STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:1964–1981
Published online 6 September 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2688

A comparison of heterogeneity variance estimators in combining results of studies

Kurex Sidik[1,‡] and Jeffrey N. Jonkman[2,*,†]

[1]*Pfizer Global Research and Development, St. Louis Lab., St. Louis, MO, U.S.A.*
[2]*Mississippi State University, Mississippi State, MS, U.S.A.*

SUMMARY

For random effects meta-analysis, seven different estimators of the heterogeneity variance are compared and assessed using a simulation study. The seven estimators are the variance component type estimator (VC), the method of moments estimator (MM), the maximum likelihood estimator (ML), the restricted maximum likelihood estimator (REML), the empirical Bayes estimator (EB), the model error variance type estimator (MV), and a variation of the MV estimator (MVvc). The performance of the estimators is compared in terms of both bias and mean squared error, using Monte Carlo simulation. The results show that the REML and especially the ML and MM estimators are not accurate, having large biases unless the true heterogeneity variance is small. The VC estimator tends to overestimate the heterogeneity variance in general, but is quite accurate when the number of studies is large. The MV estimator is not a good estimator when the heterogeneity variance is small to moderate, but it is reasonably accurate when the heterogeneity variance is large. The MVvc estimator is an improved estimator compared to the MV estimator, especially for small to moderate values of the heterogeneity variance. The two estimators MVvc and EB are found to be the most accurate in general, particularly when the heterogeneity variance is moderate to large. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: across-study variance; effect size estimation; meta-analysis; random effects model; simulation

1. INTRODUCTION

In meta-analysis, i.e. combining the results of a set of related and independent studies for assessing a treatment effect, heterogeneity in the effect sizes from the studies is common and can be substantial because of variation in study characteristics. Factors contributing to heterogeneity may include

*Correspondence to: Jeffrey N. Jonkman, Department of Mathematics and Statistics, P.O. Box MA, Mississippi State, MS 39762, U.S.A.
†E-mail: jonkman@math.msstate.edu
‡E-mail: Kurex.Sidik@pfizer.com

Received 13 May 2005
Accepted 6 July 2006

study designs, participating subjects, execution of treatment, intervention, outcome measures, and so on [1–5]. It is essential to quantify such heterogeneity in making meta-analytical inference about a treatment effect. In general, heterogeneity in effect sizes is quantified by considering the between-study variance of the effect measure estimates in a random effects model for combining the study results [1, 2, 6]. Note that the between-study variance is sometimes called the heterogeneity variance [7–11] or the random effects variance. Several methods of estimating the heterogeneity variance have been introduced. However, so far we are not aware of any literature involving comprehensive Monte Carlo comparisons of the heterogeneity variance estimators for meta-analysis. Although there are a few published papers which have made comparisons of heterogeneity variance estimators, they are either oriented toward a specific application or otherwise limited in scope, and thus do not provide a general comparison of the available estimators (e.g. References [2, 7, 9, 11–13]). In this paper, we conduct a Monte Carlo study to compare seven estimators of the heterogeneity variance.

To estimate the heterogeneity variance in a random effects meta-analysis model, Hedges [1], by analogy to the estimation of variance components in random effects analysis of variance, proposed a method of moments estimator (see also Reference [14, pp. 193–194]). This variance component (VC) type estimator is simple to compute and does not require an iterative numerical solution. It is also an unbiased estimator if the within-study variances of the effect estimates are assumed to be known. DerSimonian and Laird [2] introduced another simple estimator using the expectation of Cochran's [15, 16] statistic. Their estimator is also a method of moments (MM) estimator, and is unbiased if the weights (the within-study variances) are known. The DerSimonian and Laird [2] estimator is the most commonly used estimator in random effects meta-analysis because it is simple and non-iterative [12, 17], although it is fair to point out that the VC estimator is also simple and non-iterative. Recently, Sidik and Jonkman [11] proposed another simple and non-iterative estimator by reparameterizing the total variance of an effect statistic. The estimator is based on unbiased estimation of the model error variance in a linear model, and hence could be called a model error variance (MV) type of estimator. Besides the three simple non-iterative estimators VC, MM, and MV, several other heterogeneity variance estimators have been used for the random effects model. These estimators are less simple to compute, as they require an iterative solution in order to obtain an estimate. The maximum likelihood (ML) estimator of the heterogeneity variance, assuming a normal distribution for the effect estimates, has been used in meta-analysis (see References [2, 4, 18] etc.). The ML estimator requires some computational programming to obtain an iterative solution. Under a normal hierarchical model, Raudenbush and Bryk [6] suggested the restricted maximum likelihood (REML) estimator for the heterogeneity variance. The REML estimator is also not simple and requires an iterative method. Morris [19], in an empirical Bayes framework, introduced a method for estimating the between-study variance, which has been referred to as the empirical Bayes (EB) estimator in meta-analytical applications (see References [12, 20, 21]). The EB estimator also requires iterative computation.

To assess the empirical properties of the six estimators VC, MM, MV, ML, REML, EB, as well as a variation of the MV estimator based on an initial estimate from VC, which we denote as MVvc, we conducted Monte Carlo simulation to compare the empirical bias and mean squared error (MSE) of the estimators under a random effects model. All seven estimators considered in the comparisons are general estimators, in that they can be applied with any effect measure of interest. Estimators that are restricted in application to a specific effect measure, such as the estimators proposed for a standardized mean difference by Malzahn *et al.* [7] and (the special form of the general estimator VC) by Hedges and Olkin [14, pp. 193–194], are not considered here.

DOI: 10.1002/sim

2. THE HETEROGENEITY VARIANCE ESTIMATORS

Suppose that $Y_1, Y_2, \ldots, Y_k$ are the effect measure variables from k independent studies. Consider the random effects model

$$Y_i = \theta_i + \varepsilon_i, \quad \theta_i = \theta + \delta_i \quad \text{for } i = 1, \ldots, k \tag{1}$$

where we assume that $\varepsilon_i \sim \mathrm{N}(0, \sigma_i^2)$ and $\delta_i \sim N(0, \tau^2)$. Note that θ_i is the study-specific effect size for the ith study, σ_i^2 is the within-study variance in the ith study, θ is the overall treatment effect (viewed as the common mean of the population of study effects with θ_i for $i = 1, \ldots, k$ being samples from the population), and finally τ^2 is the random effects variance or the heterogeneity variance. Under the model, the total variance of an effect measure estimate Y_i has the two additive components σ_i^2 and τ^2. In practice, the study-specific variances σ_i^2 for $i = 1, \ldots, k$ are estimated using the sample data within studies, and are usually assumed known in meta-analytical inference. Thus, in this paper we will also use estimates of σ_i^2 for $i = 1, \ldots, k$ as in practice, and we denote the estimate for the ith study by $\hat{\sigma}_i^2$. Strictly speaking, such a practice is problematic because of uncertainty in the estimated $\hat{\sigma}_i^2$ (see References [8, 17, 22]). Nevertheless, it is common to treat the estimated $\hat{\sigma}_i^2$ as fixed and known values when combining summary statistics from several studies. Therefore, the main focus of this paper will be estimation of the heterogeneity variance τ^2, with $\hat{\sigma}_i^2$ being known.

2.1. The variance component type estimator (VC)

Using an analogue to the method of variance components estimation in a random effects analysis of variance, Hedges [1] introduced a simple estimator for the heterogeneity variance (see also Reference [14, pp. 193–194]). Specifically, set the quadratic statistic

$$S_Y^2 = \frac{1}{k-1} \sum_{i=1}^{k} (Y_i - \bar{Y})^2$$

equal to its expected value and solved for τ^2, while replacing σ_i^2 with its study-specific estimate $\hat{\sigma}_i^2$ as discussed above. Thus the VC estimator is a method of moments estimator, and is given by

$$\hat{\tau}_{\mathrm{VC}}^2 = \frac{1}{k-1} \sum_{i=1}^{k} (Y_i - \bar{Y})^2 - \frac{1}{k} \sum_{i=1}^{k} \hat{\sigma}_i^2 \tag{2}$$

We use VC to denote this estimator in order to distinguish it from the method of moments estimator proposed by DerSimonian and Laird [2]. The VC estimator may lead to a negative estimate of τ^2, and hence it is used by enforcing non-negativity in practice, i.e. $\max\{0, \hat{\tau}_{\mathrm{VC}}^2\}$. The estimator in (2) is simple to compute, and it has been found to estimate τ^2 quite accurately in some cases [4, 11]. Note that the estimator $\hat{\tau}_{\mathrm{VC}}^2$ may be viewed as the general form of the Hedges and Olkin [14] estimator, considering that a special form of the estimator has been developed for the standardized mean difference as an effects measure [1, 7, 14, 23]. The estimator has not been widely used in the random effects meta-analysis literature, except for the special form of the estimator for the standardized mean difference.

Note that other estimators of τ^2 could be developed using different quadratic functions of the Y_i instead of S_Y^2, as noted by Hedges [1]. An example of such an estimator using a similar method but

DOI: 10.1002/sim

a different quadratic function of Y_i is the method of moments estimator proposed by DerSimonian and Laird [2].

2.2. The method of moments estimator (MM)

In a study of the random effects approach for combining results from randomized clinical trials, DerSimonian and Laird [2] proposed a simple non-iterative procedure to estimate the heterogeneity variance. They used Cochran's [16] statistic

$$Q_w^2 = \sum_{i=1}^{k} w_i (Y_i - \bar{Y}_w)^2$$

where $w_i = 1/\sigma_i^2$ and $\bar{Y}_w = \sum_{i=1}^k w_i Y_i / \sum_{i=1}^k w_i$. Specifically, by equating Q_w^2 with its expected value and solving for τ^2 they found a method of moments (MM) estimator for τ^2. Using estimated values $\hat{\sigma}_i^2$ for $i = 1, \ldots, k$, in practice the MM estimator is given by

$$\hat{\tau}^2_{\mathrm{MM}} = \frac{Q_{\tilde{w}}^2 - (k-1)}{\sum_{i=1}^k \tilde{w}_i - \sum_{i=1}^k \tilde{w}_i^2 / \sum_{i=1}^k \tilde{w}_i} \tag{3}$$

Here $\tilde{w}_i = 1/\hat{\sigma}_i^2$, and $Q_{\tilde{w}}^2$ is obtained by replacing w_i by $\tilde{w}_i$ in Q_w^2 (see also Reference [8]). The estimator $\hat{\tau}^2_{\mathrm{MM}}$ may also yield a negative estimate for the heterogeneity variance, and hence the truncated version $\max\{0, \hat{\tau}^2_{\mathrm{MM}}\}$ is usually used. Note that the estimator given in (3) is a biased estimator of τ^2 because estimated weights $\tilde{w}_i$ are used. In theory, method of moments estimators are unbiased, but here that would require use of the true weights w_i. If the within-study variances are equal for all i, i.e. $\hat{\sigma}_i^2 = \hat{\sigma}^2$, then $\hat{\tau}^2_{\mathrm{MM}}$ is equal to $\hat{\tau}^2_{\mathrm{VC}}$, and is the MINQUE estimator [2, 24].

2.3. The model error variance estimators (MV)

Under the random effects model (1), Sidik and Jonkman [11] proposed a simple non-iterative estimator for τ^2 by reparameterizing the total variance of Y_i. Specifically, consider the reparameterization $\mathrm{Var}(Y_i) = \tau^2(r_i + 1)$, where $r_i = \sigma_i^2/\tau^2$ (assuming $\tau^2 \neq 0$). To estimate τ^2 in practice, *a priori* values or estimates $\hat{r}_i$ must be used in place of the ratios r_i for $i = 1, \ldots, k$. Hence, an estimator of τ^2 is obtained by the usual method of estimating the model error variance in a weighted linear model, as follows:

$$\hat{\tau}^2_{\mathrm{MV}} = \frac{1}{k-1} \sum_{i=1}^{k} \hat{v}_i^{-1} (Y_i - \bar{Y}_{\hat{v}})^2 \tag{4}$$

where $\hat{v}_i = \hat{r}_i + 1$ and $\bar{Y}_{\hat{v}} = \sum_{i=1}^k \hat{v}_i^{-1} Y_i / \sum_{i=1}^k \hat{v}_i^{-1}$. To obtain an estimate of τ^2, Sidik and Jonkman [11] used the crude ratio estimates $\hat{r}_i = \hat{\sigma}_i^2 / (\sum_{i=1}^k (y_i - \bar{y})^2/k)$, where y_i is the observed sample value of Y_i. Unlike the estimators VC and MM, the MV estimator always yields a positive estimate for τ^2, and hence truncation of the estimator at zero is not necessary. Note that if the exact r_i values were known, $\hat{\tau}^2_{\mathrm{MV}}$ would also be an unbiased estimator of τ^2.

An improvement on the estimator given in (4) may be possible by using better *a priori* values or estimates for the ratios, instead of the crude ones provided above [11]. For instance, instead of the crude ratio estimates used in $\hat{\tau}^2_{\mathrm{MV}}$, we may consider using estimated ratio values based on the VC estimator; i.e. using $\hat{r}_i = \hat{\sigma}_i^2 / \hat{\tau}^2_{\mathrm{VC}}$ in (4). Note that in this approach, a small value must be

DOI: 10.1002/sim

added to $\hat{\tau}^2_{\text{VC}}$ in order to compute $\hat{r}_i$ whenever $\hat{\tau}^2_{\text{VC}}=0$. In this study, we use $\hat{r}_i=\hat{\sigma}^2_i/0.01$ when $\hat{\tau}^2_{\text{VC}}=0$. To distinguish this estimator from the estimator MV, we denote the estimator using $\hat{\tau}^2_{\text{VC}}$ to construct $\hat{r}_i$ as the MVvc estimator.

2.4. The maximum likelihood estimator (ML)

Maximum likelihood estimation of τ^2 has also been used under the random effects model [2, 4, 12, 18]. The ML estimator is not as simple as the preceding estimators, because it requires an iterative solution. Treating the within-study variances $\hat{\sigma}^2_i$ as known (as noted earlier), we have the marginal distribution $Y_i \sim \mathrm{N}(\theta, \hat{\sigma}^2_i+\tau^2)$, leading to the log likelihood function

$$L(\theta,\tau^2)=-\frac{k}{2}\log(2\pi)-\frac{1}{2}\sum_{i=1}^{k}\log(\hat{\sigma}^2_i+\tau^2)-\frac{1}{2}\sum_{i=1}^{k}\frac{(y_i-\theta)^2}{\hat{\sigma}^2_i+\tau^2}$$

Hence, for a specified convergence criterion the ML estimator $\hat{\tau}^2_{\text{ML}}$ for τ^2 can be found as an iterative solution to the following equation:

$$\hat{\tau}^2_{\text{ML}}=\frac{\sum_{i=1}^{k}\hat{w}^2_i\{(y_i-\hat{\theta}_{\hat{w}})^2-\hat{\sigma}^2_i\}}{\sum_{i=1}^{k}\hat{w}^2_i} \tag{5}$$

where $\hat{\theta}_{\hat{w}}=\sum_{i=1}^{k}\hat{w}_i y_i/\sum_{i=1}^{k}\hat{w}_i$ and $\hat{w}_i=1/(\hat{\sigma}^2_i+\hat{\tau}^2_{\text{ML}})$, with an initial estimate of τ^2 [4, 12]. At each iteration, the estimate of τ^2 needs to be checked for a possible negative value, and non-negativity must be maintained by truncation at zero.

2.5. The restricted maximum likelihood estimator (REML)

A second likelihood approach to the estimation of τ^2 in a random effects model is the restricted maximum likelihood estimator [2, 6]. Under $Y_i \sim \mathrm{N}(\theta, \hat{\sigma}^2_i+\tau^2)$, the restricted maximum likelihood estimate of τ^2 is found using the log likelihood function [6, 25]

$$L_R(\theta,\tau^2)=-\frac{k}{2}\log(2\pi)-\frac{1}{2}\sum_{i=1}^{k}\log(\hat{\sigma}^2_i+\tau^2)-\frac{1}{2}\log\left(\sum_{i=1}^{k}\frac{1}{\hat{\sigma}^2_i+\tau^2}\right)-\frac{1}{2}\sum_{i=1}^{k}\frac{(y_i-\hat{\theta}_w)^2}{\hat{\sigma}^2_i+\tau^2}$$

where $\hat{\theta}_w=\sum_{i=1}^{k}w_i y_i/\sum_{i=1}^{k}w_i$ and $w_i=1/(\hat{\sigma}^2_i+\tau^2)$. The REML estimator $\hat{\tau}^2_{\text{RE}}$ based on $L_R(\theta,\tau^2)$ is an iterative solution of the equation

$$\hat{\tau}^2_{\text{RE}}=\frac{\sum_{i=1}^{k}\hat{w}^2_i\{(y_i-\hat{\theta}_{\hat{w}})^2+1/\sum_{j=1}^{k}\hat{w}_j-\hat{\sigma}^2_i\}}{\sum_{i=1}^{k}\hat{w}^2_i} \tag{6}$$

where $\hat{w}_i=1/(\hat{\sigma}^2_i+\hat{\tau}_{\text{RE}})$. As with ML, an initial estimate of τ^2 is required, non-negativity must be enforced at each iteration step, and iteration continues until convergence for a specified stopping criterion. As pointed out by DerSimonian and Laird [2], the REML estimator is different from the ML estimator because the REML function is adjusted to account for estimation of both θ and τ^2 from the same data.

In addition, an approximate REML estimator of τ^2 is frequently used in the random effects model, with a direct adjustment for the loss of degrees of freedom due to estimating θ in full

DOI: 10.1002/sim

maximum likelihood estimation. The approximate REML estimator of τ^2 is given as an iterative solution to the following equation [12, 19–21]:

$$\hat{\tau}^2_{\mathrm{ARE}} = \frac{\sum_{i=1}^{k} \hat{w}_i^2 \{(k/(k-1))(y_i - \hat{\theta}_{\hat{w}})^2 - \hat{\sigma}_i^2\}}{\sum_{i=1}^{k} \hat{w}_i^2} \tag{7}$$

where $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}^2_{\mathrm{ARE}})$. In our experience, the two estimators yield almost identical estimates for τ^2. Hence, we only consider $\hat{\tau}^2_{\mathrm{RE}}$ from (6) in our comparisons.

2.6. *The empirical Bayes estimator (EB)*

An estimator of τ^2 that is somewhat similar in form to the likelihood estimators has also been used in random effects meta-analysis ([12, 20, 21] etc.). The estimator was introduced for empirical Bayes inference by Morris [19], and hence it has been called the empirical Bayes estimator of τ^2 in meta-analysis (e.g. References [12, 21]). An EB estimate of τ^2 is obtained as an iterative solution to the equation [19, 20]

$$\hat{\tau}^2_{\mathrm{EB}} = \frac{\sum_{i=1}^{k} \hat{w}_i \{(k/(k-1))(y_i - \hat{\theta}_{\hat{w}})^2 - \hat{\sigma}_i^2\}}{\sum_{i=1}^{k} \hat{w}_i} \tag{8}$$

where $\hat{\theta}_{\hat{w}} = \sum_{i=1}^{k} \hat{w}_i y_i / \sum_{i=1}^{k} \hat{w}_i$ and $\hat{w}_i = 1/(\hat{\sigma}_i^2 + \hat{\tau}^2_{\mathrm{EB}})$. Like the ML and REML estimators, an initial estimate of τ^2 must be supplied, and $\hat{\tau}^2_{\mathrm{EB}} \geqslant 0$ must be enforced at each iteration to prevent a negative estimate. Note that one would obtain the EB estimator by replacing $\hat{w}_i^2$ with $\hat{w}_i$ in the approximate REML estimator (7).

By referring to an earlier paper by Fay and Herriot [26], we have noticed that the MV estimator is related to the EB estimator because both of these estimators can be linked to the equation $E\{\sum_{i=1}^{k} w_i (y_i - \hat{\theta}_w)^2\} = k - 1$ under model (1). In particular, the two estimators can both be derived by setting $\sum_{i=1}^{k} w_i (y_i - \hat{\theta}_w)^2$ equal to its expected value to produce the equation $\sum_{i=1}^{k} \hat{w}_i (y_i - \hat{\theta}_{\hat{w}})^2 = k - 1$. The two estimators may be derived by different manipulations of this equation. As a result, if one were to iterate the estimator $\hat{\tau}^2_{\mathrm{MV}}$ for the same convergence criterion, one would typically obtain a value very close to $\hat{\tau}^2_{\mathrm{EB}}$. Nevertheless, the two estimators yield different estimates of τ^2 in their presently proposed forms, and they are of different natures as estimators of τ^2. Specifically, the MV estimator in (4) as proposed by Sidik and Jonkman [11] is a non-iterative and non-negative estimator, and the EB estimator is an iterative estimator for which non-negativity must be enforced.

3. AN EXAMPLE

To illustrate the seven estimators in practice, and the possibility of considerable differences among the estimated values of τ^2, we consider the postoperative complication data from 29 randomized clinical trials for comparing laparoscopic inguinal hernia repair (LIHR) and the conventional open inguinal hernia repair (OIHR), compiled by Memon *et al.* [27]. The sample log odds ratios from the 29 trials and their asymptotic 95 per cent confidence limits are plotted in Figure 1. Note that we have

DOI: 10.1002/sim

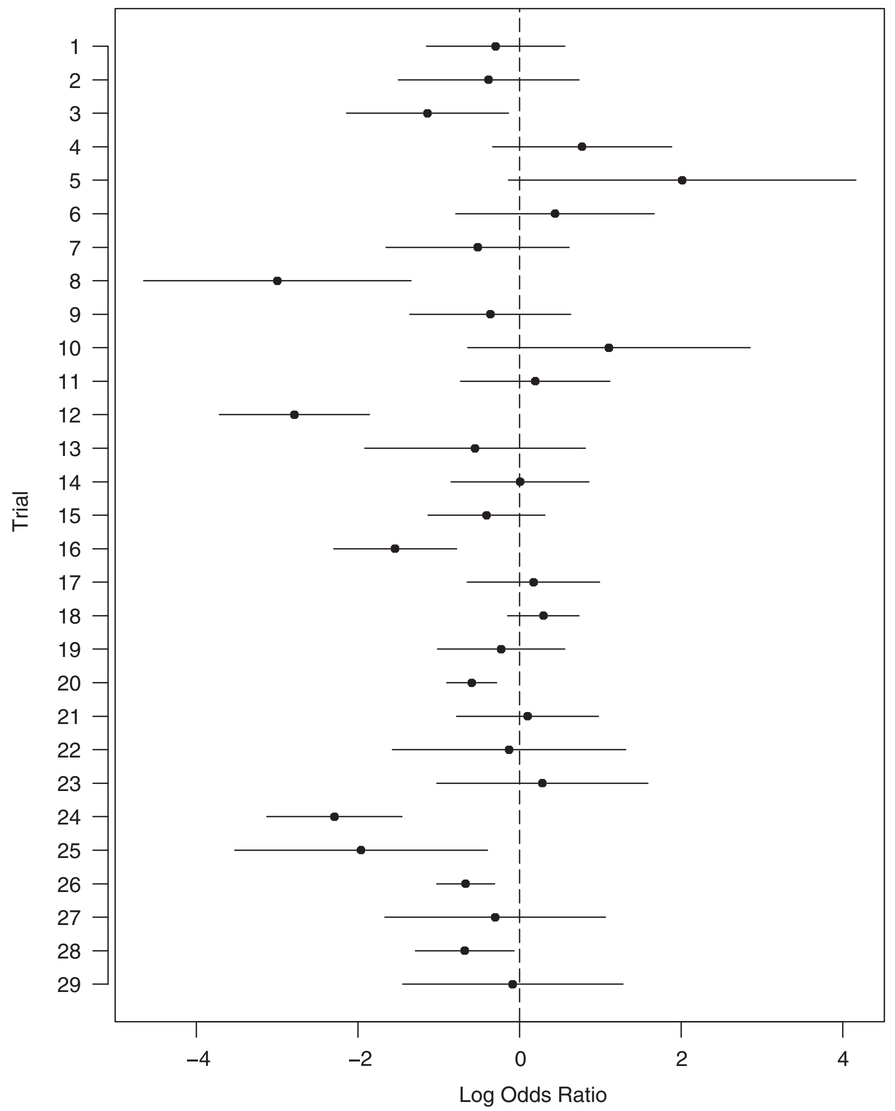

Figure 1. Odds ratio estimates and asymptotic 95 per cent confidence limits for the 29 clinical trials comparing LIHR and OIHR.

numbered the studies as $1, \ldots, 29$, where Memon *et al*. [27] used reference numbers $31, \ldots, 59$ in their paper. The complete data showing the number of postoperative complication cases and the number of patients can be found in the paper by Memon *et al*. [27]. In fact, Memon *et al.* reported a comprehensive meta-analysis study of the 29 randomized clinical trials for comparing LIHR and OIHR using six relevant outcome measures, including the postoperative complication rate. However, for illustrative comparisons of the heterogeneity variance estimators, we only consider the postoperative complication rates in this paper. Note that the log odds ratios suggest that there is

DOI: 10.1002/sim

substantial heterogeneity among the studies, as the values range from approximately −3 for study 8 (corresponding to an odds ratio of about 0.05, indicating that postoperative complications are far less likely with LIHR than OIHR) to approximately 2 for study 5 (corresponding to an odds ratio of about 7.5, suggesting that postoperative complications are more likely with LIHR, although the confidence limits for this study include zero). However, only 9 of the 29 study-specific log odds ratios differed significantly from zero.

The seven heterogeneity variance estimators discussed in Section 2 were applied to the data, using the log odds ratio to measure effect size. The estimated values of τ^2, along with the estimated overall log odds ratio $\hat{\theta}$ based on each estimator, are provided in Table I. It is apparent that $\hat{\tau}^2$ differs substantially across the seven estimators, with values ranging from $\hat{\tau}^2_{\mathrm{MM}} = 0.429$ to $\hat{\tau}^2_{\mathrm{VC}} = 0.841$. It is noteworthy that the largest estimate is nearly twice as large as the smallest estimate, particularly considering that the estimators VC and MM are both method of moments estimators and are fairly commonly used in practice (especially the MM estimator), in part because they are easy to calculate. Considering that the other estimated values of τ^2 are all substantially larger than $\hat{\tau}^2_{\mathrm{MM}}$, it is possible that $\hat{\tau}^2_{\mathrm{MM}}$ may underestimate the true value of τ^2 in this case (see also References [7, 8, 11]). Among the three iterative estimators, the REML estimate 0.598 is slightly larger than the ML estimate 0.562, and the EB estimate 0.703 is larger than both the ML and REML estimates. Finally, the model variance estimator $\hat{\tau}^2_{\mathrm{MV}} = 0.818$ yields the second largest estimate, while the MVvc estimate falls roughly midway between the EB and MV estimates. These results show that the estimated heterogeneity among studies may depend heavily on the estimator used, and suggest the utility of a comprehensive comparison study to assess the accuracy of the available estimators. Although the heterogeneity in this example may seem rather large in light of the study-specific odds ratios and the estimated values of τ^2, we note that several other example data sets in the literature exhibit similar or larger degrees of heterogeneity among log odds ratios (for example, see References [10, 12, 28]).

In addition, it should be noted that the considerable differences among the seven estimators shown in Table I led to fairly large differences among the estimated standard errors for the overall effect, although the point estimates of the overall effect ($\hat{\theta}$) did not differ greatly across the seven values of $\hat{\tau}^2$. It seems clear that differences of this magnitude among the estimated standard errors for an estimate of the same overall effect could lead to conflicting conclusions regarding the true overall effect, depending upon which estimator of τ^2 is used. In fact, for this particular example all seven 95 per cent confidence intervals for θ lead to the inference that LIHR produces a lower

Table I. Estimation for 29 clinical trials of LIHR *versus* OIHR.

				95% Confidence limits for θ	
Estimator	$\hat{\tau}^2$	$\hat{\theta}$	SE($\hat{\theta}$)	Lower limit	Upper limit
MM	0.429	−0.477	0.155	−0.782	−0.173
VC	0.841	−0.467	0.197	−0.854	−0.080
ML	0.562	−0.473	0.170	−0.807	−0.140
REML	0.598	−0.472	0.174	−0.813	−0.131
EB	0.703	−0.470	0.185	−0.832	−0.108
MV	0.818	−0.467	0.195	−0.850	−0.084
MVvc	0.747	−0.469	0.189	−0.839	−0.099

DOI: 10.1002/sim

postoperative complication rate than OIHR. However, it clearly need not be the case in general that all seven estimators will lead to the same conclusion.

4. SIMULATION STUDY

To assess the empirical properties of the heterogeneity variance estimators discussed in Section 2, we conducted Monte Carlo simulation. We compared the seven estimators, namely the four non-iterative estimators VC, MM, MV, and MVvc and the three iterative estimators ML, REML, and EB, in terms of empirical bias and mean squared error (MSE) based on the simulation. We chose to concentrate on the log odds ratio as the measure of effect size, on the grounds of its popularity in meta-analysis due to its usage in retrospective or case control studies, as noted by DerSimonian and Laird [2].

In this simulation, we selected five meta-analysis sample sizes: $k=10$, 15, 20, 30, 50. For each k, we considered three values of the overall effect, $\theta=0.5$, 0, -0.5, and eleven values of the heterogeneity variance, $\tau^2=0(0.10)0.50(0.25)1.75$. The values of k and τ^2 were selected to try and cover the full range of plausible values in meta-analysis. To place the τ^2 values in the context of odds ratios, consider that when $\tau^2=0.5$ and $\theta=0$, 95 per cent of the true log odds ratios (θ_i) would be in the interval $0 \pm 1.96\sqrt{0.5}$. In other words, 95 per cent of the true odds ratios would be in the interval $\exp(0 \pm 1.96\sqrt{0.5}) = (0.25, 4.0)$. Although this may seem to be an extremely large degree of heterogeneity, we note that larger apparent values of τ^2 do occur in practice. For example, Efron [28] obtained the estimate $\hat{\tau}^2=1.416$ for $k=39$ trials of surgical treatment of stomach ulcers, and Sidik and Jonkman [11] obtained (0.887, 2.206) as an interval estimate for τ^2 using the same data set. The value $\theta=-0.5$ was chosen because it is close to the estimated log odds ratio in the hernia repair example, and we chose $\theta=0.5$ to check whether the labelling of the treatment and control groups affected the results (because changing $\theta=-0.5$ to $\theta=0.5$ simply inverts the odds ratio, so that one would expect the results for the two values to be very similar). We also selected $\theta=0$ to investigate the behaviour of the estimators when there is no treatment effect.

For each combination of k, τ^2, and θ, we generated data for k 2×2 tables using a method similar to Berkey *et al.* [20], Platt *et al.* [29], and Knapp and Hartung [21]. Specifically, we first generated θ_i for $i=1,\ldots,k$, using $\theta_i \sim \mathrm{N}(\theta, \tau^2)$. For a given k, we selected the within-study sample sizes n_{iC} and n_{iT} for the treatment and control groups, respectively, with equal sample size for the ith study, i.e. $n_{iC}=n_{iT}=n_i$. The equal sample sizes n_i for $i=1,\ldots,k$ were determined by randomly sampling with replacement from the integers between 20 and 200. Next, the responses c_i ($i=1,\ldots,k$) for the control group were obtained from the binomial (n_{iC}, p_{iC}) distribution, where the binomial probability p_{iC} was randomly chosen from a uniform distribution on the interval [0.05, 0.65]. The responses a_i of the treatment group were generated from the binomial (n_{iT}, p_{iT}) distribution, where $p_{iT}=p_{iC}\exp\{\theta_i\}/(1-p_{iC}+p_{iC}\exp\{\theta_i\})$. Note that this maintains the log odds ratio for study i at $\theta_i=\mathrm{logit}(p_{iT})-\mathrm{logit}(p_{iC})$. Thus, for each set of k, θ, and τ^2 we obtained the 2×2 table data (a_i, b_i, c_i, d_i) for $i=1,\ldots,k$, where $b_i=n_{iT}-a_i$ and $d_i=n_{iC}-c_i$. If any zero cells were generated, we added 0.5 to each cell for all k tables, in order to avoid singularities. For each such set of parameters, we replicated this 10 000 times. At each replicate, the sample log odds ratios $y_i,\ldots,y_k$ and their estimated asymptotic variances were calculated. From these values, the seven heterogeneity variance estimators for τ^2 were computed. For each set of k, θ, and τ^2, we computed the empirical means, variances, biases, and MSEs of these estimators from their respective 10 000 estimates of τ^2.

In computing the three iterative estimators ML, REML, and EB, we set the initial value of τ^2 at zero and restricted to a maximum of 20 iterations for convergence. After each iteration, non-negativity of the estimates was checked and enforced. Convergence was determined to be attained at the jth iteration if the following stopping criterion was satisfied:

$$\frac{|\hat{\tau}^2_{(j+1)} - \hat{\tau}^2_{(j)}|}{1 + \hat{\tau}^2_{(j)}} < 0.00001$$

This stopping criterion, as well as the limit of 20 iterations, was selected by adopting a similar convergence rule proposed by Swallow and Monahan [30]. The addition of 1 to the denominator is done to prevent singularity and to keep the criterion from being overly stringent for very small values of the estimators. If the number of iterations exceeded 20 for any one of the three iterative estimators, we discarded that particular replicate for all seven estimators, in order to maintain an equal number of replicates for all estimators. Note that the three iterative estimators rarely exceeded the maximum number of iterations in this simulation study. For example, with $\theta = 0.5$ the three iterative estimators converged within 20 iterations in all cases for $k > 20$, so that all 10 000 replicates were used. For $k = 20$, only one case had fewer than 10 000 replicates—specifically, $\tau^2 = 0.10$ with 9999 replicates. For $k < 20$, all 22 cases had fewer than 10 000 replicates, but the smallest number of replicates for any case was 9974. The empirical properties of the estimators were calculated on the basis of the actual number of replicates for each set of parameters k, θ, and τ^2, instead of the intended number of replicates (i.e. 10 000).

5. RESULTS OF THE COMPARISONS

The empirical biases, variances, and MSEs of each of the seven estimators were computed from the simulation for all the combinations of k, θ, and τ^2 listed in Section 3. For $\theta = 0.5$, the biases and MSEs are presented in Tables II and III, respectively. For the other two values of θ, i.e. $\theta = 0, -0.5$, the biases and MSEs are plotted in Figures 1 and 2 for $k = 20$ and $k = 50$, respectively, as a summary.

5.1. Comparing bias

From Table II it can be seen that the three estimators MM, ML, and REML underestimate the heterogeneity variance for all values of k and τ^2, with increasing magnitude of bias as the true value of τ^2 increases. The downward bias of these estimators has also been noted in some other studies. For example, Berkey *et al.* [20] noted bias in REML; Thompson and Sharp [12] mentioned bias in ML and REML; Malzahn *et al.* [7] and Böhning *et al.* [8] noted bias in MM (see also Reference [11]). On the other hand, the VC estimator in general overestimates τ^2 for all cases with $\theta = 0.5$, although the magnitude of the bias is not as large as for the three negatively biased estimators. The MV estimator generally has some upward bias for $\tau^2 \leqslant 1$, but its bias decreases as τ^2 increases, and eventually the bias becomes negative for large values of τ^2. The EB and MVvc estimators also have upward bias for small values of k and τ^2 and downward bias for large values of τ^2, but overall these two estimators appear to be the best among all the estimators, having the smallest magnitude of bias in general. This is particularly apparent for small values of k (with all values of τ^2), and for large k with small to moderate τ^2.

DOI: 10.1002/sim

Table II. Bias of the τ^2 estimators for $\theta = 0.5$.

Estimator	τ^2 0.00	0.10	0.20	0.30	0.40	0.50	0.75	1.00	1.25	1.50	1.75
						$k = 10$					
MM	0.015	−0.002	−0.006	−0.016	−0.025	−0.045	−0.090	−0.150	−0.232	−0.329	−0.420
VC	0.040	0.016	0.017	0.020	0.026	0.032	0.039	0.053	0.062	0.047	0.067
ML	0.009	−0.021	−0.033	−0.048	−0.060	−0.077	−0.112	−0.150	−0.197	−0.258	−0.297
REML	0.013	−0.002	−0.002	−0.006	−0.007	−0.013	−0.023	−0.034	−0.054	−0.092	−0.104
EB	0.017	0.004	0.007	0.006	0.008	0.007	0.006	0.007	0.000	−0.026	−0.023
MV	0.060	0.044	0.040	0.035	0.034	0.032	0.026	0.025	0.018	−0.009	−0.005
MVvc	0.029	0.007	0.009	0.010	0.014	0.015	0.015	0.018	0.013	−0.012	−0.006
						$k = 15$					
MM	0.012	−0.002	−0.009	−0.020	−0.035	−0.045	−0.106	−0.177	−0.267	−0.359	−0.475
VC	0.034	0.015	0.015	0.019	0.025	0.036	0.044	0.052	0.058	0.065	0.071
ML	0.008	−0.016	−0.027	−0.039	−0.050	−0.054	−0.091	−0.128	−0.168	−0.208	−0.257
REML	0.011	−0.003	−0.005	−0.010	−0.014	−0.011	−0.030	−0.050	−0.073	−0.097	−0.128
EB	0.014	0.003	0.004	0.002	0.003	0.010	0.002	−0.005	−0.015	−0.024	−0.037
MV	0.063	0.049	0.042	0.037	0.034	0.039	0.029	0.021	0.011	0.002	−0.010
MVvc	0.025	0.007	0.007	0.007	0.009	0.017	0.012	0.007	−0.001	−0.008	−0.019
						$k = 20$					
MM	0.012	−0.003	−0.009	−0.021	−0.034	−0.052	−0.105	−0.181	−0.260	−0.365	−0.474
VC	0.031	0.013	0.015	0.016	0.022	0.026	0.040	0.050	0.071	0.064	0.076
ML	0.007	−0.013	−0.022	−0.033	−0.042	−0.054	−0.081	−0.117	−0.148	−0.197	−0.243
REML	0.010	−0.003	−0.005	−0.011	−0.015	−0.021	−0.036	−0.059	−0.077	−0.113	−0.146
EB	0.013	0.003	0.003	0.001	0.001	−0.001	−0.003	−0.012	−0.012	−0.034	−0.047
MV	0.066	0.052	0.045	0.037	0.034	0.030	0.026	0.018	0.018	−0.003	−0.013
MVvc	0.024	0.006	0.007	0.005	0.007	0.006	0.006	0.001	0.003	−0.017	−0.026
						$k = 30$					
MM	0.008	−0.004	−0.012	−0.022	−0.036	−0.054	−0.113	−0.193	−0.285	−0.387	−0.506
VC	0.032	0.011	0.012	0.019	0.021	0.025	0.033	0.034	0.037	0.029	0.019
ML	0.006	−0.011	−0.019	−0.026	−0.035	−0.044	−0.070	−0.106	−0.142	−0.187	−0.237
REML	0.007	−0.004	−0.008	−0.011	−0.017	−0.023	−0.040	−0.068	−0.095	−0.132	−0.174
EB	0.009	−0.001	−0.002	−0.001	−0.003	−0.005	−0.011	−0.026	−0.040	−0.063	−0.090
MV	0.068	0.055	0.046	0.043	0.037	0.034	0.025	0.009	−0.004	−0.027	−0.053
MVvc	0.022	0.001	0.002	0.005	0.003	0.003	−0.001	−0.014	−0.026	−0.048	−0.073
						$k = 50$					
MM	0.008	−0.005	−0.011	−0.022	−0.039	−0.055	−0.116	−0.194	−0.283	−0.394	−0.515
VC	0.024	0.012	0.017	0.023	0.026	0.034	0.041	0.048	0.049	0.039	0.022
ML	0.005	−0.010	−0.016	−0.022	−0.032	−0.040	−0.069	−0.102	−0.139	−0.191	−0.250
REML	0.006	−0.006	−0.009	−0.013	−0.021	−0.027	−0.050	−0.079	−0.111	−0.158	−0.212
EB	0.009	0.000	0.000	0.000	−0.003	−0.003	−0.014	−0.027	−0.044	−0.075	−0.112
MV	0.072	0.058	0.050	0.045	0.040	0.038	0.025	0.012	−0.004	−0.033	−0.068
MVvc	0.019	0.005	0.006	0.007	0.005	0.006	−0.002	−0.013	−0.028	−0.057	−0.091

More specifically, we note that the REML estimator has smaller bias than both MM and ML in all the cases considered, except for $\tau^2 = 0$ and $\tau^2 = 0.1$, the smallest τ^2 values considered, for which MM and REML have essentially the same bias. As expected, the difference in bias between

DOI: 10.1002/sim

Table III. MSE of the τ^2 estimators for $\theta = 0.5$.

	τ^2										
Estimator	0.00	0.10	0.20	0.30	0.40	0.50	0.75	1.00	1.25	1.50	1.75
						$k=10$					
MM	0.001	0.007	0.018	0.032	0.050	0.069	0.135	0.222	0.327	0.448	0.614
VC	0.006	0.016	0.033	0.055	0.080	0.117	0.216	0.345	0.527	0.678	0.917
ML	0.000	0.006	0.017	0.031	0.048	0.071	0.142	0.237	0.363	0.491	0.667
REML	0.001	0.007	0.020	0.036	0.056	0.082	0.164	0.270	0.412	0.544	0.741
EB	0.001	0.009	0.022	0.040	0.061	0.091	0.178	0.292	0.449	0.586	0.798
MV	0.006	0.011	0.023	0.041	0.062	0.091	0.178	0.291	0.448	0.583	0.795
MVvc	0.003	0.011	0.025	0.043	0.065	0.095	0.184	0.298	0.457	0.594	0.808
						$k=15$					
MM	0.001	0.005	0.012	0.022	0.033	0.049	0.092	0.153	0.232	0.332	0.473
VC	0.004	0.011	0.022	0.036	0.054	0.079	0.144	0.235	0.339	0.449	0.589
ML	0.000	0.005	0.012	0.022	0.035	0.051	0.100	0.168	0.249	0.341	0.454
REML	0.001	0.005	0.013	0.024	0.038	0.056	0.108	0.179	0.262	0.356	0.468
EB	0.001	0.006	0.015	0.026	0.041	0.062	0.117	0.194	0.283	0.381	0.500
MV	0.006	0.009	0.016	0.027	0.042	0.062	0.117	0.194	0.284	0.381	0.500
MVvc	0.002	0.008	0.017	0.028	0.043	0.064	0.120	0.198	0.288	0.386	0.506
						$k=20$					
MM	0.001	0.004	0.009	0.016	0.026	0.036	0.071	0.121	0.196	0.292	0.422
VC	0.003	0.008	0.017	0.026	0.040	0.055	0.106	0.169	0.259	0.344	0.437
ML	0.000	0.004	0.010	0.017	0.027	0.038	0.076	0.123	0.193	0.267	0.352
REML	0.000	0.004	0.010	0.018	0.028	0.040	0.079	0.126	0.197	0.269	0.349
EB	0.001	0.005	0.011	0.019	0.031	0.043	0.085	0.136	0.212	0.286	0.366
MV	0.006	0.008	0.013	0.020	0.031	0.043	0.086	0.137	0.213	0.286	0.366
MVvc	0.002	0.006	0.013	0.021	0.032	0.045	0.087	0.139	0.215	0.290	0.369
						$k=30$					
MM	0.000	0.002	0.006	0.011	0.017	0.024	0.052	0.093	0.156	0.247	0.372
VC	0.003	0.007	0.013	0.020	0.029	0.038	0.072	0.110	0.159	0.208	0.260
ML	0.000	0.002	0.006	0.011	0.018	0.026	0.053	0.086	0.131	0.185	0.245
REML	0.000	0.003	0.006	0.012	0.018	0.026	0.053	0.086	0.128	0.179	0.234
EB	0.000	0.003	0.007	0.013	0.020	0.028	0.056	0.089	0.132	0.180	0.230
MV	0.006	0.006	0.009	0.015	0.021	0.029	0.057	0.089	0.131	0.177	0.225
MVvc	0.001	0.004	0.008	0.014	0.021	0.029	0.058	0.091	0.133	0.180	0.229
						$k=50$					
MM	0.000	0.002	0.004	0.007	0.011	0.016	0.038	0.073	0.128	0.214	0.336
VC	0.001	0.004	0.007	0.012	0.017	0.024	0.044	0.067	0.093	0.120	0.155
ML	0.000	0.002	0.004	0.008	0.012	0.017	0.034	0.057	0.085	0.124	0.176
REML	0.000	0.002	0.004	0.008	0.011	0.016	0.033	0.055	0.081	0.116	0.164
EB	0.000	0.002	0.005	0.008	0.012	0.017	0.034	0.054	0.077	0.106	0.144
MV	0.006	0.005	0.007	0.010	0.013	0.019	0.034	0.053	0.076	0.102	0.136
MVvc	0.001	0.003	0.005	0.009	0.013	0.018	0.035	0.054	0.077	0.104	0.141

ML and REML for a given value of τ^2 is decreasing as k increases, because $k/(k-1)$ is the only factor distinguishing these two estimators. Note that the ML estimator has relatively large bias for small to moderate values of k, because it fails to account for the loss in degrees of freedom due to

DOI: 10.1002/sim

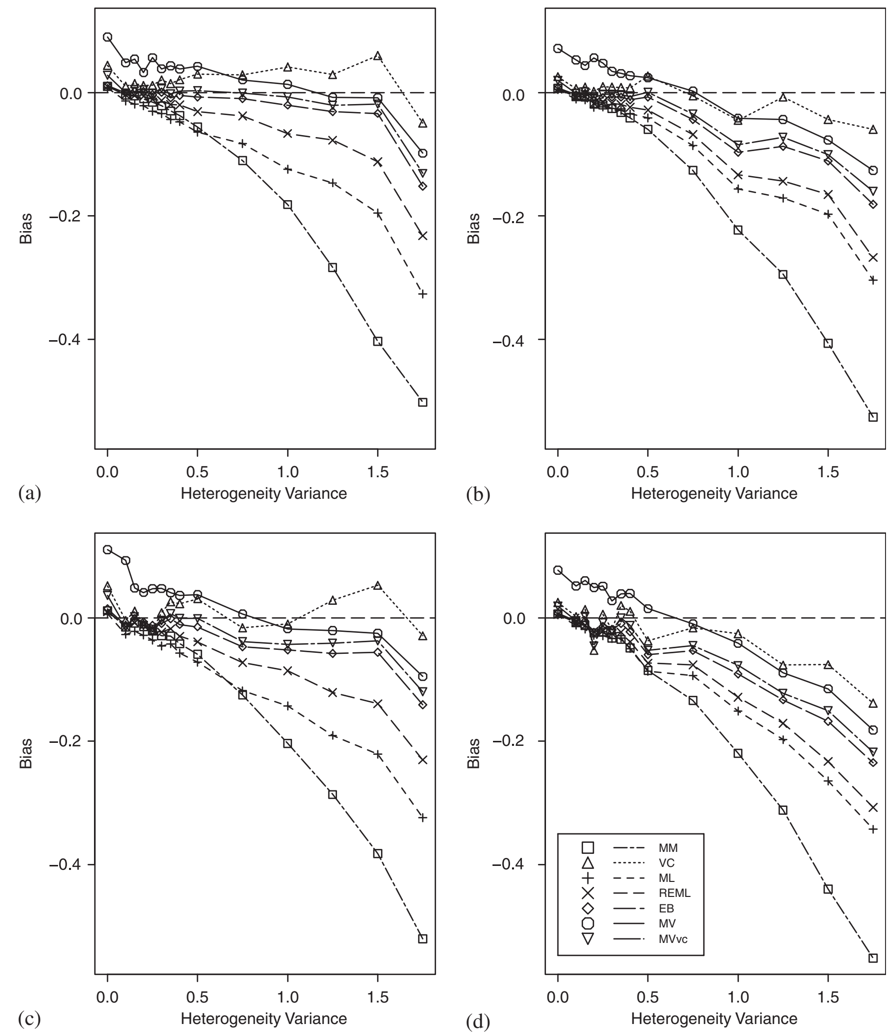

Figure 2. Empirical bias of τ^2 estimators: (a) $\theta=0$ and $k=20$; (b) $\theta=0$ and $k=50$; (c) $\theta=-0.5$ and $k=20$; and (d) $\theta=-0.5$ and $k=50$ (for legend see (d)).

estimation of θ (see also Reference [2]). Comparing the VC and MV estimators, we note that VC has smaller positive bias than MV for $\tau^2 \leqslant 0.5$ and all k. For $\tau^2 > 0.5$, MV generally has a smaller magnitude of bias than VC, apart from a few cases where both k and τ^2 are large. In particular, the VC estimator performs as well as any other estimator in terms of overall bias when $k=50$. The MVvc and EB estimators have consistently smaller bias than the MM, ML, and REML estimators, except for some very small values of τ^2 and k, in which REML or MM have slightly smaller bias. Furthermore, MVvc is a better estimator than MV in terms of bias. The bias reduction from using MVvc instead of MV (i.e. from using $\hat{\tau}^2_{\mathrm{VC}}$ to construct the estimated ratios, instead of the crude ratio estimate) is considerable for small to moderate values of τ^2, although MV has slightly smaller

DOI: 10.1002/sim

bias than MVvc for large τ^2. Clearly, the model variance estimator is improved in terms of bias by using the VC estimator as an initial value in estimating the r_i, particularly for small to moderately small values of τ^2. Similarly, EB has smaller bias than MV for small to moderate values of τ^2, but for large values of τ^2 the MV estimator has roughly the same or smaller magnitude of bias than both EB and MVvc. For $\tau^2 \leqslant 0.5$, the range in which the VC estimator outperforms the MV estimator, both the EB and MVvc estimators have smaller bias than the VC estimator.

Finally, note that the empirical biases of the seven estimators for $\theta = 0, -0.5$ are fairly similar to the case of $\theta = 0.5$. For a graphical summary of the empirical bias, see Figure 2 for $k = 20, 50$. However, for $\theta = 0, -0.5$ it may be important to note that the VC estimator performs as well in general as the MVvc and EB estimators, even for $k = 20$, whereas for $\theta = 0.5$ this was true only for large values of k. Also, for $\theta = 0$ and $\theta = -0.5$ the EB estimator has downward bias rather than upward bias for small values of τ^2, particularly for $\theta = -0.5$.

5.2. Comparing MSE

As shown in Table III, the MM, ML, and REML estimators have smaller MSEs than the VC, EB, MV, and MVvc estimators when the number of studies is relatively small ($k \leqslant 20$), except when τ^2 is large. For small τ^2 and k, the REML estimator tends to have larger MSE than the MM and ML estimators, which have very similar MSEs. As the number of studies increases, the MSEs of MM, ML, and REML become comparable to the MSEs of the other four estimators, again except for cases where τ^2 is large. For large τ^2 and moderate to large k the MM, ML, and REML estimators, and in particular the MM estimator, have the largest MSEs among the seven estimators. Clearly, the increasing bias of MM with increasing values of τ^2 eventually becomes sufficiently large to dominate its variance in calculating MSE. In general, the VC estimator tends to have the largest MSE among the seven estimators, except for large values of τ^2 and moderate to large k. In fact, the empirical variance of $\hat{\tau}^2_{\mathrm{VC}}$ is consistently larger than the empirical variances of the other six estimators for all cases considered. The MV and MVvc estimators have nearly the same MSEs in most cases, although the MSE tends to be slightly smaller for the MV estimator than for the MVvc estimator as τ^2 increases. The EB estimator also has approximately the same MSEs as both the MV and MVvc estimators, although its MSE is slightly smaller than MV for small values of τ^2, and slightly larger than MV for large values of τ^2. Note that the variances of the EB and MV estimators are essentially identical for all τ^2 and k. Hence, the slight superiority of EB over MV in terms of MSE for small to moderately large τ^2, and the slight superiority of MV over EB for large τ^2, are caused by the differences in bias noted in the preceding section.

For $\theta = 0$ and $\theta = -0.5$, the results of comparisons concerning the empirical MSEs of the seven estimators are fairly similar to the comparisons for $\theta = 0.5$ discussed above. For a graphical summary of the MSEs of the seven estimators for $\theta = 0, -0.5$ and $k = 20, 50$, see Figure 3.

5.3. Choosing an estimator: bias and MSE

We have noted that the three estimators MM, ML, and REML have the smallest MSEs in many of the comparisons among the seven estimators. The key features of these three estimators are that they all are forcibly bound below by zero, they all have negative biases for all the values of τ^2 and k considered, and the amount of bias increases considerably as τ^2 increases. We also note that these three estimators tend to have the smallest empirical variances in general among the seven estimators. A similar phenomenon was noted by Swallow and Monahan [30] in their comparison of estimators for variance components. It appears that the downward biases of the

DOI: 10.1002/sim

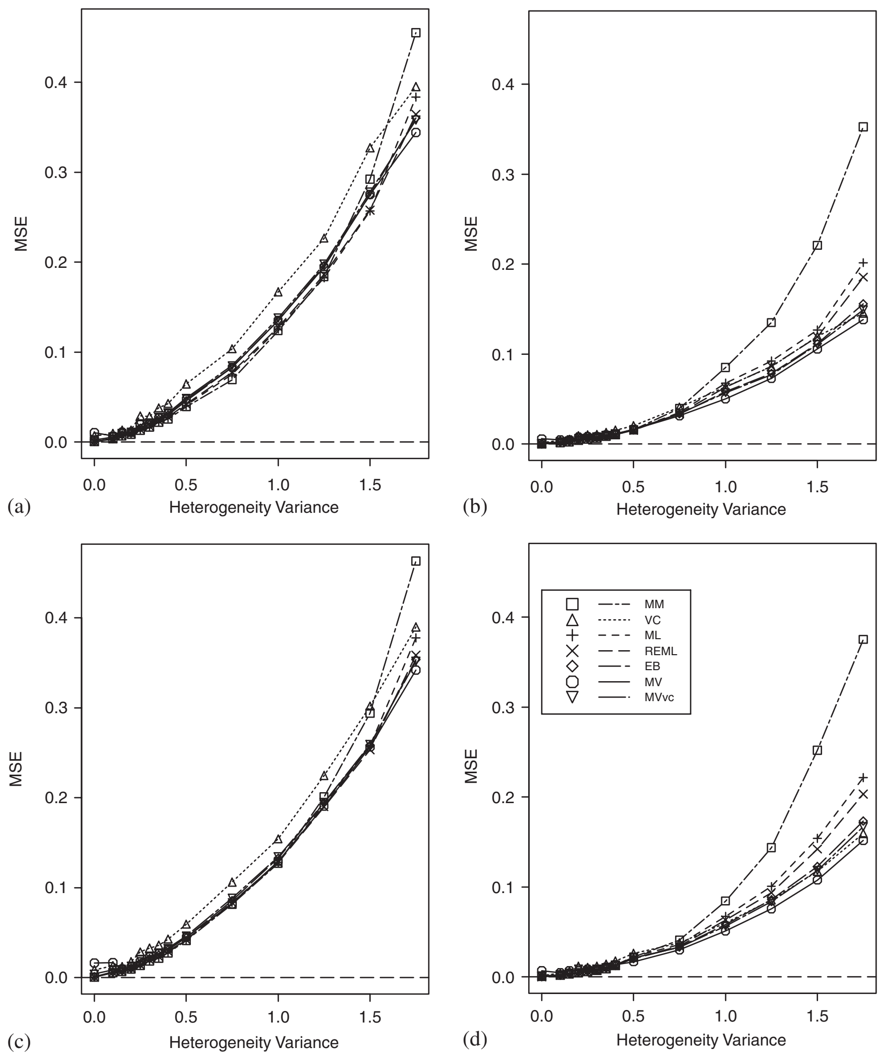

Figure 3. Empirical MSE of τ^2 estimators: (a) $\theta=0$ and $k=20$; (b) $\theta=0$ and $k=50$; (c) $\theta=-0.5$ and $k=20$; and (d) $\theta=-0.5$ and $k=50$ (for legend see (d)).

MM, ML, and REML estimators, along with their enforced non-negativity, tend to concentrate their sampling distributions, and hence their variances are reduced sufficiently to offset the squared bias. This in turn leads to a decrease in MSE, as explained by Swallow and Monahan [30] (see also Reference [11]). This interesting phenomenon may support the arguments of Casella and Berger [31, p. 305] that MSE may not be a good criterion when estimating scale parameters, and strengthens our belief that bias is more informative than MSE in assessing variance estimators.

In summary, then, the MM, ML, and REML estimators underestimate the heterogeneity variance τ^2 for all values of k and τ^2 considered in the simulation, with increasing magnitude of bias as the

DOI: 10.1002/sim

true value of τ^2 increases. This finding seems to agree with the results and observations of some other studies (e.g. References [7, 8, 11, 12, 20]), some of which used continuous rather than discrete outcome measures. Although these three estimators of τ^2 generally have smaller MSEs compared to the other four estimators, the substantial negative bias of these estimators with increasing τ^2, and the effects of the negative bias on the MSEs of these estimators described in the preceding paragraph, may lead one to conclude that MM, ML, and REML are not good estimators for τ^2 unless the true heterogeneity variance is small. To see the effects of the negative bias in these estimators, consider the case with $k=10$, $\theta=0.5$, and $\tau^2=0.5$. The true values $\theta=0.5$ and $\tau^2=0.5$ imply that 95 per cent of the true odds ratios will be in the interval $\exp(0.5\pm 1.96\sqrt{0.5})=(0.41, 6.59)$. The empirical bias of the MM estimator from the simulations is -0.045 in this case, which translates to an average value of $\hat{\tau}^2_{\mathrm{MM}}=0.455$, so we would estimate (assuming $\hat{\theta}=0.5$) that 95 per cent of the true odds ratios would be in the interval (0.44, 6.18). For the ML estimator, the empirical bias is -0.077, which would lead to the estimate that 95 per cent of the true odds ratios would be in the interval (0.46, 5.90). Although these differences may not seem excessively large, they are large enough to be of concern, and large enough to potentially cause inaccurate inferences. Furthermore, because the heterogeneity variance appears in all the study weights and is used in estimation of the variance of $\hat{\theta}$, it is likely to have a large influence on meta-analytic inference [12, 22]. It seems logical that underestimation of τ^2 would likely lead to underestimation of the variance of $\hat{\theta}$, which in turn could lead to falsely significant tests of the overall effect, along with confidence intervals that have coverage lower than the nominal level.

On the other hand, the VC estimator generally overestimates τ^2 in all cases with $\theta=0.5$. However, it is simple to calculate, and is one of the more accurate estimators for τ^2 in terms of bias as long as k is large ($k\geqslant 30$), although it has the largest MSE among the seven estimators in general, due to its large variance. Friedman [13] derived variances for the MM and VC estimators, and found that the MM estimator would have smaller variance than the VC estimator for small τ^2, while the VC estimator would have smaller variance for large τ^2. We have found an approximately opposite result in our simulations, namely that for large τ^2 the MM estimator has very small variance but large bias, and that the VC estimator has larger variance even when τ^2 is large. However, Friedman did not consider the fact that both $\hat{\tau}^2_{\mathrm{MM}}$ and $\hat{\tau}^2_{\mathrm{VC}}$ are truncated at zero in practice, so we may not necessarily expect his results to hold in the simulations. The MV estimator has larger bias than the other six estimators for relatively small values of τ^2 (i.e. $\tau^2 < 0.5$). However, the magnitude of its bias relative to the other six estimators tends to decrease as τ^2 increases, and it becomes one of the best estimators for large τ^2 in terms of MSE and especially in terms of bias. The MVvc estimator improves substantially on the MV estimator in terms of bias unless τ^2 is large, has approximately the same MSEs as the MV and EB estimators, and ranks with the EB estimator as the best overall in terms of bias. Moreover, it is a computationally simple, non-iterative estimator. Finally, the EB estimator is also generally quite accurate, but is not as simple to compute and requires an iterative solution.

6. CONCLUSIONS

In this paper, we consider seven heterogeneity variance estimators for combining summary statistics from a series of independent studies, under the random effects model typically assumed in meta-analysis. We found that three estimators, namely MM, ML, and REML, have downward bias as estimators for the true heterogeneity variance. The underestimation can be severe if the

DOI: 10.1002/sim

heterogeneity variance is moderate to large, especially for the MM and ML estimators. On the other hand, these three estimators have smaller MSEs than the other four estimators in many of the cases considered. However, the smaller MSEs of these estimators are mainly due to their negative biases, as mentioned in Section 5. In particular, downward bias in a non-negativity-enforced estimator tends to reduce the variance of the estimator, which in turn may dominate the squared bias and yield a small MSE (see Reference [30]). Because of this systematic underestimation, the MM, ML, and REML estimators may only be recommended if the analyst is fairly certain that the heterogeneity in effect measures among the studies is relatively small.

The VC estimator generally has some upward bias, and generally has the largest MSE among the seven estimators. However, VC provides accurate estimation of τ^2 in terms of bias when a large number of studies are being combined (i.e. $k>30$), despite its relatively large MSE. In general, we found that the EB, MV, and MVvc estimators performed the best overall in terms of bias. These three estimators have similar MSEs over the range of cases considered. The EB and MVvc estimators have smaller (generally downward) bias than the MV estimator for small to moderate values of τ^2, and the MV estimator has the smallest (generally upward) bias of the three for relatively large τ^2.

In drawing practical conclusions regarding an estimator for the heterogeneity variance, we note that in practice, of course the analyst will not know the true heterogeneity variance. However, it may be reasonable to assume that it is possible to say whether the degree of heterogeneity is small, moderate, or large, based on the magnitude and direction of the effect estimates from the studies and perhaps the relative magnitude of Cochran's [16] statistic to its degrees of freedom. In the simulation study, the EB and MVvc estimators performed best in terms of bias when the heterogeneity was small to moderate. Thus, we recommend that one of these two estimators be used for estimating τ^2 when the degree of heterogeneity is expected to be small or moderate. In particular, the MVvc estimator may be preferred because of its computational advantage over the EB estimator. In view of the slight superiority of MV over MVvc in terms of bias when τ^2 is large, we suggest that the MV estimator may be chosen when the degree of heterogeneity is expected to be large, unless the number of studies is also quite large, in which case the VC estimator may be used. Finally, given that the log odds ratios are assumed to be approximately normal, and the similarity of our findings to some previously published studies with continuous outcome data (e.g. References [7, 8]), we would expect that these conclusions might also be approximately correct for normally distributed continuous outcome data, such as standardized mean differences. However, further study would be required to draw firm conclusions regarding continuous data, or to make recommendations for non-normally distributed effects or for meta-regression.

REFERENCES

1. Hedges LV. A random effects model for effect sizes. *Psychological Bulletin* 1983; **93**:388–395.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
3. Morris CN, Normand SL. Hierarchical models for combining information and for meta-analyses. In *Bayesian Statistics*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds), vol. 4. Oxford University Press: Oxford, 1992; 321–344 (with discussion).
4. Raudenbush SW. Random effects models. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: NY, 1994; 301–321.
5. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**: 1539–1558.
6. Raudenbush SW, Bryk AS. Empirical Bayes meta-analysis. *Journal of Educational Statistics* 1985; **10**:75–98.

DOI: 10.1002/sim

7. Malzahn U, Böhning D, Holling H. Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika* 2000; **87**:619–632.
8. Böhning D, Malzahn U, Dietz E *et al.* Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics* 2002; **3**:445–457.
9. Böhning D, Sarol J, Rattanasiri S, Viwatwongkasem C, Biggeri A. A comparison of non-iterative and iterative estimators of heterogeneity variance for the standardized mortality ratio. *Biostatistics* 2004; **5**:61–74.
10. Makambi KH. The effect of the heterogeneity variance estimator on some tests of treatment efficacy. *Journal of Biopharmaceutical Statistics* 2004; **14**:439–449.
11. Sidik K, Jonkman JN. Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society*, *Series C*: *Applied Statistics* 2005; **54**:367–384.
12. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
13. Friedman L. Estimators of random effects variance components in meta-analysis. *Journal of Educational and Behavioral Statistics* 2000; **25**:1–12.
14. Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Academic Press: California, 1985.
15. Cochran WG. Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society* 1937; **4**(Suppl.):102–118.
16. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; **10**:101–129.
17. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; **20**:825–840.
18. Hardy R, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
19. Morris CN. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association* 1983; **78**:47–55.
20. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
21. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 2003; **22**:2693–2710.
22. Sidik K, Jonkman JN. Robust variance estimation for random effects meta-analysis. *Computational Statistics and Data Analysis* 2006; **50**:3681–3701.
23. Sidik K, Jonkman JN. On constructing the confidence intervals for a standardized mean difference in meta-analysis. *Communications in Statistics–Simulation and Computation* 2003; **32**:1191–1203.
24. Rao PS, Kaplan J, Cochran WG. Estimators for the one-way random effects model with unequal error variances. *Journal of the American Statistical Association* 1981; **76**:89–97.
25. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 1977; **72**:320–339.
26. Fay IIIrd RE, Herriot RA. Estimates of income for small places: an application of James–Stein procedures to census data. *Journal of the American Statistical Association* 1979; **74**:269–277.
27. Memon MA, Cooper NJ, Memon B, Memon MI, Abrams KR. Meta-analysis of randomized clinical trials comparing open and laparoscopic inguinal hernia repair. *British Journal of Surgery* 2003; **90**:1479–1492.
28. Efron B. Empirical Bayes methods for combining likelihoods. *Journal of the American Statistical Association* 1996; **91**:538–550.
29. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. *Statistics in Medicine* 1999; **18**:643–654.
30. Swallow WH, Monahan JF. Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. *Technometrics* 1984; **26**:47–57.
31. Casella G, Berger RL. *Statistical Inference*. Duxbury Press: California, 1990.

DOI: 10.1002/sim