Meta-analysis without study-specific variance information: Heterogeneity case

Patarawan Sangnawakij,[1] Dankmar Böhning,[2] Sa-Aat Niwitpong,[3] Stephen Adams,[4] Michael Stanton[4] and Heinz Holling[5]

Abstract

The random effects model in meta-analysis is a standard statistical tool often used to analyze the effect sizes of the quantity of interest if there is heterogeneity between studies. In the special case considered here, meta-analytic data contain only the sample means in two treatment arms and the sample sizes, but no sample standard deviation. The statistical comparison between two arms for this case is not possible within the existing meta-analytic inference framework. Therefore, the main objective of this paper is to estimate the overall mean difference and associated variances, the between-study variance and the within-study variance, as specified as the important elements in the random effects model. These estimators are obtained using maximum likelihood estimation. The standard errors of the estimators and a quantification of the degree of heterogeneity are also investigated. A measure of heterogeneity is suggested which adjusts the original suggested measure of Higgins' I^2 for within study sample size. The performance of the proposed estimators is evaluated using simulations. It can be concluded that all estimated means converged to their associated true parameter values, and its standard errors tended to be small if number of the studies involved in the meta-analysis was large. The proposed estimators could be favorably applied in a meta-analysis on comparing two surgeries for asymptomatic congenital lung malformations in young children.

1 Introduction and motivation

Meta-analysis is a statistical methodology used to summarize the results from multiple independent studies on the same topic. Recently, there has been increasing interest in the construction of estimators for quantifying effect sizes of populations in meta-analysis inference, especially in the areas of medical, biostatistical, and epidemiological research, and social science.[1,2] In general, two statistical models are commonly used to integrate the available statistical information: the fixed effect model (or common effect model) and the random effects model.[3–5] These are outlined briefly below. In the fixed effect model, it is assumed that the various contributing studies on the problem of interest share the same true effect size, so that the estimated parameter is a common effect size for all studies. For considering k studies, a general model is defined by

$$Y_i = \mu_i + \varepsilon_i,$$

where Y_i is the effect size estimate and μ_i is the true effect size of the study i. This could be called a *fixed effects* model as there are several fixed effects as one has in analysis of variance. However, in our situation, no replication is available which would not leave any degrees of freedom for estimating the error variance.[6] Hence in this case $\mu_i = \mu$, for $i = 1, 2, \ldots, k$, which leads to the fixed effect or common effect model. The sampling error ε_i is assumed to be normally distributed, $\varepsilon_i \sim N(0, \sigma_i^2)$. In contrast, the random effects model treats the true effect size of the studies involved in the meta-analysis as a random sample from the population effect size distribution and hence allows each study to contribute a different effect. Thus, each effect size μ_i is given by

$$\mu_i = \mu + \xi_i,$$

where the error term ξ_i is again assumed to be normally distributed, $\xi_i \sim N(0, \tau^2)$. If the between-study variance or heterogeneity variance $\tau^2 = 0$, the fixed effect model is a particular case of the random effects

[1] Department of Mathematics and Statistics, Thammasat University, Thailand

[2] Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton, UK

[3] Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Thailand

[4] Department of Paediatric Surgery, Southampton General Hospital, UK

[5] Statistics and Quantitative Methods, Faculty of Psychology and Sports Science, University of Münster, Germany

Corresponding author:
Dankmar Böhning, Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK.
Email: D.A.Bohning@soton.ac.uk

model. The random effects model is often used in the meta-analytic practice if there is variability across studies which frequently occurs. [7] Hence, it is not surprising that the random effects model has gained wide popularity and is considered as the more appropriate approach in most cases. More details of the random effects model in meta-analysis are given in the following section.

In the classical meta-analytic approach of the fixed effect model and the random effects model, [8,9] the overall effect estimate is based on the weighted average of the outcome measures which is computed from the inverse of variance of the effect size. As minimum information, meta-analytic inference underpinning the fixed and random effects models requires the quantity of interest or outcome from each study together with its estimated variance. This reflects the complete case analysis. If only partially complete data are available from some of the studies, imputation often used to handle the missing sample variances. [10,11] However in some practical applications, only information on the quantity of interest and the sample size of the study are reported in the published evidence. No other information on uncertainty quantification is available from which the standard deviation can be obtained, such as the confidence interval or p-value as discussed in Higgins and Green. [12] This is very often the case in settings where proper clinical trials are rarely done for a variety of reasons such as in the situation of congenital lung malformations in young children, often of age less than one year. Here information is often coming from collected case reports on the surgeries under comparison. The question arises how meta-analytic inference can proceed in these cases where standard tools including available statistical packages fail to deliver solutions. Approaches to address this problem have been discussed in Hozo et al., Wan et al., Bland, and Chowdhry et al. [13–16] In Sangnawakij et al., [17] a solution for estimating the mean difference has been suggested assuming homogeneity of effect across studies, in other words, under the fixed effect model. Here we will approach the more realistic but also more complex case of heterogeneity and will show, under mild regularity assumptions, that the more appropriate random effects model can be fitted.

The motivation for this study stems from our wish to consider the differences in process and outcome for young children undergoing either thoracoscopic or open procedures to remove asymptomatic congenital lung malformations (CLMs). CLMs are found in routine antenatal scans of 1 in 2500 live infants. [18] Prior to the 1990s, the open procedure (thoracotomy) was usually performed on infants to excise these lesions. [19] More recently, a new approach called thoracoscopy (or key-hole surgery) has been frequently applied to this operation. [20] The latter surgery involves making several small incisions in the chest through which a fibre-optic camera and operating instruments can be passed. Thus, the procedure makes smaller incisions than the open procedure. However, there have been doubts about the benefits and harms of the new approach in comparison to the open procedure. The comparative outcomes of thoracoscopic and open procedures for CLM should therefore be demonstrated. In a meta-analysis, 38 studies for the period 2004-2015 were provided. The published reports presented only the sample sizes and sample means of the clinical endpoints of interest for open and thoracoscopic surgeries, including

Table 1. The mean length of hospital stay (days) and sample size for thoracoscopic and open operations

Study	Thoracoscopic		Open	
	n	mean	n	mean
1	12	2.00	24	5.00
2	12	3.50	24	4.00
3	14	.	8	.
4	49	3.00	13	3.00
5	39	6.95	28	11.96
6	14	2.95	14	2.60
7	7	6.10	27	8.10
8	5	6.00	4	12.00
9	5	7.00	.	.
10	8	4.60	.	.
11	12	5.80	.	.
12	97	2.40	.	.
13	75	2.40	.	.
14	50	6.00	.	.
15	7	.	.	.
16	12	4.00	.	.
17	20	2.00	.	6.00
18	6	3.50	.	.
19	12	2.00	.	.
20	6	4.50	.	.
21	6	7.00	.	.
22	30	8.00	.	.
23	.	.	105	3.00
24	.	.	16	.
25	.	.	14	7.20
26	.	.	5	.
27	.	.	8	15.00
28	.	.	6	.
29	.	.	6	3.00
30	.	.	3	7.00
31	.	.	9	4.00
32	.	.	35	11.00
33	26	5.30	28	9.60
34	9	.	2	.
35	144	2.80	.	.
36	15	4.20	.	.
37	100	3.00	188	3.10
38	112	3.00	146	4.00

Note. Symbol '.' indicates not reported.

the number of days that drains are left in the chest, length of hospital stay, and weight of the child. As an example, the available information on the mean length of hospital stay is shown in Table 1 together with the associated sample size. No other information is available in the original papers, and the sample

standard deviations cannot be computed given the data reported. Note that several studies are one-armed. For this case study, prior work [17] on meta-analysis without study-specific variance information has only investigated several clinical endpoints of interest including the frequency of operative complications, length of operation, and duration of children stay in hospital under the assumption that the fixed effect model holds. Under the assumption of equal variance across studies the fixed effect model is identifiable and can be fitted easily. However, effect heterogeneity might be more realistic. This leads to the random effects meta-analysis we would like to address here.

In this research, we are interested in constructing the overall mean difference in the random effects model allowing for heterogeneity for the case that only sample means and sample sizes of the two groups to be compared are available in the meta-analytic data. No sample variance information for the studies involved nor confidence intervals for the mean differences are assumed to be available. In the next section, the random effects model for the mean difference is introduced. The estimators for the parameters of the random effects model are developed using a likelihood approach, and then its variance estimates are derived. A measure of heterogeneity which adjusts the original suggested measure of Higgins [21] is discussed in the paper. We investigate the properties of the developed estimators in a simulation study for various settings. The proposed estimators are applied to estimate the overall mean difference of the case study on meta-analysis of key-hole versus open surgery. The last section contains some concluding remarks and discussions.

2 Random effects model

Suppose that we consider k independent studies from two treatment arms with the sample means $(\bar{X}_i)$ and sample sizes (n_i) available in the study i, for $i = 1, 2, \ldots, k$, but no sample standard deviation. The estimated mean difference of the study i, which is the effect size estimate of interest here can be computed by $D_i = \bar{X}_i^T - \bar{X}_i^C$, where $\bar{X}_i^T$ and $\bar{X}_i^C$ are the sample means in the treatment group and control or comparison group, respectively. Here, $\bar{X}_i^T$ and $\bar{X}_i^C$ are independent and identically distributed as $\bar{X}_i^T \sim N(\mu^T, \sigma^2/n_i^T)$ and $\bar{X}_i^C \sim N(\mu^C, \sigma^2/n_i^C)$.

The random effects model for dealing with the estimate of effect size D_i is given by

$$D_i = \mu + \xi_i + \varepsilon_i, \tag{1}$$

where the true mean difference of the study i, defined by $\mu_i = \mu + \xi_i$, is random, μ is the overall mean difference parameter, ξ_i is the random error of μ_i, and ε_i is the sampling error of D_i. The error terms ξ_i and ε_i are independent, and are assumed to be normally distributed with $\xi_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma^2 u_i)$, where τ^2 is the variance due to the study or the between-study variance, $\sigma^2 u_i$ is the

within-study variance, and $u_i = 1/n_i^T + 1/n_i^C$ is a non-random. We point out again that σ^2 is assumed to be identical across all studies. Thus, it follows that the probability density function of D_i is given by

$$f(D_i; \mu, \sigma^2, \tau^2) = \frac{1}{\sqrt{2\pi(\tau^2 + \sigma^2 u_i)}} \exp\left\{-\frac{(D_i - \mu)^2}{2(\tau^2 + \sigma^2 u_i)}\right\}, \tag{2}$$

denoted as $D_i \sim \mathrm{N}(\mu, \tau^2 + \sigma^2 u_i)$. In general, the parameters μ, σ^2, and τ^2 are unknown values, and need to be estimated. Maximum likelihood estimation is used to formulate the estimators in the following section.

3 Identifiability and maximum likelihood estimation

As mentioned in the beginning section, we focus on building the estimators for the parameters in model (1). In order to do inference, we first investigate identifiability of the statistical model. Let D_i be a random variable from a population with the probability density function shown in equation (2). Assume that $\nu_1 = (\mu_1, \sigma_1^2, \tau_1^2)$ and $\nu_2 = (\mu_2, \sigma_2^2, \tau_2^2)$. By the definition of identifiability,[22,23] we have to show that

$$\prod_{i=1}^{k} f(D_i; \nu_1) = \prod_{i=1}^{k} f(D_i; \nu_2), \text{ for all } i, \tag{3}$$

implies $\nu_1 = \nu_2$. Now from equation (3) and the identifiability of the multivariate normal, this implies that $\mu_1 = \mu_2$ and also

$$\tau_1^2 + \sigma_1^2 u_i = \tau_2^2 + \sigma_2^2 u_i, \text{ for all } i,$$

which is only possible if $\sigma_1^2 = \sigma_2^2$ and $\tau_1^2 = \tau_2^2$, assuming that at least two studies have different sample sizes or, more formally, at least one pair i and j, $i \neq j$, exists for which $u_i \neq u_j$. We therefore conclude that the probability model of D_i is identifiable, if not all sample sizes are identical. Parameter estimation for the random effects model is considered as follows.

From equation (2), the log-likelihood function of $\nu = (\mu, \sigma^2, \tau^2)$ is given by

$$l(\nu) = -\frac{1}{2} \sum_{i=1}^{k} \left[\frac{(D_i - \mu)^2}{\tau^2 + \sigma^2 u_i} + \log(\tau^2 + \sigma^2 u_i) + c\right], \tag{4}$$

where $c = \log(2\pi)$. To find the maximum likelihood estimator, we first take the partial derivative of equation (4) with respect to μ, and set it equal to zero. It can be simply written as

$$\frac{\partial l(\nu)}{\partial \mu} = \sum_{i=1}^{k} \frac{D_i - \mu}{\tau^2 + \sigma^2 u_i} = 0.$$

Solving the above equation, the expression for μ is obtained as

$$\mu = \frac{\sum_{i=1}^{k} D_i/(\tau^2 + \sigma^2 u_i)}{\sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)}. \tag{5}$$

Note that equation (5) provides only an implicit estimate for μ as it depends on σ^2 and τ^2.

Similarly, the partial derivatives with respect to σ^2 and τ^2 are taken of equation (4), respectively. Those are given by

$$\frac{\partial l(\nu)}{\partial \sigma^2} = \sum_{i=1}^{k} \frac{(D_i - \mu)^2 u_i}{(\tau^2 + \sigma^2 u_i)^2} - \sum_{i=1}^{k} \frac{u_i}{\tau^2 + \sigma^2 u_i} = 0$$

$$\sum_{i=1}^{k} \frac{(D_i - \mu)^2 u_i}{(\tau^2 + \sigma^2 u_i)^2} = \sum_{i=1}^{k} \frac{(\tau^2 + \sigma^2 u_i) u_i}{(\tau^2 + \sigma^2 u_i)^2}$$

and

$$\frac{\partial l(\nu)}{\partial \tau^2} = \sum_{i=1}^{k} \frac{(D_i - \mu)^2}{(\tau^2 + \sigma^2 u_i)^2} - \sum_{i=1}^{k} \frac{1}{\tau^2 + \sigma^2 u_i} = 0$$

$$\sum_{i=1}^{k} \frac{(D_i - \mu)^2}{(\tau^2 + \sigma^2 u_i)^2} = \sum_{i=1}^{k} \frac{\tau^2 + \sigma^2 u_i}{(\tau^2 + \sigma^2 u_i)^2}.$$

Hence we obtain the expressions for σ^2 and τ^2:

$$\sigma^2 = \frac{\sum_{i=1}^{k}((D_i - \mu)^2 u_i - \tau^2 u_i)/(\tau^2 + \sigma^2 u_i)^2}{\sum_{i=1}^{k} u_i^2/(\tau^2 + \sigma^2 u_i)^2} \tag{6}$$

and

$$\tau^2 = \frac{\sum_{i=1}^{k}((D_i - \mu)^2 - \sigma^2 u_i)/(\tau^2 + \sigma^2 u_i)^2}{\sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)^2}. \tag{7}$$

We can see that equations (5)–(7) have no closed-form solutions, so that the maximum likelihood estimators for μ, σ^2, and τ^2 cannot be given in a simple formula. Thus, a numerical approach based on algorithms[24] is needed to solve the likelihood equations. In fact, we will use fixed point iteration building on equations (5)–(7). More precisely, for given values of μ_s, σ_s^2, and τ_s^2, let

$$\mu_{s+1} = \frac{\sum_{i=1}^{k} D_i/(\tau_s^2 + \sigma_s^2 u_i)}{\sum_{i=1}^{k} 1/(\tau_s^2 + \sigma_s^2 u_i)}, \tag{8}$$

$$\sigma_{s+1}^2 = \frac{\sum_{i=1}^{k}((D_i - \mu_s)^2 u_i - \tau_s^2 u_i)/(\tau_s^2 + \sigma_s^2 u_i)^2}{\sum_{i=1}^{k} u_i^2/(\tau_s^2 + \sigma_s^2 u_i)^2}, \tag{9}$$

and

$$\tau_{s+1}^2 = \frac{\sum_{i=1}^k \left((D_i - \mu_s)^2 - \sigma_s^2 u_i\right)/(\tau_s^2 + \sigma_s^2 u_i)^2}{\sum_{i=1}^k 1/(\tau_s^2 + \sigma_s^2 u_i)^2}. \tag{10}$$

These steps are repeated until convergence, more precisely, until no significant difference ϵ between all three estimated parameters of iteration s and $s+1$. This means that the iterative procedure is run until $|\theta_{s+1} - \theta_s| < \epsilon$, where θ is a generic parameter. A small value for ϵ is used to make sure that we have reached the optimal estimated value for a parameter. Note that if the starting value is chosen by setting $\tau_0^2 = 0$, so that

$$\hat{\mu}_0 = \frac{\sum_{i=1}^k D_i/u_i}{\sum_{i=1}^k 1/u_i} \quad \text{and} \quad \hat{\sigma}_0^2 = \frac{1}{k}\sum_{i=1}^k \frac{(D_i - \hat{\mu}_0)^2}{u_i}$$

become the estimators in the fixed effect meta-analysis. We have no guarantee that the fixed point algorithm presented in equations (8)–(10) converges in theory, but it is easy to show that the algorithm will increase the likelihood at each iteration.

4   Estimation of the overall mean difference

In section 3, the estimators for the difference of mean and related variances in the random effects model are proceeded using a numerical approach. So, we now estimate the variance of D_i by substituting the variance estimates into its parameters. It is given by

$$\widehat{Var}(D_i) = \hat{\tau}^2 + \hat{\sigma}^2 u_i.$$

This estimate is an essential quantity for computing the weight of each study's effect size and concluding the outcomes in meta-analysis inference. Typically, the weight estimate is considered by inverse of variance of the effect size of the study. Thus in the case we are interested here, the mean difference for each study, D_i, is weighted by

$$v_i = \frac{1/(\hat{\tau}^2 + \hat{\sigma}^2 u_i)}{\sum_{i=1}^k 1/(\hat{\tau}^2 + \hat{\sigma}^2 u_i)}.$$

By the weighted average of the mean differences, the overall mean difference is therefore accomplished as

$$\hat{\mu}_{overall} = \frac{\sum_{i=1}^k D_i/(\hat{\tau}^2 + \hat{\sigma}^2 u_i)}{\sum_{i=1}^k 1/(\hat{\tau}^2 + \hat{\sigma}^2 u_i)}. \tag{11}$$

Obviously, $\hat{\mu}_{overall}$ corresponds to the mean estimator shown in equation (5) at convergence. The standard errors of the mean effect and associated variance estimates in the random effects model are discussed in the following section.

5 Standard errors

The Fisher information matrix [25] is a symmetric matrix obtained by taking the negative of the expectation of the second partial derivative matrix or Hessian of the log-likelihood function. In this paper, it can be obtained by taking the derivatives of equation (4) with respect to the parameters μ, σ^2, and τ^2, which is given by

$$I(\nu) = \begin{bmatrix} \sum_{i=1}^{k} \frac{1}{\tau^2 + \sigma^2 u_i} & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^{k} \frac{u_i^2}{(\tau^2 + \sigma^2 u_i)^2} & \frac{1}{2} \sum_{i=1}^{k} \frac{u_i}{(\tau^2 + \sigma^2 u_i)^2} \\ 0 & \frac{1}{2} \sum_{i=1}^{k} \frac{u_i}{(\tau^2 + \sigma^2 u_i)^2} & \frac{1}{2} \sum_{i=1}^{k} \frac{1}{(\tau^2 + \sigma^2 u_i)^2} \end{bmatrix}.$$

The asymptotic variances and covariances of the maximum likelihood estimators can be found by deriving the inverse of $I(\nu)$ as follows:

$$I^{-1}(\nu) = \begin{bmatrix} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}^2) & Cov(\hat{\mu}, \hat{\tau}^2) \\ Cov(\hat{\mu}, \hat{\sigma}^2) & Var(\hat{\sigma}^2) & Cov(\hat{\sigma}^2, \hat{\tau}^2) \\ Cov(\hat{\mu}, \hat{\tau}^2) & Cov(\hat{\sigma}^2, \hat{\tau}^2) & Var(\hat{\tau}^2), \end{bmatrix}, \tag{12}$$

where

$$Var(\hat{\mu}) = \frac{1}{\sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)},$$

$$Var(\hat{\sigma}^2) = \frac{2 \sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)^2}{\sum_{i=1}^{k} u_i^2/(\tau^2 + \sigma^2 u_i)^2 \sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)^2 - (\sum_{i=1}^{k} u_i/(\tau^2 + \sigma^2 u_i)^2)^2},$$

$$Var(\hat{\tau}^2) = \frac{2 \sum_{i=1}^{k} u_i^2/(\tau^2 + \sigma^2 u_i)^2}{\sum_{i=1}^{k} u_i^2/(\tau^2 + \sigma^2 u_i)^2 \sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)^2 - (\sum_{i=1}^{k} u_i/(\tau^2 + \sigma^2 u_i)^2)^2},$$

$$Cov(\hat{\sigma}^2, \hat{\tau}^2) = \frac{-2 \sum_{i=1}^{k} u_i/(\tau^2 + \sigma^2 u_i)^2}{\sum_{i=1}^{k} u_i^2/(\tau^2 + \sigma^2 u_i)^2 \sum_{i=1}^{k} 1/(\tau^2 + \sigma^2 u_i)^2 - (\sum_{i=1}^{k} u_i/(\tau^2 + \sigma^2 u_i)^2)^2},$$

$$Cov(\hat{\mu}, \hat{\tau}^2) = Cov(\hat{\mu}, \hat{\sigma}^2) = 0.$$

We therefore write the asymptotic distribution of the maximum likelihood estimators as $\hat{\nu} \sim N(\nu, I^{-1}(\nu))$, where $\hat{\nu}$ is the vector of the estimators $\hat{\mu}$, $\hat{\sigma}^2$, and $\hat{\tau}^2$. The standard errors of the estimators are obtained by the square roots of the diagonal elements of $I^{-1}(\nu)$.

6   Higgins' I^2

Heterogeneity of effect has become an important issue in meta-analysis. As different studies are likely to experience more variation than we usually have in a single study, it has become almost routine to test for heterogeneity. In the general meta-analytic setting, using D_i for the effect size and σ_i^2 for the associate sampling variance in the study i, the heterogeneity test statistic is defined as

$$Q = \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{\sigma_i^2},$$

where $\bar{D} = (\sum_{i=1}^{k} D_i/\sigma_i^2)/(\sum_{i=1}^{k} 1/\sigma_i^2)$. In our case, σ_i^2 corresponds to $\sigma^2 u_i$. Under normality of the effect size estimator $D_i \sim \mathrm{N}(\mu, \sigma_i^2)$, where μ is the true effect size in the fixed effect model and σ_i^2 is known, Q follows the chi-square distribution with $k-1$ degrees of freedom. If σ_i^2 is estimated, the distribution of this statistic is the approximately chi-squared with $k-1$ degrees of freedom. The details are given in Appendix. The results are similar to those reported in Biggerstaff and Jackson[26], Chen et al.[28], and Hoaglin[27] w.r.t. the distribution of Q. However, in most meta-analytic applications Q is significant, so that the hypothesis of homogeneity of effect has to be rejected. But how much heterogeneity is there? This question inspired Higgins and Thompson[21] to suggest a measure to quantify the amount of heterogeneity present at any meta-analysis. The suggested measure, also called Higgins' I^2, is defined as

$$I^2 = \frac{Q - (k-1)}{Q}.$$

They also present a different form of I^2, namely

$$I^2 = \frac{\tau^2}{\tau^2 + S\bar{E}^2},$$

where $S\bar{E}^2$ is close to an harmonic mean of the study-specific sampling variances σ_i^2, see Borenstein et al.[29] for more details. The benefit of the latter version is that it shows that I^2 is the proportion of total variation due to heterogeneity. Hence this measure is given often in percentage points. It also relates to a variance component interpretation. The total variance is partitioned into between-study and within-study variances, and the measure I^2 can be viewed as the proportion of the between-study variance and total variances of effect size.

However, the measure is often misinterpreted and misused as pointed out by Rücker et al.[30] The point is that the measure is confounded by the study-specific sample sizes involved in the meta-analysis. More precisely, the variance σ_i^2 involves the size n_i of the study, for example, if μ is an overall mean, then $\sigma_i^2 = \sigma^2/n_i$. Hence, a meta-analysis with large sized individual studies will show a larger I^2 than a

meta-analysis with small sized studies although both would have the same underlying τ^2. Similar points have been also made in Borenstein et al.[29,31] Hence it seems appropriate to remove the *sample size effect* from the measure. As

$$Var(D_i) = \tau^2 + \sigma^2 u_i,$$

it appears natural to define

$$J^2 = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{13}$$

which is free of any sample size. In this paper, we will use equation (13) as heterogeneity measure. The interpretation is quite similar to a two component variance model and gives the proportion of variance due to heterogeneity as, marginally, $Var(\bar{X}_i^T - \bar{X}_i^C) = \tau^2 + \sigma^2(1/n_i^T + 1/n_i^C)$. If both arms are balanced with $n_i^T = n_i^C = 2$, the minimum size to allow computation of a variance, we have that $Var(\bar{X}_i^T - \bar{X}_i^C) = \tau^2 + \sigma^2$ which is the basis of our definition for J^2. Note that this form has also the character of a parameter and possesses a straightforward plug-in estimate given by

$$\hat{J}^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2}. \tag{14}$$

We would like to point out that this construction is similar to the coefficient of reliability for a continuous outcome with replications. To be more precise, let Y_{ij} denote the outcome measure for the $j - th$ replication on object i, i might go from 1 to k and j from 1 to n_i. Then, it is standard to assume the conventional random effects model $Y_{ij} = \mu + \delta_i + \epsilon_{ij}$, where δ_i and ϵ_{ij} are independent $N(0, \tau^2)$ and $N(0, \sigma^2)$. It follows that $Var(Y_{ij}) = \tau^2 + \sigma^2$ which is basis for the coefficient of reliability $\tau^2/(\tau^2 + \sigma^2)$. For the mean $\bar{Y}_{i.} = (\sum_j Y_{ij})/n_i$ per object we have that

$$Var(\bar{Y}_{i.}) = \tau^2 + \sigma^2/n_i,$$

just like the meta-analytic situation we consider here.

7 Simulation studies

The performance of the proposed estimators was investigated by simulation using the R statistical language.[32] The setting of the simulation is as follows. The data of ξ_i and D_i were generated as independent and identically distributed with $N(\mu, \tau^2)$ and $N(\xi_i, \sigma^2 u_i)$, respectively. We simulated the weight u_i from uniform$(0.02, 0.20)$ in order to produce reasonable sample sizes of the study. Note that this implies that n_i (assuming balanced studies) will range between 10 and 100. In addition, number of the studies k was varied as 10, 30, 50, and 100. The parameter constellation (μ, σ^2, τ^2) was considered

as (0,12,4), (0,9,4), (0,4,4), and (0,2,6), with the heterogeneity measure J^2 of 0.25, 0.32, 0.50, and 0.75, respectively.

The estimators for μ, σ^2, and τ^2 were computed using the numerical method described in section 3. For the iterative process, we used the following stopping criteria:

$$|\mu_{s+1} - \mu_s| < 0.00001,$$

$$|\sigma^2_{s+1} - \sigma^2_s| < 0.00001,$$

and

$$|\tau^2_{s+1} - \tau^2_s| < 0.00001,$$

where s is the sth iteration. For 10,000 replications of generated dataset, the estimated mean, bias, and standard error for each estimator were calculated in the usual way:

$$\widehat{E(\hat{\theta})} = \sum_{i=1}^{10,000} \hat{\theta}_i / 10,000,$$

$$\text{Bias}(\hat{\theta}) = \widehat{E(\hat{\theta})} - \theta,$$

and

$$\text{SE}(\hat{\theta}) = \sqrt{\sum_{i=1}^{10,000} (\hat{\theta}_i - \widehat{E(\hat{\theta})})^2 / 10,000},$$

where $\hat{\theta}$ is the estimator for a generic parameter θ.

The results in terms of mean, bias, and standard error for all estimators are presented in Table 2. The major results from the simulation studies are summarized as follows:

1. With respect to the mean of the estimators, the overall mean difference $\hat{\mu}$ was fairly close to the true mean effect for all cases in the study (see (a) of Figure 1). The estimated variances $\hat{\sigma}^2$ and $\hat{\tau}^2$ showed some underestimation in general cases. The bias of these two variance estimators depended on k involved whereas the mean estimator seemed to be unaffected by k. This fits to the fact that the mean estimator is unbiased if there is no heterogeneity.
2. To consider the standard error of the estimators, $\hat{\mu}$ provided smaller of the standard error if k was increased, but the standard error did not depend on J^2 (see (b) of Figure 1). $\hat{\sigma}^2$ and $\hat{\tau}^2$ had smaller of standard error for large k. Moreover, $\hat{\sigma}^2$ and $\hat{\tau}^2$ behaved similar across the various degrees of heterogeneity.

Table 2. Mean, bias, and standard error for the maximum likelihood estimators from the simulations for various settings of heterogeneity J^2 and number of the studies k

J^2	k	Mean of			Bias of			Standard error of		
		$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
0.25	10	0.0145	11.5173	3.5961	0.0145	-0.4827	-0.4039	0.8985	50.2959	5.5567
	30	-0.0046	11.7879	3.8589	-0.0046	-0.2121	-0.1411	0.4454	30.9276	3.5410
	50	-0.0023	12.1452	3.8977	-0.0023	0.1452	-0.1023	0.3316	21.6385	2.1109
	100	0.0018	11.9352	3.9492	0.0018	-0.0648	-0.0508	0.2305	14.0151	1.5253
0.32	10	0.0082	7.0515	3.7013	0.0082	-1.9485	-0.2987	0.9003	38.2609	4.9630
	30	-0.0012	8.7462	3.8774	-0.0012	-0.2538	-0.1226	0.4355	28.6170	3.1680
	50	-0.0045	9.2592	3.8944	-0.0045	0.2592	-0.1056	0.3245	20.7052	2.4068
	100	-0.0008	8.9820	3.9544	-0.0008	-0.0180	-0.0456	0.2251	13.8505	1.5707
0.50	10	0.0003	2.5700	3.7879	0.0003	-1.4300	-0.2121	0.8379	39.2310	5.2741
	30	-0.0026	3.5187	3.9180	-0.0026	-0.4813	-0.0820	0.4166	26.0939	3.1109
	50	-0.0053	3.9104	3.9256	-0.0053	-0.0896	-0.0744	0.3066	18.4738	2.1327
	100	0.0020	3.9155	3.9626	0.0020	-0.0845	-0.0374	0.2104	12.3303	1.3812
0.75	10	0.0018	0.6830	5.7004	0.0018	-1.3170	-0.2996	0.9702	77.3782	9.3725
	30	0.0055	2.1001	5.8055	0.0055	0.1001	-0.1945	0.4940	39.2915	4.3828
	50	-0.0025	1.8716	5.9245	-0.0025	-0.1284	-0.0755	0.3624	25.8234	3.3537
	100	0.0022	1.9565	5.9263	0.0022	-0.0435	-0.0737	0.2508	16.4556	1.9859

Table 3. Standard error for the maximum likelihood estimators computed from the inverse of Hessian matrix

J^2	k	Standard error of		
		$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
0.25	10	0.8290	42.1864	4.7553
	30	0.3888	28.7642	3.0663
	50	0.3104	23.4202	2.5542
	100	0.2242	13.5165	1.4548
0.32	10	0.9051	42.3244	5.2547
	30	0.3748	25.3836	2.9801
	50	0.3033	17.0084	1.9818
	100	0.2199	14.4694	1.7027
0.50	10	0.7859	40.2160	4.5538
	30	0.3761	29.5240	2.7175
	50	0.2888	15.8014	2.0168
	100	0.2059	12.5895	1.3374
0.75	10	0.8654	69.6343	7.4421
	30	0.4163	31.7833	3.5287
	50	0.3405	22.7031	2.9165
	100	0.2454	17.0483	2.1262

3. From Table 3, the standard errors of $\hat{\mu}$, $\hat{\sigma}^2$, and $\hat{\tau}^2$ evaluated from equation (12) decreased as k increased. This conclusion is similar to that of the standard error based on the basis of simulation as noted above.

Table 4. Mean, bias, and standard error from the simulations for various setting number of the studies k and $\sigma_i^2 \sim N(\sigma^2 = 12, \varphi)$, where $\mu = 0$ and $\tau^2 = 4$

φ	k	Mean of			Bias of			Standard error of		
		$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$
1	10	-0.0242	9.2004	3.5459	-0.0242	-2.7996	-0.4541	0.9143	64.3924	8.2825
	30	-0.0062	11.4807	3.8342	-0.0062	-0.5193	-0.1658	0.4612	31.6823	3.4389
	50	-0.0011	11.7146	3.9536	-0.0011	-0.2854	-0.0464	0.3307	22.1839	2.1868
	100	-0.0022	12.1487	3.9475	-0.0022	0.1487	-0.0525	0.2276	15.5601	1.6524
3	10	0.0117	10.3958	3.8399	0.0117	-1.6042	-0.1601	0.9342	72.0829	9.9928
	30	-0.0013	11.5679	3.8860	-0.0013	-0.4321	-0.1140	0.4518	30.3783	3.1606
	50	-0.0026	12.1083	3.9010	-0.0026	0.1083	-0.0990	0.3344	20.6507	2.2282
	100	-0.0025	12.0647	3.9390	-0.0025	0.0647	-0.0610	0.2327	14.0622	1.6190
5	10	0.0233	8.1661	3.1308	0.0233	-3.8339	-0.8692	0.9451	74.6134	9.6298
	30	-0.0050	11.6849	3.9353	-0.0050	-0.3151	-0.0647	0.4392	30.6386	3.2872
	50	0.0036	11.8223	3.9325	0.0036	-0.1777	0.0675	0.3434	22.0911	2.3371
	100	-0.0032	12.1294	3.9556	-0.0032	0.1293	-0.0444	0.2312	14.9574	1.5481

In summary, the estimators $\hat{\mu}$, $\hat{\sigma}^2$, and $\hat{\tau}^2$ appear to behave well. To be more precise, $\hat{\mu}$ has small bias for all k and all heterogeneity cases. In fact, the bias is zero within 2 decimal places (except the one case where $k = 10$ and $J^2 = 0.25$) which is conventionally considered as zero-bias in simulation work. This is an important finding as μ is the parameter of interest in applications. Bias for $\hat{\tau}^2$ and $\hat{\sigma}^2$ is not ignorable for $k = 10$ but settles down for $k = 30$ and larger number of studies. The standard errors are as they are and there are no benchmark or competitor estimators to compare too. However, it is important to note that, as Table 2 and Table 3 show, the standard errors can validly be estimated via the Hessian matrix, at least for values of k larger than 10, so that confidence intervals and other inferential procedures requiring standard errors can be trustworthy applied.

The question arises how robust are the suggested estimators if variance homogeneity across trials is violated. To mimic at least one realistic scenario, we generated ξ_i from $N(\mu, \tau^2)$ and D_i from $N(\xi_i, \sigma_i^2 u_i)$. σ_i^2 is the population squared standard deviation which was generated from $N(\sigma^2, \varphi)$, where the variance φ was varied as 1, 3, and 5. Marginally we still have a common variance, but study-specific variances differ. Estimators were computed under the common study-specific variance assumption. In Table 4, the mean, bias, and standard error are presented. They were compared with the related values shown in the first row of Table 2. If we compare the estimator for σ^2 with the one in Table 2, we see the occurrence of some bias which diminishes with increasing number of studies. There is also a general increase in the standard error for $\hat{\sigma}^2$. However, the inference for the parameter of interest, μ, remains largely unchanged.
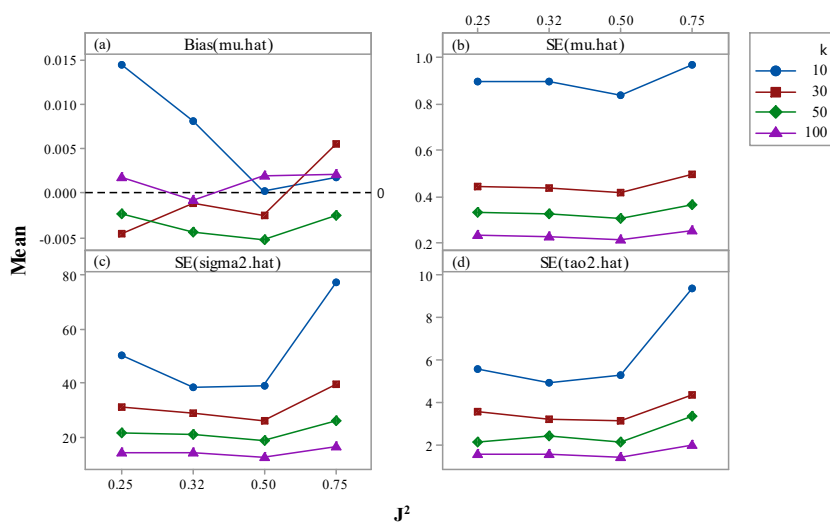
Figure 1. Simulation-based graphs of bias for $\hat{\mu}$ (a) and standard errors for $\hat{\mu}$ (b), $\hat{\sigma}^2$ (c), and $\hat{\tau}^2$ (d) against number of the studies k

8 Case study

To illustrate the estimators presented in this paper, we continue the discussion of the meta-analytic data on thoracoscopic and open operations, the two surgeries for treating congenital lung malformation motivated in section 1. As this data set contains a number of studies or reports with information on one arm, however we consider only those with both arms information. The continuous variables considered in the analysis are consisted of length of hospital stay (days), number of chest tube days (days), and weight of child (kgs).

The estimators are computed for μ, σ^2, and τ^2 based on iteration procedure proposed in the previous section. The numerical results of the estimation procedure using a meta-analysis for the random effects model are presented in Table 5. It can be seen that the estimated values of $\hat{J}^2$ for two variables, length of hospital stay and number of chest tube days, show low heterogeneity with values of 15% and 18%, respectively. A very small value of $\hat{J}^2$, very close to zero, is found for weight of child. So, there is some evidence of heterogeneity for length of stay and number of chest tube days, none for child's weight. The 95% confidence intervals for the mean difference and the variances computed from equation (15) are shown in Table 6. It can be concluded that length of the children stay in hospital and number of chest
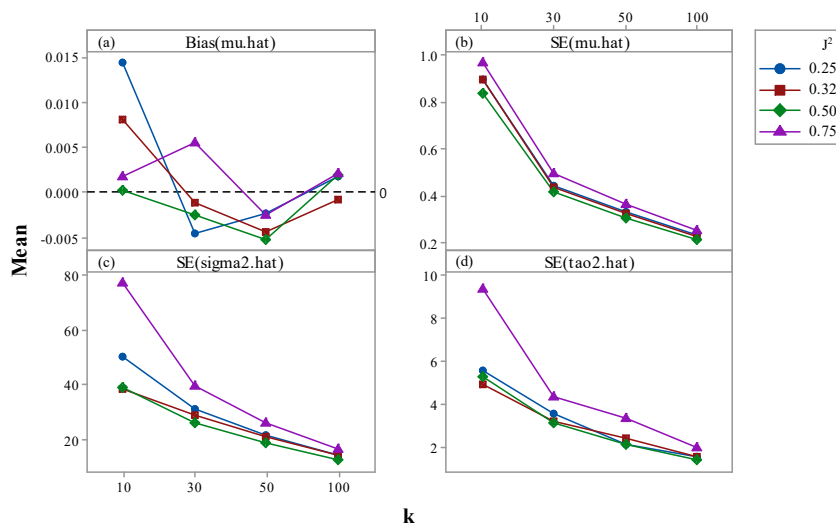
Figure 2. Simulation-based graphs of bias for $\hat{\mu}$ (a) and standard errors for $\hat{\mu}$ (b), $\hat{\sigma}^2$ (c), and $\hat{\tau}^2$ (d) against the heterogeneity measure J^2

Table 5. The results for the meta-analysis comparing thoracoscopic and open operations

Variable	k	Estimator			
		$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\%\hat{J}^2$
length of hospital stay	11	-2.10	12.77	2.83	18.00
number of chest tube days	9	-1.17	6.42	1.15	15.00
weight of child	7	0.41	15.03	0.07	0.46

tube days are better for thoracoscopic resection. Length of stay is 2.10 days shorter. Number of chest tube days is 1.17 days shorter. Weight of child is not different between two surgeries.

The $(1 - \alpha)100\%$ confidence interval for the parameter of interest is given by

$$\hat{\theta} \pm Z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta})}, \tag{15}$$

where $\hat{\theta}$ is the estimator for a generic parameter θ and $Z_{\alpha/2}$ is the $(\alpha/2)100$th percentile of the standard normal distribution.

Table 6. The 95% confidence intervals for the meta-analysis comparing thoracoscopic and open operations

Variable	95% confidence intervals for		
	μ	σ^2	τ^2
length of hospital stay	(-3.32,-0.88)	(0.00,57.56)	(0.00,8.00)
number of chest tube days	(-2.07,-0.27)	(0.00,29.21)	(0.00,3.84)
weight of child	(-0.22,1.04)	(0.00,38.61)	(0.00,0.83)

9   Discussion and conclusion

In this paper, we develop the overall mean difference (effect size) and variance estimators for the random effects meta-analysis using the maximum likelihood approach in a setting where only mean differences of the quantity of interest in the two treatment arms and associate sample sizes are available, but no other information. In particular, estimated standard errors or confidence intervals cannot be computed given the data reported. The derived estimators are built under the assumption of equal variances in the study arms and across studies. This is a crucial assumption and might not be appropriate, in particular if trials are included in the meta-analysis with variation of scale of measurement. In cases, where the same scale is used, one such as the case study here, it seems a plausible assumption. In addition, we allow heterogeneity across studies as this seems most appropriate for meta-analytic applications. In the approach we address, the estimators cannot be given in a simple formula, and the same amount of information provided by sample mean differences is used to estimate all the three parameters: true effect size, within-study and between-study variances.

To avoid misunderstanding we just emphasize that we do not propose to ignore study-specific variance information. In fact, it is shown in previous work that using partially available study-specific variance information can crucially improve efficiency.[10,11,16,17] In particular we suggest, of course, to report study specific variability whenever possible. However, just for settings, where this information is not available we propose the solutions offered, here so that the available information can be used in a meaningful way. We believe that the problem of entirely missing variance information occurs more frequently than yet documented. We mention here a meta-analysis of changes in renal function after living kidney donation[33]. It might be also the case that often meta-analyses are not undertaken because of missing variance information.

We point out again that, in a typical setting with study-specific variances available, the information should be used to estimate the two parameters of interest (mean and heterogeneity variance). Here, the classical method is not applicable since no variance information is available for all studies. The use of the available information is fostered in the approach proposed here to be able to make limited inference for the data available. It seems that the proposed methodology provided reasonable accurate estimators

although the price is to pay is a loss of efficiency. In summary, it appears that the proposed methodology performs reasonably well.

Since no simple, closed-form expressions are available for the proposed estimators, a numerical method based on algorithm is applied to obtain estimated values. This method is very similar to an EM–type–of–iteration in the sense that it is nondecreasing at every iteration. Simulation was used to evaluate the performance of the estimators. We used the bias and standard error as performance indicators. The simulation results showed that the values of bias and standard error for all estimators did not depend on the measure J^2 (see Figure 2). Moreover, if the number of studies was increased, the standard errors of all estimators decreased and tended to be small. These results are as expected for maximum likelihood estimators as sample sizes increase.[22] In conclusion, the random effects model for the mean difference in an application for a meta-analysis without estimated study-specific variance information can be well estimated using the approach outlined here. In addition, R-code is available in the supplementary material to readily implement all estimators for any other meta-analysis.

References

1. Brockwell SE and Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; **20**: 825–840.

2. Huang EP, Wang XF, Choudhury KR, et al. Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology. *Stat Methods Med Res* 2015; **24**: 141–174.

3. Hedges LV and Vevea JL. Fixed and random effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486–504.

4. Schulze R, Holling H and Böhning D. *Meta-analysis: new developments and applications in medical and social sciences*. Göttingen: Hogrefe & Huber, 2003.

5. Borenstein M, Hedges LV, Higgins JPT, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Syn Meth* 2010; **1**: 97–111.

6. Higgins JPT, Thompson SG and Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A* 2009; **172**: 137–159.

7. Hunter JE and Schmidt FL. Fixed effects vs random effects meta-analysis model: implication for cumulative research knowledge. *Int J Sel Assess* 2000; **8**: 275–292.

8. Sutton AJ, Abrams KR, Jones DR, et al. *Methods for meta-analysis in medical research*. New York: Wiley, 2000.

9. Böhning D, Kuhnert R and Rattanasiri S. *Meta-analysis of binary data using profile likelihood*. London: Chapman & Hall, 2008.

10. Furukawa TA, Barbui C, Cipriani A, et al. Imputing missing standard deviations in meta analyses can provide accurate results. *J Clin Epidemiol* 2006; **59**: 7–10.

11. Idris NRN and Robertson C. The effects of imputing the missing standard deviations on the standard error of meta analysis estimates. *Commun Stat Simul Comput* 2009; **38**: 513–526.

12. Higgins J and Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley and Sons, 2008.

13. Hozo SP, Djulbegovic B and Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Meth* 2005; **5**: 1–10.

14. Wan X, Wang W, Liu J, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Meth* 2014; **14**: 1–13.

15. Bland M. Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum. *Int J Stat Med Res* 2015; **4**: 57–64.

16. Chowdhry AK, Dworkinb RH and McDermott MP. Meta-analysis with missing study-level sample variance data. *Stat Med* 2016; **35**: 3021–3032.

17. Sangnawakij P, Böhning D, Adams S, et al. Statistical methodology for estimating the mean difference in a meta-analysis without study-specific variance information. *Stat Med* 2017; **36**: 1395–1413.

18. Stocker LJ, Wellesley DG, Stanton MP, et al. The increasing incidence of fetal echogenic congenital lung malformations: an observational study. *Prenat Diagn* 2015; **35**: 148–153.

19. Stanton M, Njere I, Adeajayi N, et al. Systematic review and meta-analysis of the postnatal management of congenital cystic lung lesions. *J Pediatr Surg* 2009; **44**: 1027–1033.

20. Albanese CT, Sydorak RM, Tsao K, et al. Thoracoscopic lobectomy for prenatally diagnosed lung lesions. *J Pediatr Surg* 2003; **38**: 553–555.

21. Higgins JP and Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–1558.

22. Casella G and Berger RL. *Statistical inference*. California: Duxbury Press, 2002.

23. Zwiernik P. *Semialgebraic statistics and latent tree models*. London: Chapman & Hall, 2016.

24. Eliason SR. *Maximum likelihood estimation: logic and practice*. California: Sage Publications, 1993.

25. Rohatgi VK. *An introduction to probability theory and mathematical statistics*. New York: Wiley and Sons, 2001.

26. Biggerstaff BJ and Jackson D. The exact distribution of Cochran's heterogeneity statistic in one-way random effects meta-analysis. *Stat Med* 2008; **27**: 6093–6110.

27. Hoaglin DC. Misunderstandings about Q and Cochran's Q test in meta-analysis. *Stat Med* 2016; **35**: 485–495.

28. Chen Z, Ng HKT, Nadarajah S. A note on Cochran test for homogeneity in one-way ANOVA and meta-analysis. *Statistical Papers* 2014; **55**: 301–310.

29. Borenstein M, Hedges LV, Higgins JPT, et al. *Introduction to meta-analysis*. Chichester: Wiley & Sons, 2009.

30. Rücker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Meth* 2008; **8**: 1-9.

31. Borenstein M, Higgins JP, Hedges LV, et al. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Res Synth Methods* 2017. doi: 10.1002/jrsm.1230.

32. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/ (accessed 9 July 2016).

33. Philbrook HT, Barrowmanb N and Garga AX. Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *J Clin Epidemiol* 2007; **60**: 228–240.

Appendix

In this appendix, the distribution of the heterogeneity test statistic Q in section 6 is derived. The test statistic is defined as

$$Q = \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{\sigma_i^2} = \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{\sigma^2 u_i},$$

where $D_i \sim \mathrm{N}(\mu, \sigma_i^2 = \sigma^2 u_i)$ is the mean difference, $\bar{D} = (\sum_{i=1}^{k} D_i/\sigma_i^2)/(\sum_{i=1}^{k} 1/\sigma_i^2)$, and $u_i = 1/n_i^T + 1/n_i^C$. Here, we have

$$\sum_{i=1}^{k} \frac{(D_i - \mu)^2}{\sigma^2 u_i} = \frac{1}{\sigma^2} \sum_{i=1}^{k} \frac{(D_i - \bar{D} + \bar{D} - \mu)^2}{u_i}$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{u_i} + \frac{1}{\sigma^2} \sum_{i=1}^{k} \frac{(\bar{D} - \mu)^2}{u_i}.$$

This can be written as

$$\frac{1}{\sigma^2} \sum_{i=1}^{k} \left(\frac{D_i - \mu}{\sqrt{u_i}} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{k} \left(\frac{D_i - \bar{D}}{\sqrt{u_i}} \right)^2 + \frac{1}{\sigma^2} \left[(\bar{D} - \mu) \sqrt{\sum_{i=1}^{k} 1/u_i} \right]^2$$

$$= \left(\frac{k-1}{\sigma^2} \right) \frac{1}{k-1} \sum_{i=1}^{k} \left(\frac{D_i - \bar{D}}{\sqrt{u_i}} \right)^2 + \frac{1}{\sigma^2} \left[(\bar{D} - \mu) \sqrt{\sum_{i=1}^{k} 1/u_i} \right]^2$$

$$= \left(\frac{k-1}{\sigma^2} \right) S^2 + \frac{1}{\sigma^2} \left[(\bar{D} - \mu) \sqrt{\sum_{i=1}^{k} 1/u_i} \right]^2$$

$$U = U_1 + U_2.$$

As in standard inferential statistics, S^2 and $\bar{D}$ are independent, hence the terms U_1 and U_2 are. This implies for the respective moment generating functions that $M_{U_1+U_2}(t) = M_{U_1}(t)M_{U_2}(t)$.

Since $E(\bar{D}) = \mu$ and $\text{Var}(\bar{D}) = 1/\sum_{i=1}^{k} 1/\sigma^2 u_i$, U and U_2 follow the chi-square distributions with k and one degrees of freedom, respectively. Using moment generating functions, we obtain that $M_{U_1}(t) = 1/(1-2t)^{(k-1)/2}$. Therefore, under $\tau^2 = 0$ and σ^2 known, $Q = U_1$ has an exact chi-square distribution with $k-1$ degrees of freedom. Note that the result from the current study are the same as that reported in Biggerstaff and Jackson[26].

Suppose that σ^2 is an unknown parameter and it is estimated by the variance estimate $\hat{\sigma}^2$ as presented in equation (9). In this case, the heterogeneity test statistic is given as

$$\hat{Q} = \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{\hat{\sigma}^2 u_i} = \frac{\sigma^2}{\hat{\sigma}^2} \sum_{i=1}^{k} \frac{(D_i - \bar{D})^2}{\sigma^2 u_i}.$$

As the maximum likelihood estimator $\hat{\sigma}^2$ is consistent, $\hat{\sigma}^2$ converges in probably to σ^2. Hence, $\hat{Q}$ converges in distribution to Q by Slutsky's theorem. In consequence, the limiting distribution of $\hat{Q}$ is chi-square with $k-1$ degrees of freedom.