in the bias of correlation estimates from meta-analyses employing approximations were consistent with the relative accuracy of the approximations. Results from meta-analyses that overestimated or underestimated within-study correlations by 50 per cent (the least erroneous approximations) tended to be closest to those from the reference case. Those assuming independence and employing a negative correlation (the most erroneous approximation) to derive within-study covariances yielded increasingly overestimated between-study correlations, and declining coverage of confidence intervals. The same does not appear to occur when the within-study correlation was over-approximated by 50 per cent, since the median biases were positive and the replications seemed to produce larger positively biased estimates. To verify this further, we ran some simulations where the magnitude of the error in approximation was greater (over-approximated by 75 per cent). We found that the between-study correlation tended to be underestimated more frequently; for instance, for scenario 3b, the median (5th–95th percentiles) were -21.5 per cent (-186.1 to 112.1). This suggests that, the model may be compensating for errors in approximating within-study correlations, by over- or under-estimating the correlation across studies to *maintain* the *total* correlation measured in the data. This *balancing* of within- and between-study correlations may explain why summary treatment effect parameters were not affected by the errors in approximation: their estimation relies only on the total covariance matrices ($V_i = D + S_i$), which will tend to be accurate since estimates of D seem to counter errors in the specification of the S_i.

Although we simulated meta-analyses of continuous endpoints, our findings should also apply when multivariate models are used for two or more log-odds-ratio, log-rate-ratio or other measures. Within-study correlations are harder to specify for these measures, since by definition, they apply to populations or groups. That is, it is difficult to interpret within-study correlations between two-log-odds-ratios in an individualistic sense, in the same way that one can with continuous measures (e.g. the correlation between a patient's blood pressure and lipid levels). The relationship between dichotomous outcomes may be more easily described with probabilistic arguments (e.g. probability of having both outcomes *versus* probability of only one). Consequently, multivariate models based on this type of structure may be more useful and provide more information about the relationship between the outcomes in this setting.

In summary, multivariate meta-analysis models appear to be fairly robust to errors made in approximating within-study covariances when only summary effect estimates are of interest. In fact, assuming independence within studies and accounting for the correlations in the data through the random effects for each outcome can do as well as when within-study covariances are observed. Multivariate meta-analyses do not appear to be as reliable when interest lies in estimating correlations between outcomes; even when within-study covariances were observed, estimates of the correlations were prone to relatively large biases and lacked precision. The direction of the bias in between-study correlation estimates depended on whether within-study correlations were under- or over-approximated. Furthermore, the correlations measured across studies may not reflect the underlying association between treatment effects [17].

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

This work was supported by a doctoral fellowship from the Natural Sciences and Engineering Research Council of Canada to K. I., a Chercheur-boursier award from the Fonds de la Recherche en Santé du Québec to R. P., a Senior Scientist award from the Canadian Institutes of Health Research to L. J. and grants from the Natural Sciences and Engineering Research Council of Canada and the Fonds Québécois de la Recherche sur la Nature et les Technologies to J. H. The authors thank the reviewer and the associate editor for their very helpful comments, which greatly improved the manuscript.

REFERENCES

1. Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. *Psychological Bulletin* 1988; **103**(1):111–120.
2. Berkey CS, Anderson JJ, Hoaglin DC. Multiple outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**:537–557.
3. 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.
4. Nam I, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Statistics in Medicine* 2003; **22**:2309–2333. DOI: 10.1002/sim.1410
5. Dear KBG. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; **50**:989–1002.
6. Arends LR, Voko Z, Stijnen T. Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* 2003; **22**:1335–1353. DOI: 10.1002/sim.1370
7. Van Houwelingen HC, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine* 1993; **12**:2272–2284.
8. 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
9. McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* 1996; **15**:1713–1728.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
11. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). *Annual Review of Public Health* 1996; **17**:1–23.
12. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; **20**:825–840.
13. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
14. 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* 2006; **26**:78–97.
15. Demidenko E. *Mixed Models. Theory and Applications*. Wiley: New Jersey, 2004.
16. Spiegelhalter D, Thomas A, Best N, Gilks W, Lunn D. *BUGS*: *Bayesian Inference using Gibbs Sampling*. MRC Biostatistics Unit, Cambridge, England. www.mrc-bsu.cam.ac.uk/bugs/
17. Ishak K. Meta-analysis of multiple outcomes: fundamentals and applications. *Ph.D. Thesis*, McGill University, 2006.