Journal of Educational and Behavioral Statistics
Fall 2005, Vol. 30, No. 3, pp. 261–293

Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model

Wolfgang Viechtbauer
University of Maastricht

The meta-analytic random effects model assumes that the variability in effect size estimates drawn from a set of studies can be decomposed into two parts: heterogeneity due to random population effects and sampling variance. In this context, the usual goal is to estimate the central tendency and the amount of heterogeneity in the population effect sizes. The amount of heterogeneity in a set of effect sizes has implications regarding the interpretation of the meta-analytic findings and often serves as an indicator for the presence of potential moderator variables. Five population heterogeneity estimators were compared in this article analytically and via Monte Carlo simulations with respect to their bias and efficiency.

Keywords: *heterogeneity estimation, meta-analysis, random-effects model*

1. Introduction

Although the roots of meta-analytic methodology reach back further in time than the conception of the term "meta-analysis" itself (Chalmers, Hedges, & Cooper, 2002; Glass, 1976), most of the research conducted in developing this technique has taken place during the last two decades. Compared to other statistical methods, we may, therefore, consider meta-analysis to be a relatively recent development. However, spurred by the information explosion in the scientific literature (Adair & Vohra, 2003), researchers have been eager to put forth various methods for aggregating the results from studies that provide commensurable evidence about a particular measurable effect (Bangert-Drowns, 1986). Much work still remains to be done in determining which of the proposed methods, if any, provides unbiased and efficient estimates of the effects, which researchers are trying to measure when conducting a meta-analysis.

Which parameters are going to be estimated depends on the statistical model adopted for the analysis. There seems to be a growing consensus that the random- and mixed-effects models should be preferred over the decidedly more simple fixed-effects model (Erez, Bloom, & Wells, 1996; Hunter & Schmidt, 2000; National

The author thanks Howard Wainer and two anonymous reviewers for their valuable comments. This article was written as part of the author's PhD dissertation at the University of Illinois, Urbana-Champaign.

Research Council, 1992). The meta-analytic random-effects model assumes that the variability in effect size estimates drawn from a set of studies can be decomposed into two parts: heterogeneity due to random population effects and sampling variance. In this context, the primary goal is to estimate the central tendency and the amount of heterogeneity in the population effect sizes. Various estimators for the amount of heterogeneity in the population effect sizes have been suggested, but, aside from a few notable exceptions (e.g., Friedman, 2000), little work has been done to compare and contrast the statistical properties of these estimators directly. The present article attempts to expand on what is known about these variance estimators and in particular, to determine whether one should be preferred over the others.

The general outline of this article is as follows. In the second section, I briefly outline the random-effects model in meta-analysis and emphasize the need for a careful analysis of the various estimators for the amount of heterogeneity in the population effect sizes. Next, formulas for five variance estimators, their bias, and sampling variances are given. In particular, the estimators include one suggested by Hunter and Schmidt (1990) from the validity generalization literature, one proposed by Hedges (1983, 1989), an estimator by DerSimonian and Laird (1986), the maximum likelihood estimator, and the restricted maximum likelihood estimator. Two illustrative examples in section 4 demonstrate that the estimators can yield noticeably divergent results. Analytic comparisons between the estimators in terms of their sampling variance and mean squared error are given in section 5. Furthermore, the bias in the estimators due to truncation of negative estimates is discussed. However, the analytic comparisons require some restrictive assumptions, which might not hold in practice. The sixth section, therefore, presents the results from Monte Carlo simulations that were conducted to complement the analytic results. Some final remarks conclude the article.

2. The Random-Effects Model in Meta-Analysis

Assume that $i = 1, \ldots, k$ independent effect size estimates, $ES_1, \ldots, ES_k$, have been derived from a set of studies. The random-effects model assumes that each effect size estimate can be decomposed into two variance components by a two-stage hierarchical process. We assume that θ_i, the population effect size in the ith study, is drawn from a distribution of population effect sizes with mean μ_θ and variance σ^2_θ. The size of σ^2_θ indicates the degree of heterogeneity in the population effect sizes, whereas μ_θ describes their central tendency. Let τ_i represent an error term by which the population effect size in the ith study differs from μ_θ. Furthermore, let ϵ_i represent an error term by which ES_i differs from θ_i. Then we can write the random-effects model as

$$ES_i = \theta_i + \epsilon_i = \mu_\theta + \tau_i + \epsilon_i. \tag{1}$$

The assumptions of the meta-analytic random-effects model are: (a) $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$, (b) $\tau_i \overset{iid}{\sim} N(0, \sigma^2_\theta)$, (c) $\text{Cov}[\epsilon_i, \epsilon_j] = 0$ for $i \neq j$, and (d) $\text{Cov}[\epsilon_i, \tau_j] = 0$ for all i and j. It follows from these assumptions that $ES_i \sim N(\mu_\theta, \sigma^2_\theta + \sigma^2_{\epsilon_i})$. The goal in this context is to estimate μ_θ and σ_θ^2 in an optimal manner. Criteria for optimality include the (un)biasedness and efficiency, and more generally, the mean squared error (MSE) of the estimators for these parameters.

The Cramer–Rao lower bounds of unbiased estimators for μ_θ and σ_θ^2 are

$$\left(\sum_{i=1}^{k} \frac{1}{\sigma_\theta^2 + \sigma_{\epsilon_i}^2} \right)^{-1} \tag{2}$$

and

$$2\left[\sum_{i=1}^{k} \frac{1}{\left(\sigma_\theta^2 + \sigma_{\epsilon_i}^2\right)^2} \right]^{-1}, \tag{3}$$

respectively. Estimators for μ_θ are usually of the form

$$\overline{ES} = \frac{\sum_{i=1}^{k} m_i ES_i}{\sum_{i=1}^{k} m_i}, \tag{4}$$

where m_i are weights assigned to each of the effect sizes. For $m_i = 1/(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$, we note that $\mathrm{Var}[\overline{ES}]$ achieves the Cramer–Rao lower bound and, therefore, is the uniformly minimum variance unbiased estimator (UMVUE) of μ_θ (Viechtbauer, 2002).

However, in practice, estimates must be substituted into m_i for the unknown parameters σ_θ^2 and $\sigma_{\epsilon_i}^2$. For many commonly used effect size measures, it is possible to calculate unbiased estimates of the $\sigma_{\epsilon_i}^2$ values from little more than the observed effect sizes and knowledge about the sample sizes on which the observed effect sizes are based. The usual practice is to treat such estimates as the true $\sigma_{\epsilon_i}^2$ values and to ignore any associated sampling error. Furthermore, as shown below, an unbiased estimate of σ_θ^2 can also be obtained without too much difficulty. Again, the sampling variance of such an estimate is usually ignored.

The consequence of calculating the m_i values using estimates of $\sigma_{\epsilon_i}^2$ and σ_θ^2 is as follows. Let $\hat{\sigma}_\theta^2$ and $\hat{\sigma}_{\epsilon_i}^2$ be unbiased estimates of the corresponding parameters. After extending the results by Li, Shi, and Roth (1994), it can be shown that

$$E\left[\left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}_\theta^2 + \hat{\sigma}_{\epsilon_i}^2} \right)^{-1} \right] \leq \left(\sum_{i=1}^{k} \frac{1}{\sigma_\theta^2 + \sigma_{\epsilon_i}^2} \right)^{-1}. \tag{5}$$

In other words, substituting unbiased estimates of σ_θ^2 and $\sigma_{\epsilon_i}^2$ into m_i results in an estimate of the sampling variance of $\overline{ES}$ that is negatively biased. As a consequence of this negative bias, the sampling variance of $\overline{ES}$ will be underestimated on average, and researchers will attribute unwarranted precision to their estimate of μ_θ.

Using estimates of σ_θ^2 and $\sigma_{\epsilon_i}^2$ with small sampling variance should reduce the extent of the negative bias. Therefore, if one has the option of choosing among several reasonable estimators of σ_θ^2 and $\sigma_{\epsilon_i}^2$, it would be valuable to know which of them have small sampling variance. The current article focuses in particular on estimators of σ_θ^2. Future work could address the efficiency of $\sigma_{\epsilon_i}^2$ estimates for various effect size measures.

A second point to consider is the fact that the sampling variance of $\overline{ES}$ is an increasing function of σ_θ^2. Consequently, the sampling variance of $\overline{ES}$ will be underestimated on average when using a negatively biased estimator of σ_θ^2. On the other hand, positive bias in σ_θ^2 will lead researchers to understate the accuracy of their estimate of μ_θ. Also, the parameter σ_θ^2 is of interest not only for calculating $\overline{ES}$ and its sampling variance, but also because it directly indicates any heterogeneity in the effect sizes that cannot be accounted for by sampling error in the effect size estimates alone. In fact, $\sigma_\theta^2 > 0$ can result from (a) heterogeneous population effect sizes as described by the random-effects model, (b) differences in the population effect sizes due to the presence of moderator variables, or (c) a combination of random population effect sizes and moderator effects. Proper handling of the latter two cases requires modeling of moderator variables effects and leads, respectively, to the fixed-effects model with moderators and the mixed-effects model (Viechtbauer, 2002). Modeling of moderator effects is not discussed in the present article, but the interested reader should consult, for example, Hedges (1994), Overton (1998), and Raudenbush (1994).

Suffice it to say, any heterogeneity in the population effect sizes, whether caused by moderator effects or random variation, will be reflected in estimates of σ_θ^2 being greater than zero. On the other hand, estimates of σ_θ^2 equal to zero are usually interpreted as indicating the absence of moderators and random variation within the population effects, in which case, the random-effects model reduces to the simple fixed-effects model (estimates of σ_θ^2 smaller than zero are usually truncated to zero). In that case, μ_θ no longer represents the central tendency of the population effects, but rather a fixed population effect size θ that all k effect sizes estimate jointly. Therefore, choice and interpretation of the appropriate model is influenced to a large extent by the estimate of σ_θ^2. In fact, Hunter and Schmidt (1990) advise against the use of statistical hypothesis tests to determine the homogeneity of effect sizes (which would rule out moderator effects and population heterogeneity) and instead favor a critical examination of σ_θ^2 estimates in terms of their magnitude.

Consequently, the effects of bias in estimates of σ_θ^2 can either lead researchers to search for and discover moderators that do not really exist or lead researchers to ignore the presence of actual moderator effects. This is of particular concern now because researchers are becoming increasingly aware of the fact that detection and estimation of moderator effects is often the most valuable contribution of meta-analysis to the research domains in which it is applied. Because estimates of σ_θ^2 play such a crucial role in meta-analysis, the bias and efficiency of σ_θ^2 estimators must be considered in this context as well. More confidence can be placed in (unbiased) estimators of σ_θ^2 with smaller sampling variance because they will reflect, on average, more accurately the true value of the amount of heterogeneity in the population effect sizes.

In summary, then, it is worthwhile to examine the statistical properties of the various population heterogeneity estimators because (a) using efficient estimators of σ_θ^2 reduces the possibility that we overestimate the precision of $\overline{ES}$, (b) using biased estimators of σ_θ^2 leads to over or underestimation of the precision of $\overline{ES}$, and (c) estimates of σ_θ^2 are relevant for model choice and moderator analysis. Finally, as is demonstrated in section 4 with two examples, the estimators can yield divergent or even conflicting results.

3. Heterogeneity Estimators in the Random-Effects Model

In this section, five population heterogeneity estimators are introduced, including a commonly used estimator in validity generalization research, which can be attributed to Hunter and Schmidt (1990), an estimator proposed by Hedges (1983, 1989), an estimator by DerSimonian and Laird (1986), and two estimators based on maximum likelihood estimation, the regular maximum likelihood estimator and the restricted maximum likelihood estimator.

3.1. Hunter–Schmidt Estimator

The Hunter–Schmidt (HS) estimator is given by

$$\hat{\sigma}_\theta^{2\,(HS)} = \frac{\sum_{i=1}^{k} w_i\left(ES_i - \overline{ES}\right)^2}{\sum_{i=1}^{k} w_i} - \frac{\sum_{i=1}^{k} w_i \sigma_{\epsilon_i}^2}{\sum_{i=1}^{k} w_i}, \tag{6}$$

where the w_i values are weights that do not have to coincide with the m_i weights used to calculate $\overline{ES}$ as defined in Equation 4. However, setting $w_i = m_i$ is the usual practice and will be assumed from now on. Because σ_θ^2 is unknown, it is not possible to use the optimal $w_i = 1/(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$ weights in Equation 6. Instead, w_i and m_i are initially set equal to (a) the fixed-effects model weights, namely, $1/\sigma_{\epsilon_i}^2$; (b) the sample size on which the ith effect size is based (as an approximation to $1/\sigma_{\epsilon_i}^2$); or (c) unity, which would provide an unweighted estimate of σ_θ^2. Once $\hat{\sigma}_\theta^{2(HS)}$ has been obtained, a final estimate of μ_θ is calculated using the approximately optimal weights for $\overline{ES}$, namely $w_i = 1/(\hat{\sigma}_\theta^{2(HS)} + \hat{\sigma}_{\epsilon_i}^2)$.

However, the estimator given by Equation 6 has been shown to be negatively biased (Viechtbauer, 2002). The bias is equal to

$$\text{Bias}\left[\hat{\sigma}_\theta^{2\,(HS)}\right] = -\frac{\sum_{i=1}^{k} w_i^2\left(\sigma_\theta^2 + \sigma_{\epsilon_i}^2\right)}{\left(\sum_{i=1}^{k} w_i\right)^2}, \tag{7}$$

which increases for larger σ_θ^2 and tends to decrease as $k \to \infty$ for bounded values of $(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$. For homogeneous sampling variances and weights, the bias will be equal to $-(\sigma_\theta^2 + \sigma_\epsilon^2)/k$, where σ_ϵ^2 denotes the common sampling variance of the k effect sizes. When k is small, this could potentially lead to substantial underestimation of the population heterogeneity.

The sampling variance of $\hat{\sigma}_\theta^{2(HS)}$ is obtained by writing $\Sigma w_i(ES_i - \overline{ES})^2$ as a quadratic form (Viechtbauer, 2002) and then using a known theorem in linear algebra (Searle, 1971, p. 55) to derive the variance thereof. After some tedious algebra, we find that

$$\mathrm{Var}\left[\hat{\sigma}_\theta^{2(HS)}\right] = \frac{2}{\left(\sum_{i=1}^k w_i\right)^2} \left[\sum_{i=1}^k w_i^2 v_i^2 - 2\frac{\sum_{i=1}^k w_i^3 v_i^2}{\sum_{i=1}^k w_i} + \frac{\left(\sum_{i=1}^k w_i^2 v_i\right)^2}{\left(\sum_{i=1}^k w_i\right)^2} \right], \tag{8}$$

where $v_i = (\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$. The bias and the variance of $\hat{\sigma}_\theta^{2(HS)}$ go to zero as $k \to \infty$ for bounded values of v_i and, consequently we can conclude that the HS estimator is consistent. For the common case of $w_i = 1/\sigma_{\epsilon_i}^2$, we can write Equation 8 as

$$\frac{2}{\left(\sum_{i=1}^k w_i\right)^4} \left\{ \sum_{j=1}^k w_j^2 \left[\left(\sum_{h \neq j}^k w_h \right)^2 + \sum_{h \neq j}^k w_h^2 \right] \right\} \sigma_\theta^4 + \frac{4c}{\left(\sum_{i=1}^k w_i\right)^2} \sigma_\theta^2 + \frac{2(k-1)}{\left(\sum_{i=1}^k w_i\right)^2}, \tag{9}$$

where

$$c = \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}. \tag{10}$$

Because using inverse sampling variance weights, or an approximation thereof, is the most common practice when calculating the HS estimator, any subsequent reference to this estimator assumes the use of such weights.

3.2. Hedges Estimator

Let $\overline{ES}$ be the unweighted average of the ES_i values. Then an unbiased estimate of σ_θ^2 is given by

$$\hat{\sigma}_\theta^{2(HE)} = \frac{\sum_{i=1}^k \left(ES_i - \overline{ES}\right)^2}{k-1} - \frac{1}{k} \sum_{i=1}^k \sigma_{\epsilon_i}^2. \tag{11}$$

The Hedges (HE) estimator is unbiased not only when the exact sampling variances are known, but also when substituting unbiased estimates for the $\sigma_{\epsilon_i}^2$ values in Equation 11.

The sampling variance of the HE estimator is equal to

$$\mathrm{Var}\left[\hat{\sigma}_{\theta}^{2(HE)}\right]=\frac{2}{(k-1)^2}\left[\sum_{i=1}^{k} v_i^2-2\frac{\sum_{i=1}^{k} v_i^2}{k}+\frac{\left(\sum_{i=1}^{k} v_i\right)^2}{k^2}\right]. \quad (12)$$

Again, the variance of the estimator decreases to zero as $k \to \infty$ and, therefore, $\hat{\sigma}_{\theta}^{2(HE)}$ is consistent. Friedman (2000) showed that Equation 12 can be written as

$$\frac{2}{(k-1)}\sigma_{\theta}^4+\left(\frac{4}{k(k-1)}\sum_{i=1}^{k}\sigma_{\epsilon_i}^2\right)\sigma_{\theta}^2+\frac{2}{k^2}\left(\sum_{i=1}^{k}\sigma_{\epsilon_i}^4+\frac{1}{(k-1)^2}\sum_{i=1}^{k}\sum_{j\neq i}^{k}\sigma_{\epsilon_i}^2\sigma_{\epsilon_j}^2\right). \quad (13)$$

3.3. DerSimonian–Laird Estimator

DerSimonian and Laird (1986) suggested the estimator

$$\hat{\sigma}_{\theta}^{2(DL)}=\frac{\sum_{i=1}^{k} w_i\left(ES_i-\overline{ES}\right)^2-(k-1)}{c}, \quad (14)$$

with c as defined in Equation 10 and $w_i = m_i = 1/\sigma_{\epsilon_i}^2$. The DerSimonian–Laird (DL) estimator is unbiased under the assumption that the $\sigma_{\epsilon_i}^2$ values are known.

The sampling variance of $\hat{\sigma}_{\theta}^{2(DL)}$ is equal to

$$\mathrm{Var}\left[\hat{\sigma}_{\theta}^{2(DL)}\right]=\frac{2}{c^2}\left[\sum_{i=1}^{k} w_i^2 v_i^2-2\frac{\sum_{i=1}^{k} w_i^3 v_i^2}{\sum_{i=1}^{k} w_i}+\frac{\left(\sum_{i=1}^{k} w_i^2 v_i\right)^2}{\left(\sum_{i=1}^{k} w_i\right)^2}\right]. \quad (15)$$

The DL estimator is also consistent. Friedman (2000) showed that Equation 15 can be written as

$$\frac{2}{\left[\left(\sum_{i=1}^{k} w_i\right)^2-\sum_{i=1}^{k} w_i^2\right]^2}\left\{\sum_{j=1}^{k} w_j^2\left[\left(\sum_{h\neq j}^{k} w_h\right)^2+\sum_{h\neq j}^{k} w_h^2\right]\right\}\sigma_{\theta}^4+\frac{4}{c}\sigma_{\theta}^2+\frac{2(k-1)}{c}. \quad (16)$$

3.4. Maximum-Likelihood Estimator

The random-effects model given by Equation 1 is just a special case of the general linear mixed-effects model (GLMM) of the form

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{\gamma}+\boldsymbol{e},$$

where $\boldsymbol{y}$ is a $(k \times 1)$ vector of random variables, $\boldsymbol{X}$ is a $(k \times p)$ matrix of known constants for the $(p \times 1)$ fixed effects parameter vector $\boldsymbol{\beta}$, $\boldsymbol{Z}$ is the $(k \times q)$ design matrix for the $(q \times 1)$ random effects parameter vector $\boldsymbol{\gamma}$, and $\boldsymbol{e}$ is a $(k \times 1)$ vector of random error terms. We assume $\mathrm{E}[\boldsymbol{\gamma}] = \mathbf{0}$, $\mathrm{E}[\boldsymbol{e}] = \mathbf{0}$, and $\mathrm{Cov}[\boldsymbol{\gamma}, \boldsymbol{e}] = \mathbf{0}$. Define $\boldsymbol{D}$ as the $(q \times q)$ covariance matrix of the random effects parameters in $\boldsymbol{\gamma}$ and $\boldsymbol{R}$ as the $(k \times k)$ covariance matrix of $\boldsymbol{e}$. Then $\boldsymbol{V}$, the $(k \times k)$ covariance matrix of $\boldsymbol{y}$, is equal to $\boldsymbol{ZDZ}' + \boldsymbol{R}$. After assuming normality of the random terms in the model, we obtain $\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V})$. Denoting the variance components in $\boldsymbol{V}$ by the vector $\boldsymbol{\sigma}^2$, we can write the log-likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$ as

$$\ln L\left(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}\right) = -\frac{1}{2} \ln |\boldsymbol{V}| - \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{Xb}\right)' \boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{Xb}\right),$$

leaving out the additive constant.

For the meta-analytic random-effects model, $\boldsymbol{y}$ consists of the k effect size estimates, $\boldsymbol{X}$ is a $(k \times 1)$ vector composed entirely of 1s, $\boldsymbol{\beta}$ includes only the grand mean μ_θ, $\boldsymbol{Z}$ is the $(k \times k)$ identity matrix, $\boldsymbol{\gamma}$ is comprised of the τ_i values at the population level, and $\boldsymbol{e}$ includes the random error terms, $\epsilon_1, \ldots, \epsilon_k$. Then $\boldsymbol{V}$ is diagonal with $v_i = (\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$ and $\boldsymbol{y} \sim N(\mathbf{1}\mu_\theta, \boldsymbol{V})$.

Treating the sampling variances as known, the log-likelihood function of μ_θ and σ_θ^2 therefore simplifies to

$$\ln L\left(\mu_\theta, \sigma_\theta^2 | \boldsymbol{ES}\right) = -\frac{1}{2}\sum_{i=1}^{k} \ln\left(\sigma_\theta^2 + \sigma_{\epsilon_i}^2\right) - \frac{1}{2}\sum_{i=1}^{k} \frac{(ES_i - \mu_\theta)^2}{\sigma_\theta^2 + \sigma_{\epsilon_i}^2}.$$

Setting partial derivatives with respect to μ_θ and σ_θ^2 equal to zero and solving the likelihood equations for the two parameter to be estimated, we obtain Equation 4 as $\hat{\mu}_\theta^{(ML)}$ and

$$\hat{\sigma}_\theta^{2\,(ML)} = \frac{\sum_{i=1}^{k} w_i^2\left[\left(ES_i - \hat{\mu}_\theta^{(ML)}\right)^2 - \sigma_{\epsilon_i}^2\right]}{\sum_{i=1}^{k} w_i^2}, \tag{17}$$

with $w_i = m_i = 1/(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$.

Solutions to these equations are obtained by iterating between $\hat{\mu}_\theta^{(ML)}$ and $\hat{\sigma}_\theta^{2(ML)}$, starting either with an initial guess for $\hat{\sigma}_\theta^{2(ML)}$ (such as given by any of the noniterative methods) or setting $\hat{\sigma}_\theta^{2(ML)} = 0$ (Erez, Bloom, & Wells, 1996; National Research Council, 1992). This process continues until the parameter estimates do not change from one iteration to the next. Convergence usually occurs rapidly within less than ten iterations (Erez et al., 1996).

Occasionally, choice of particular effect size measures imposes constraints on the parameter space of μ_θ. Moreover, σ_θ^2 is constrained to be non-negative in all cases. When the solutions converge to values outside the parameter space, then one should check whether a maximum of the log-likelihood function occurs at the boundaries of the parameter space. Solutions inside the parameter space should be evaluated at $\hat{\mu}_\theta^{(ML)}$ and $\hat{\sigma}_\theta^{2(ML)}$ via the Hessian matrix, given by

$$\boldsymbol{H} = \begin{bmatrix} -\sum_{i=1}^{k} w_i & -\sum_{i=1}^{k} w_i^2 (ES_i - \mu_\theta) \\ -\sum_{i=1}^{k} w_i^2 (ES_i - \mu_\theta) & \frac{1}{2}\sum_{i=1}^{k} w_i^2 - \sum_{i=1}^{k} w_i^3 (ES_i - \mu_\theta)^2 \end{bmatrix},$$

to ensure that the determinant of $\boldsymbol{H}$ is positive.

3.5. Restricted Maximum Likelihood Estimator

The maximum likelihood estimator of σ_θ^2 tends to underestimate the population heterogeneity in finite samples by failing to account for the fact that μ_θ in Equation 17 is also estimated from the data. In fact, maximum likelihood estimates of variance components are known to be negatively biased in many cases (Patterson & Thompson, 1974; Corbeil & Searle, 1976). The restricted maximum likelihood (REML) estimator compensates for this underestimation by using a linear combination of the $\boldsymbol{y}$ vector, so that the transformed data are free of the fixed effects in $\boldsymbol{\beta}$. Specifically, let $\boldsymbol{M}$ be equal to $(k - 1)$ linearly independent columns of $\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}$. Then $\boldsymbol{M'}\,\boldsymbol{y}$ is independent of $\boldsymbol{\beta}$ in the sense that $\boldsymbol{M'}\,\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{M'VM})$. In fact, we can take any matrix $\boldsymbol{M}$ of full rank, as long as $\boldsymbol{M'}\,\boldsymbol{X} = 0$. Then the log-likelihood function to be maximized is given by

$$\ln L(\boldsymbol{\sigma}^2|\boldsymbol{y}) = -\frac{1}{2}\ln|\boldsymbol{V}| - \frac{1}{2}\ln|\boldsymbol{X'V}^{-1}\boldsymbol{X}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}),$$

where $\tilde{\boldsymbol{\beta}}$ is a maximum likelihood solution of $\boldsymbol{\beta}$ for fixed $\boldsymbol{\sigma}^2$ (Harville, 1977). This simplifies to

$$\ln L(\sigma_\theta^2|\boldsymbol{ES}) = -\frac{1}{2}\sum_{i=1}^{k}\ln(\sigma_\theta^2 + \sigma_{\epsilon_i}^2) - \frac{1}{2}\ln\sum_{i=1}^{k}\frac{1}{(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)} - \frac{1}{2}\sum_{i=1}^{k}\frac{(ES_i - \hat{\mu}_\theta^{(ML)})^2}{(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)}.$$

for the meta-analytic random-effects model. The REML estimator of σ_θ^2 is then given by

$$\hat{\sigma}_\theta^{2\,(REML)} = \frac{\sum_{i=1}^{k} w_i^2\left[(ES_i - \hat{\mu}_\theta^{(ML)})^2 - \sigma_{\epsilon_i}^2\right]}{\sum_{i=1}^{k} w_i^2} + \frac{1}{\sum_{i=1}^{k} w_i} \tag{18}$$

and is obtained in the same iterative manner as described for the regular maximum likelihood estimator.

Restricted maximum likelihood estimation does not yield estimates of the fixed effects. However, as suggested by Searle, Casella, and McCulloch (1992), it seems reasonable to evaluate the regular maximum likelihood estimates of the fixed effects parameters using the REML variance component estimates. In fact, finding a solution for Equation 18 requires that we obtain exactly this type of estimate as part of the iteration procedure, which we now define as the REML estimator of μ_θ.

The restricted maximum likelihood estimator has frequently been defined as

$$\hat{\sigma}_\theta^{2\,(REML)} = \frac{\sum_{i=1}^{k} w_i^2\left[(k/(k-1))\left(ES_i - \hat{\mu}_\theta^{(ML)}\right)^2 - \sigma_{\epsilon_i}^2\right]}{\sum_{i=1}^{k} w_i^2} \tag{19}$$

in the literature (Berkey, Hoaglin, Mosteller, & Colditz, 1995; Normand, 1999; Thompson & Sharp, 1999). It appears that Equation 19 originated with Morris (1983), who suggested it as an "approximate" REML estimate (p. 53). When the sampling variance are homogeneous (i.e., $\sigma_{\epsilon_i}^2 = \sigma_\epsilon^2, \forall i = 1, \ldots, k$), then the approximation given by Equation 19 and the exact REML estimate given by Equation 18 are equal to each other, but this is rare in practice, and, therefore, Equation 18 should be preferred.

According to the usual principles of maximum likelihood estimation, the regular MLEs of μ_θ and σ_θ^2 are expected to be consistent and asymptotically fully efficient. From the information matrix of μ_θ and σ_θ^2, we can obtain the asymptotic sampling variances. For the maximum likelihood estimates, they are given by

$$\text{Var}_\infty\left[\hat{\mu}_\theta^{(ML)}\right] = \left(\sum_{i=1}^{k} w_i\right)^{-1}$$

and

$$\text{Var}_\infty\left[\hat{\sigma}_\theta^{2\,(ML)}\right] = 2\left(\sum_{i=1}^{k} w_i^2\right)^{-1}.$$

Not surprisingly, these are of the same form as the Cramer–Rao lower bounds for μ_θ and σ_θ^2 (see Equations 2 and 3). On the other hand, the asymptotic sampling variance of $\hat{\sigma}_\theta^{2\,(REML)}$ is

$$\text{Var}_\infty\left[\hat{\sigma}_\theta^{2(REML)}\right] = 2\left[\sum_{i=1}^{k} w_i^2 - 2\frac{\sum_{i=1}^{k} w_i^3}{\sum_{i=1}^{k} w_i} + \frac{\left(\sum_{i=1}^{k} w_i^2\right)^2}{\left(\sum_{i=1}^{k} w_i\right)^2}\right]^{-1},$$

which can be shown to be greater than $\text{Var}_\infty[\hat{\sigma}_\theta^{2\,(ML)}]$. Consequently, the restricted maximum likelihood estimator is less efficient than the regular maximum likelihood estimator in finite samples. Estimates of these sampling variances are obtained by evaluating the equations with the corresponding estimates of σ_θ^2, namely, by setting w_i to $1/(\hat{\sigma}_\theta^{2\,(ML)} + \hat{\sigma}_{\epsilon_i}^2)$ and $1/(\hat{\sigma}_\theta^{2\,(REML)} + \hat{\sigma}_{\epsilon_i}^2)$, respectively.

Occasionally, it is possible that the ML or REML estimator will converge to $-\sigma_{\epsilon_i}^2$ for one of the $\sigma_{\epsilon_i}^2$ values. Then $w_i = 1/(\sigma_\theta^2 + \sigma_{\epsilon_i}^2)$ is undefined, and the iteration procedure breaks down. Such negative variance estimates lie outside the parameter space and are usually truncated to zero. Also, occasionally, the iterative estimation procedures will not converge and instead continue cycling between several values of $\hat{\sigma}_\theta^2$. This seems more likely to happen when k is small. Usually, the cycle is confined to two values, but higher cycles also occur. In this case, one can employ a direct maximization technique without derivatives, such as the algorithms suggested by Nelder and Mead (1965) and by Byrd, Lu, Nocedal, and Zhu (1995). Constraints on the parameter space of μ_θ can also be incorporated into these optimization procedures.

4. Examples

The two examples in the present section help to illustrate that the five estimators can provide noticeably divergent or conflicting estimates. The first data set, given in Table 1, provides the results for $k = 10$ studies that examined the effectiveness of open versus traditional education programs on student creativity (Hedges & Olkin, 1985, p. 25). The table lists the effect size ES_i, the estimated sampling variance $\hat{\sigma}_{\epsilon_i}^2$, and the inverse sampling variance weight $w_i = 1/\hat{\sigma}_{\epsilon_i}^2$ for each study.

When using the HS estimator, σ_θ^2 is estimated to be equal to 0.23, while the HE and DL estimators yield the values 0.15 and 0.28, respectively. These calculations can be easily verified with a pocket calculator. The ML estimate converges quickly to 0.20 within seven iterations when the stopping criterion is taken to be a change in $\hat{\sigma}_\theta^{2\,(ML)}$ less than 10^{-5} from one iteration to the next. On the other hand, the REML estimator converges within five iterations to the value 0.22. Clearly, these five estimates are far from being unanimous, although they all indicate substantial heterogeneity in the effect sizes over and beyond what one would expect based on sampling error alone.

TABLE 1
Results for 10 Studies of the Effectiveness of Open Versus Traditional Education on Student Creativity

Study	Effect Size (ES_i)	Variance ($\hat{\sigma}_{\epsilon_i}^2$)	Weight ($w_i = 1/\hat{\sigma}_{\epsilon_i}^2$)
1	−0.581	0.023	43.478
2	0.530	0.052	19.231
3	0.771	0.060	16.667
4	1.031	0.115	8.696
5	0.553	0.095	10.526
6	0.295	0.203	4.926
7	0.078	0.200	5.000
8	0.573	0.211	4.739
9	−0.176	0.051	19.608
10	−0.232	0.040	25.000

Source: Hedges and Olkin (1985), p. 25.

TABLE 2
Results for 18 Studies of the Effectiveness of Open Versus Traditional Education on Student Self-Concept

Study	Effect Size (ES_i)	Variance ($\hat{\sigma}^2_{\epsilon_i}$)	Weight ($w_i = 1/\hat{\sigma}^2_{\epsilon_i}$)
~~1~~	~~–0.581~~	~~0.023~~	~~43.478~~
1	0.100	0.016	62.500
2	–0.162	0.015	66.667
3	–0.090	0.050	20.000
4	–0.049	0.050	20.000
5	–0.046	0.032	31.250
6	–0.010	0.052	19.231
7	–0.431	0.036	27.778
8	–0.261	0.024	41.667
9	0.134	0.034	29.412
10	0.019	0.033	30.303
11	0.175	0.031	32.258
12	0.056	0.034	29.412
13	0.045	0.039	25.641
14	0.103	0.167	5.988
15	0.121	0.134	7.463
16	–0.482	0.096	10.417
17	0.290	0.016	62.500
18	0.342	0.035	28.571

Source: Hedges and Olkin (1985), p. 25.

In the first example, there is disagreement among the estimators with respect to the extent of the heterogeneity. However, even more problematic are cases where the amount of population heterogeneity is estimated to be equal to zero (or negative) with some of the estimators and positive with others. Consider Table 2, which provides results for $k = 18$ studies comparing open versus traditional education using student self-concept as the outcome variable (Hedges & Olkin, 1985, p. 25).

For the most part, the estimators yield similar results, namely 0.010 for the HS estimator, 0.013 when using the DL or ML estimator, and 0.016 with the REML estimator. The ML and REML estimators again converge quickly in less than five iterations. Each of these four values indicates a modest amount of population heterogeneity. However, when using the HE estimator, we obtain an estimate of σ^2_θ equal to –0.002, which would indicate the absence of any variability in the population effects.

5. Analytic Comparisons of the Estimators

The fact that the various estimators can lead to divergent or conflicting results immediately raises the question whether one should be preferred over the others. In the remainder of the article, I attempt to address this particular question. However, before giving some general results about the estimators, it is instructive to consider three special cases. In particular, I examine the behavior of the estimators

(a) when the sampling variances are homogeneous, (b) when the sample sizes on which the effect sizes are based become very large, and (c) when the population heterogeneity is zero.

5.1. Homogeneous Sampling Variances

When the sampling variances of the k effect sizes are homogeneous, we obtain the notable result that

$$\hat{\sigma}_\theta^{2\,(HS)} = \hat{\sigma}_\theta^{2\,(ML)} = \frac{\sum_{i=1}^{k}\left(ES_i - \overline{ES}\right)^2}{k} - \sigma_\epsilon^2$$

and

$$\hat{\sigma}_\theta^{2\,(HE)} = \hat{\sigma}_\theta^{2\,(DL)} = \hat{\sigma}_\theta^{2\,(REML)} = \frac{\sum_{i=1}^{k}\left(ES_i - \overline{ES}\right)^2}{k-1} - \sigma_\epsilon^2,$$

where $\overline{ES}$ is the unweighted average of the effect size estimates and σ_ϵ^2 denotes the common sampling variance of the effect sizes.

Therefore, $\hat{\sigma}_\theta^{2\,(ML)}$, like the HS estimator, is negatively biased for finite k. On the other hand, $\hat{\sigma}_\theta^{2\,(REML)}$ is unbiased. These results do not come as a surprise, as maximum likelihood estimates of variance components often exhibit negative bias (Corbeil & Searle, 1976; Patterson & Thompson, 1974) and in fact, restricted maximum-likelihood estimation was suggested as a means for reducing or eliminating the bias in the ML estimates.

The sampling variances of the estimators simplify to

$$\mathrm{Var}\left[\hat{\sigma}_\theta^{2\,(HS/ML)}\right] = \frac{2(k-1)}{k^2}\left(\sigma_\theta^2 + \sigma_\epsilon^2\right)^2 \tag{20}$$

and

$$\mathrm{Var}\left[\hat{\sigma}_\theta^{2\,(HE/DL/REML)}\right] = \frac{2}{(k-1)}\left(\sigma_\theta^2 + \sigma_\epsilon^2\right)^2. \tag{21}$$

Moreover, the Cramer–Rao lower bound of unbiased σ_θ^2 estimators is now equal to $(2/k)(\sigma_\theta^2 + \sigma_\epsilon^2)^2$, and we can conclude that: (a) the HS and ML estimators approach the Cramer–Rao lower bound from below, (b) the HE, DL, and REML estimators approach the bound from above; and (c) the estimators are asymptotically fully efficient. Note that $\hat{\sigma}_\theta^{2\,(HS)}$ and $\hat{\sigma}_\theta^{2\,(ML)}$ might actually have sampling variances that fall below the Cramer–Rao lower bound, which is a consequence of the bias in these estimators. Finally, from commonly known results (Lehmann & Casella, 1998), we can infer that the HE, DL, and REML estimators are the UMVUE of σ_θ^2 when the sampling variances are known and homogeneous.

The MSE of $\hat{\sigma}_\theta^{2(HS/ML)}$ is now given by

$$MSE\left[\hat{\sigma}_\theta^{2(HS/ML)}\right] = \frac{2k-1}{k^2}\left(\sigma_\theta^2 + \sigma_\epsilon^2\right)^2,$$

while the MSE of $\hat{\sigma}_\theta^{2(HE/DL/REML)}$ is equal to Equation 21, as these three estimators are unbiased. The relative efficiency of the HS and ML estimators compared to the HE, DL, and REML estimators is given by $((k-1)/k)^2$, which approaches 1 as k increases. However, as Table 3 shows, for small to moderately large values of k, differences in the relative efficiency are quite noticeable. The relative MSE is given by $(2k-1)(k-1)/(2k^2)$, which also approaches 1 for increasing k. As Table 3 demonstrates, the HS and ML estimators have lower MSE than the three unbiased estimators. In other words, these results indicate that

$$\text{Var}\left[\hat{\sigma}_\theta^{2(HS/ML)}\right] < \text{Var}\left[\hat{\sigma}_\theta^{2(HE/DL/REML)}\right] \tag{22}$$

and

$$MSE\left[\hat{\sigma}_\theta^{2(HS/ML)}\right] < MSE\left[\hat{\sigma}_\theta^{2(HE/DL/REML)}\right]. \tag{23}$$

Finally, it should be mentioned that we can again infer from commonly known results (Lehmann & Casella, 1998) that the HS and ML estimators do not minimize the MSE. Instead, the estimator $\hat{\sigma}_\theta^2 = (1/(k+1)) \sum (ES_i - \overline{ES})^2 - \sigma_\epsilon^2$ dominates the HS and ML estimators in terms of MSE for all values of σ_θ^2 when the sampling variances are known and homogeneous.

5.2. Infinite Sample Sizes

If we keep k fixed and let the sample sizes on which the ES_i values are based go to infinity, then $\sigma_{\epsilon_i}^2 \to 0$ for $i = 1, \ldots, k$. In other words, if we imagine each study having infinite sample size, then no sampling error remains in the effect size estimates. In that case, $\hat{\sigma}_\theta^{2(HE)}$ and $\hat{\sigma}_\theta^{2(REML)}$ are simply equal to $\sum(ES_i - \overline{ES})^2/(k-1)$, the UMVUE of the variance in the effect sizes. On the other hand, $\hat{\sigma}_\theta^{2(ML)}$ simpli-

TABLE 3
Relative Efficiency and MSE of the HS and ML Estimators Compared with the DL, HE, and REML Estimators of σ_θ^2 Assuming Homogeneous Sampling Variances

	Number of Effect Sizes (k)				
Parameter	5	10	20	40	80
Relative efficiency	0.64	0.81	0.90	0.95	0.98
Relative MSE	0.72	0.86	0.93	0.96	0.98

fies to $\sum(ES_i - \overline{ES})^2/k$, the typical maximum likelihood estimate of the variance. Therefore,

$$\text{Var}\left[\hat{\sigma}_\theta^{2\,(ML)}\right] = \frac{2(k-1)}{k^2}\,\sigma_\theta^4 \tag{24}$$

and

$$\text{Var}\left[\hat{\sigma}_\theta^{2\,(HE/REML)}\right] = \frac{2}{(k-1)}\,\sigma_\theta^4. \tag{25}$$

On the other hand, when $\sigma_{\epsilon_i}^2 = 0$ for any one of the effect sizes, both the HS and DL estimators and their sampling variances are undefined. However, if we let $\sigma_{\epsilon_i}^2$ get arbitrarily close to 0 for all k effect sizes, then lower bounds of Var[$\hat{\sigma}_\theta^{2\,(HS)}$] and Var[$\hat{\sigma}_\theta^{2\,(DL)}$] are given by Equations 24 and 25, respectively, but these lower bounds are never actually quite reached, even as k becomes large. In fact, the relative efficiency of the HE and REML estimators compared to the DL estimator will consistently fall slightly below 1, no matter how large k becomes. The same applies to the relative efficiency of the ML estimator when compared to the HS estimator.

From these results, and the fact that the Cramer–Rao lower bound is equal to $(2/k)\sigma_\theta^4$ when the sampling variances are zero, we can conclude that: (a) Var[$\hat{\sigma}_\theta^{2\,(HE/REML)}$] and Var[$\hat{\sigma}_\theta^{2\,(DL)}$] exceed the Cramer–Rao lower bound and consequently, the HE, DL, and REML estimators are less than fully efficient for finite k, even when the $\sigma_{\epsilon_i}^2$ values become arbitrarily close to zero; (b) for large k, the HE, ML, and REML estimators are fully efficient; and (c) the HS and DL estimators are not fully efficient even as k becomes large.

5.3. Homogeneous Population Effect Sizes

When the population effect sizes are homogeneous, then $\sigma_\theta^2 = 0$. The sampling variances and MSEs of the three noniterative estimators then simplify to

$$\text{Var}\left[\hat{\sigma}_\theta^{2\,(HS)}\right] = \frac{2(k-1)}{\left(\sum_{i=1}^{k} w_i\right)^2},$$

$$MSE\left[\hat{\sigma}_\theta^{2\,(HS)}\right] = \frac{2k-1}{\left(\sum_{i=1}^{k} w_i\right)^2},$$

$$\text{Var}\left[\hat{\sigma}_\theta^{2(HE)}\right] = MSE\left[\hat{\sigma}_\theta^{2(HE)}\right] = \frac{2}{k^2}\left[\sum_{i=1}^{k} w_i^{-2} + \frac{1}{(k-1)^2}\sum_{i=1}^{k}\sum_{j\neq i}^{k} w_i^{-1} w_j^{-1}\right],$$

$$\text{Var}\left[\hat{\sigma}_\theta^{2(DL)}\right] = MSE\left[\hat{\sigma}_\theta^{2(DL)}\right] = \frac{2(k-1)}{c},$$

where $w_i = 1/\sigma^2_{\epsilon_i}$. The Cramer–Rao lower bound of σ^2_θ is now equal to $2/\sum w_i^2$. It is easy to show that $\text{Var}[\hat{\sigma}^{2\,(HS)}_\theta]$ can be smaller than $2/\sum w_i^2$, which again illustrates that the sampling variance of the HS estimator can actually fall below the Cramer–Rao lower bound of unbiased estimators. On the other hand, $\text{Var}[\hat{\sigma}^{2\,(HE)}_\theta]$ and $\text{Var}[\hat{\sigma}^{2\,(DL)}_\theta]$ are always greater than the Cramer–Rao lower bound for finite k. Finally, it can be shown that

$$\text{Var}\left[\hat{\sigma}^{2\,(HS)}_\theta\right] < \text{Var}\left[\hat{\sigma}^{2\,(DL)}_\theta\right] < \text{Var}\left[\hat{\sigma}^{2\,(HE)}_\theta\right]$$

and

$$MSE\left[\hat{\sigma}^{2\,(HS)}_\theta\right] < MSE\left[\hat{\sigma}^{2\,(DL)}_\theta\right] < MSE\left[\hat{\sigma}^{2\,(HE)}_\theta\right].$$

5.4. General Results for the Noniterative Estimators

From Equations 13 and 16, we see that $\text{Var}[\hat{\sigma}^{2\,(DL)}_\theta] - \text{Var}[\hat{\sigma}^{2\,(HE)}_\theta]$ can be written as a quadratic equation of the form $A(\sigma^2_\theta)^2 + B(\sigma^2_\theta) + C$. By determining that B and C are nonpositive and A non-negative, Friedman (2000) showed that the DL estimator is more efficient for smaller values of σ^2_θ, while the *HE* estimator will be more efficient when the amount of heterogeneity becomes large. Because both of these estimators are unbiased, the same conclusion applies to the MSE. Friedman's results are now extended to include the HS estimator.

From Equations 8 and 15, we see that $\text{Var}[\hat{\sigma}^{2\,(HS)}_\theta]$ and $\text{Var}[\hat{\sigma}^{2\,(DL)}_\theta]$ only differ by a multiplicative term. Moreover, it is easy to prove that $1/c^2 > 1/(\sum w_i)^2$ and, therefore, the HS estimator always has smaller sampling variance than the DL estimator.

A comparison between the HE and the HS estimator reveals results that are not as unequivocal. From Equation 9, we see that $\text{Var}[\hat{\sigma}^{2\,(HS)}_\theta] - \text{Var}[\hat{\sigma}^{2\,(HE)}_\theta]$ is also of the form $A(\sigma^2_\theta)^2 + B(\sigma^2_\theta) + C$. It can be shown that C and B are nonpositive. However A can take on positive or negative values, depending on k and the w_i values. When the sampling variances are homogeneous, A is always negative, which implies that the HS estimator is more efficient than the HE estimator (see also section 5.1). As the sampling variances become increasingly heterogeneous, A eventually becomes positive and will do so more rapidly for larger k. When A is positive, then the HS estimator is more efficient for smaller values of σ^2_θ, while the HE estimator is more efficient for larger values of σ^2_θ.

The MSE of the HS estimator is also a quadratic equation in σ^2_θ. Comparing the MSE of the HS and the DL estimators reveals that C and B are always nonpositive. The value of A is also always nonpositive for homogeneous sampling variances (see section 5.1). Without this assumption, A will remain to be nonpositive, except under some unusual circumstances. In particular, when the w_i values are relatively homogeneous except for one comparatively large value, then A can become positive. In other words, A might become positive when the sample sizes of the studies on which the effect sizes are based are roughly homogeneous except for one study with a very large sample size. In this case, the MSE of the HS estimator might exceed that of the DL estimator. However, in most other cases, the HS estimator will have smaller MSE.

When comparing the MSE of the HE estimator with that of the HS estimator, we obtain essentially the same pattern of results as we did when comparing their sampling variances. In the equation $A(\sigma_\theta^2)^2 + B(\sigma_\theta^2) + C$, the values of B and C will again be nonpositive, with A being negative for homogeneous sampling variances (see section 5.1) and A taking on positive values as the sampling variances become increasingly heterogeneous.

To summarize, we can conclude that; (a) for small values of σ_θ^2, the HS estimator has lower sampling variance and MSE than the DL estimator, which in turn is more efficient and has lower MSE than the HE estimator; (b) for homogeneous sampling variances, the HE and DL estimators have the same sampling variance and MSE, and are always less efficient and have higher MSE than the HS estimator; (c) for heterogeneous sampling variances and sufficiently large σ_θ^2, the HE estimator will be more efficient and have smaller MSE than the HS and DL estimators; and (d) for heterogeneous sampling variances and sufficiently large σ_θ^2, the HS estimator will be more efficient and have smaller MSE than the DL estimator, unless the heterogeneity is caused by a single very small sampling variance, in which case the MSE of the DL will be lower than that of the HS estimator.

5.5. Bias Due to Truncation of Negative Variance Estimates

It is possible to obtain negative estimates of σ_θ^2 by all of the estimators considered in this article. In fact, for $\sigma_\theta^2 = 0$, we should expect to obtain a negative estimate about half of the time when using an unbiased estimator, while a negatively biased estimator would be expected to result in negative estimates more than half of the time. Because a negative variance estimate is inadmissible, the common practice is to truncate such values to zero. Such truncation, however, introduces a certain amount of positive bias into the estimators. This bias will increase with smaller k and larger values of $\sigma_{\epsilon_i}^2$, because estimates of σ_θ^2 become more variable under these conditions and, therefore, cases requiring truncation become more prevalent. On the other hand, as σ_θ^2 moves further away from zero, the bias caused by truncation will decrease because the probability that $\hat{\sigma}_\theta^2$ is smaller than zero then shrinks accordingly.

Deriving the exact amount of bias in the estimates of σ_θ^2 due to truncation is possible after imposing some restrictive assumptions. For the case $\sigma_\theta^2 = 0$, the bias in $\hat{\sigma}_\theta^{2\,(DL)}$ due to truncation of negative estimates is given by

$$\text{Bias}\left[\hat{\sigma}_\theta^{2\,(DL)}\right] = \int_{k-1}^{\infty} \frac{q-(k-1)}{c} f(q)dq,$$

where $f(x)$ is the probability density function of a chi-square random variable with $k-1$ degrees of freedom. The bias in $\hat{\sigma}_\theta^{2\,(HS)}$ is equal to

$$\text{Bias}\left[\hat{\sigma}_\theta^{2\,(HS)}\right] = \int_{k}^{\infty} \frac{q-k}{\sum_{i=1}^{k} w_i} f(q)dq.$$

Two different factors now cause $\hat{\sigma}_\theta^{2\,(HS)}$ to be biased. From Equation 7 we know that $\hat{\sigma}_\theta^{2\,(HS)}$ is negatively biased in its unconstrained form. On the other hand, the truncation introduces some positive bias into the estimator. One might hope that these two sources of bias cancel each other out, but as will be shown momentarily, this will be the case only in very particular circumstances.

If we assume that the sampling variances are homogeneous across the k studies, then we can obtain results for all five estimators and for values of $\sigma_\theta^2 > 0$. The bias is then given by

$$\text{Bias}\left[\hat{\sigma}_\theta^{2\,(DL/HE/REML)}\right] = \frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2}\int_{(k-1)\sigma_\epsilon^2}^{\infty}\left(\frac{s}{k-1} - \sigma_\epsilon^2\right) f\left(\frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2}\, s\right) ds - \sigma_\theta^2$$

and

$$\text{Bias}\left[\hat{\sigma}_\theta^{2\,(HS/ML)}\right] = \frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2}\int_{k\sigma_\epsilon^2}^{\infty}\left(\frac{s}{k} - \sigma_\epsilon^2\right) f\left(\frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2}\, s\right) ds - \sigma_\theta^2.$$

Hedges and Vevea (1998) gave an equivalent expression for the bias in $\hat{\sigma}_\theta^{2\,(DL)}$. Values of the bias in these estimators can be obtained by numerical integration.

Figure 1 illustrates the extent of the bias in the estimators under the assumption of homogeneous sampling variances ($\sigma_\epsilon^2 = 1$) for various values of k and σ_θ^2. We note that the HE, DL, and REML estimators are always positively biased when truncated. The HS and ML estimators are also positively biased when $\sigma_\theta^2 = 0$, but to a lesser extent because the inherent negative bias of these estimators counteracts some of the positive bias caused by truncation. For larger values of σ_θ^2 (when truncation becomes less prevalent), the negative bias in $\hat{\sigma}_\theta^{2(HS/ML)}$ takes over and negatively biases these estimators. There are only a few cases where the two types of biases cancel each other out. For example, for $k = 10$, the positive and negative bias cancel each other out when $\sigma_\theta^2 \approx .302$ and for $k = 20$, this occurs when $\sigma_\theta^2 \approx .275$. Moreover, Figure 1 demonstrates that as long as σ_θ^2 is not too large, the absolute bias is lower for $\hat{\sigma}_\theta^{2(HS/ML)}$ than for $\hat{\sigma}_\theta^{2\,(DL/HE/REML)}$. Finally, as $k \to \infty$, the bias decreases for all estimators toward zero.

Whether the positive bias caused by truncation is important or not depends on how an estimator of σ_θ^2 will be used. When $\hat{\sigma}_\theta^2$ is simply an indicator of the existence and extent of population heterogeneity, then truncation bias is irrelevant because truncation would not change the conclusions. However, truncated and, therefore, positively biased estimates will introduce bias into Var[$\overline{ES}$] that will usually lead researchers to understate the accuracy of their estimate of μ_θ (at least, when the estimator of $\hat{\sigma}_\theta^2$ without truncation is unbiased).

6. Monte Carlo Simulations

The analytic comparisons between the estimators were supplemented with Monte Carlo simulations because of several reasons. First of all, it was necessary to assume normally distributed effect size measures to derive the sampling vari-

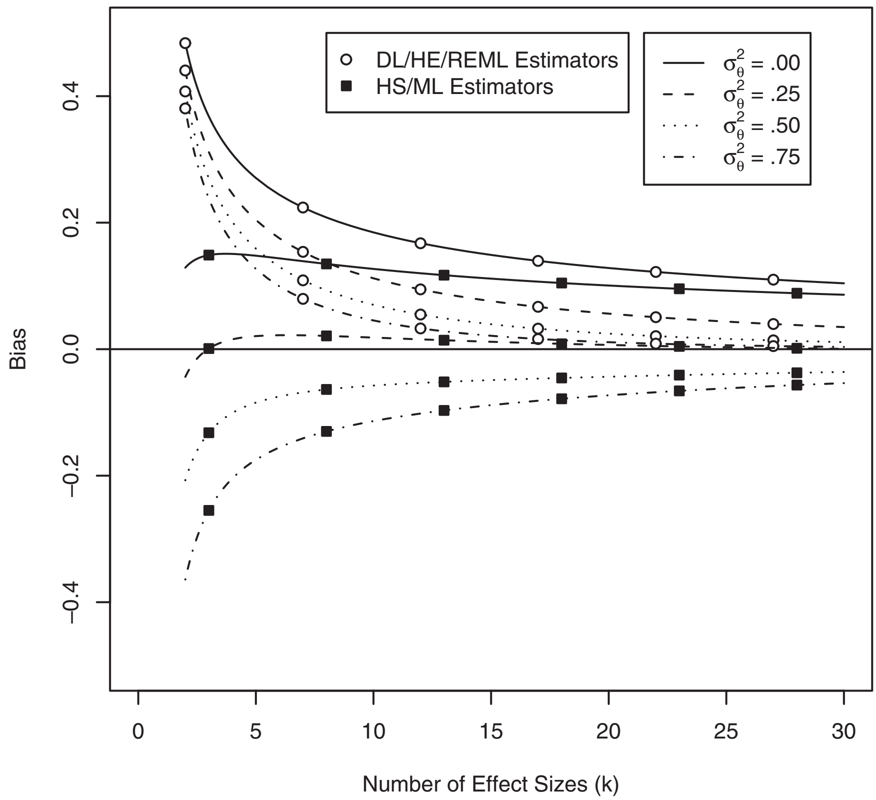

FIGURE 1. *Bias in the Hunter–Schmidt/maximum likelihood (HS/ML) and DerSimonian–Laird/Hedges/restricted maximum likelihood (DL/HE/REML) estimators of* σ_θ^2 *when truncating negative estimates (homogeneous sampling variances assumed).*

ances of the σ_θ^2 estimators. However, the normality assumption only holds asymptotically for many commonly used effect sizes. Moreover, the assumption of known $\sigma_{\epsilon_i}^2$ values will be violated in practice, as the sampling variances usually must be estimated from the data. Also, not all effect size measures provide unbiased estimates and for some, $\sigma_{\epsilon_i}^2$ is known to depend on θ_i, thereby introducing some dependency between ϵ_i and τ_i. In addition, some of the results were obtained under the assumption of homogeneous sampling variances, which might not hold in practice. Finally, without some actual numerical results, it remains unclear to what extent estimates of σ_θ^2, $\overline{ES}$, and Var[$\overline{ES}$] are influenced by the various methods for estimating σ_θ^2 under realistic conditions.

Simulations were conducted with two different effect size measures: the unstandardized and the standardized mean difference. The parameters k, μ_θ, σ_θ^2, and $\sigma_{\epsilon_i}^2$ were manipulated systematically to determine how they affect estimates of σ_θ^2. The factors were completely crossed and for each condition, 100,000 meta-analyses were simulated. On each iteration, k values of θ_i were first generated from $N(\mu_\theta, \sigma_\theta^2)$. Then k values of ES_i and $\hat{\sigma}_{\epsilon_i}^2$ were generated from the appropriate distributions. Estimates of σ_θ^2 were obtained with each of the five methods discussed.

Estimates of σ_θ^2 were not truncated for calculating the bias, efficiency, and MSE. Trials where the ML or REML estimators did not converge were skipped and replaced by an additional trial to ensure that exactly 100,000 iterations were run for each condition. Overall, this occurred in less than 0.04% of the trials and, therefore, should not have a substantial impact on the results. After constraining the $\hat{\sigma}_\theta^2$ values to be non-negative, $\overline{ES}$ and Var[$\overline{ES}$] were calculated with each of the five variance estimators using Equations 2 and 4. The usual large-sample test of H_0: $\mu_\theta = 0$ is given by $z = \overline{ES}/SE/[\overline{ES}]$, which has an asymptotic standard normal distribution under the null hypothesis. The value of z was obtained for each method and tested for significance at $\alpha = .05$.

6.1. Unstandardized Mean Difference

6.1.1. Methods

Studying the unstandardized mean difference (UMD) has several advantages: its distribution is exactly normal, it can be estimated unbiasedly, and its sampling variance is independent of the population effect size. Therefore, the UMD allows us to investigate the various estimators of the population heterogeneity under what might be considered ideal conditions.

Let X_{ij}^C and X_{ij}^E be the jth observations from a control and an experimental group in the ith study. Assume that $X_{ij}^C \sim N(\mu_i^C, \sigma_i^2)$ and $X_{ij}^E \sim N(\mu_i^E, \sigma_i^2)$. For the ith study, we define the UMD as $\theta_i = \mu_i^E - \mu_i^C$. Given n_i^C and n_i^E observations from the control and experimental group, respectively, we can estimate θ_i unbiasedly by $ES_i = \bar{X}_i^E - \bar{X}_i^C$, which is distributed $N(\mu_i^E - \mu_i^C, \sigma_i^2(1/n_i^E + 1/n_i^C))$. The sampling variance $\sigma_{\epsilon_i}^2$ of ES_i can be estimated unbiasedly by $s_i^2(1/n_i^E + 1/n_i^C)$, where s_i^2 is the typical pooled within-group variance.

To make the number of simulated conditions more manageable, it was assumed that $n_i = n_i^C = n_i^E$ and σ_i^2 was set to 10. Finally, μ_i^C was set to zero and μ_i^E was sampled from $N(\mu_\theta, \sigma_\theta^2)$ to generate heterogeneous values of θ_i. The following factors were manipulated in the simulations: $k = (5, 10, 20, 40, 80)$, $\mu_\theta = (0, 1, 2, 4)$, $\sigma_\theta^2 = (0, 0.125, 0.25, 0.5, 1)$, and $\bar{n}_i = (20, 40, 80, 160, 320)$, and consequently, $\sigma_{\epsilon_i}^2 = (1, 0.5, 0.25, 0.125, 0.0625)$. To simulate heterogeneous sampling variances, the values of n_i were sampled from a normal distribution with mean $\bar{n}_i$ and standard deviation $\bar{n}_i/3$.

6.1.2. Results

The estimates of σ_θ^2 were insensitive to the different values of μ_θ. The HE estimator was unbiased as expected. On the other hand, the DL and REML estimators were slightly positively biased for small $\bar{n}_i$, that is when the $\hat{\sigma}_{\epsilon_i}^2$ values are larger and less efficient estimates of the true sampling variances. For $\bar{n}_i = 20$, the bias fluctuated between 0.05 and 0.09 and for $\bar{n}_i = 40$ between 0.01 and 0.02, with no improvement as k increased. The bias in the HS and ML estimators was roughly the same and quite substantial in some conditions. The bias in these estimators increased with

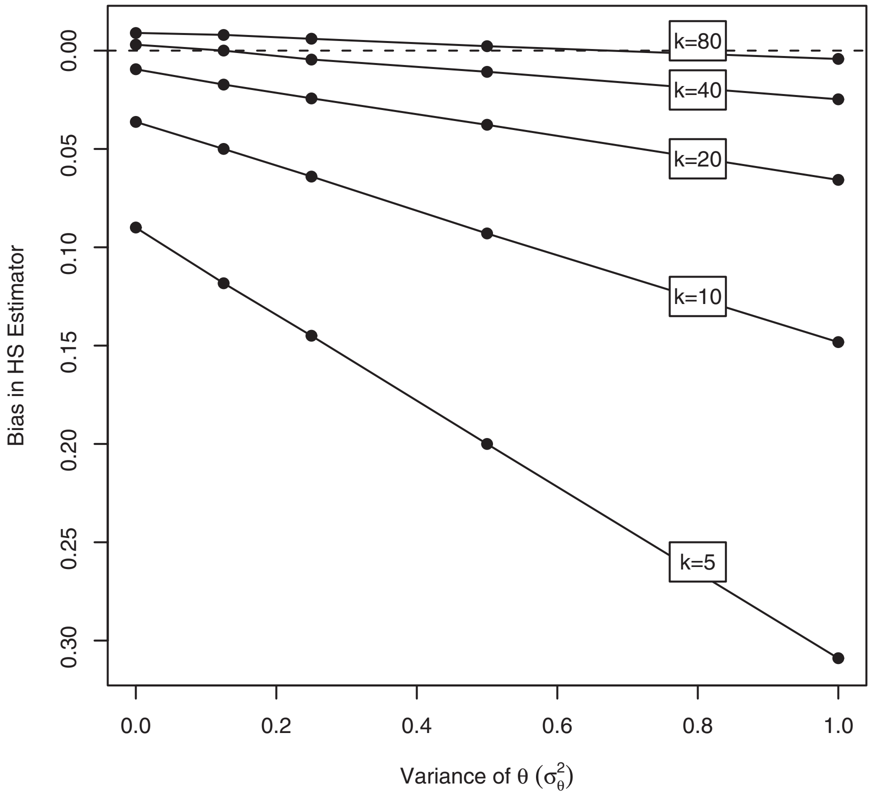

FIGURE 2. *Bias in the Hunter–Schmidt (HS) estimator of* σ_θ^2 *when using the unstandardized mean difference (UMD) effect size* ($\bar{n}_i = 40$).

larger values of σ_θ^2 and decreased as k and $\bar{n}_i$ increased. Figure 2 illustrates the extent of the bias in the HS estimator for various values of σ_θ^2 and k when $\bar{n}_i = 40$.

Not surprisingly, the sampling variability and MSE of the heterogeneity estimators decreased as k and $\bar{n}_i$ increased. On the other hand, the sampling variability and MSE increased with larger values of σ_θ^2. In Figure 3, the MSE of the five estimators is plotted against σ_θ^2 for some representative values of k and $\bar{n}_i$, which reveals these trends. Moreover, we can sort the estimators into three groups according to their efficiency and MSE. The first group is comprised of the HS and ML estimators, which were roughly equally efficient and had the same MSE across all values of k, $\bar{n}_i$, and σ_θ^2. The DL and REML estimators form the second group, as these two estimators also differed insubstantially in efficiency and MSE. However, the HS and ML estimators were more efficient and had lower MSE than the DL and REML estimators, which can be seen in Figure 3. However, it is not readily apparent from the figure that the *relative* efficiency and MSE of these estimators

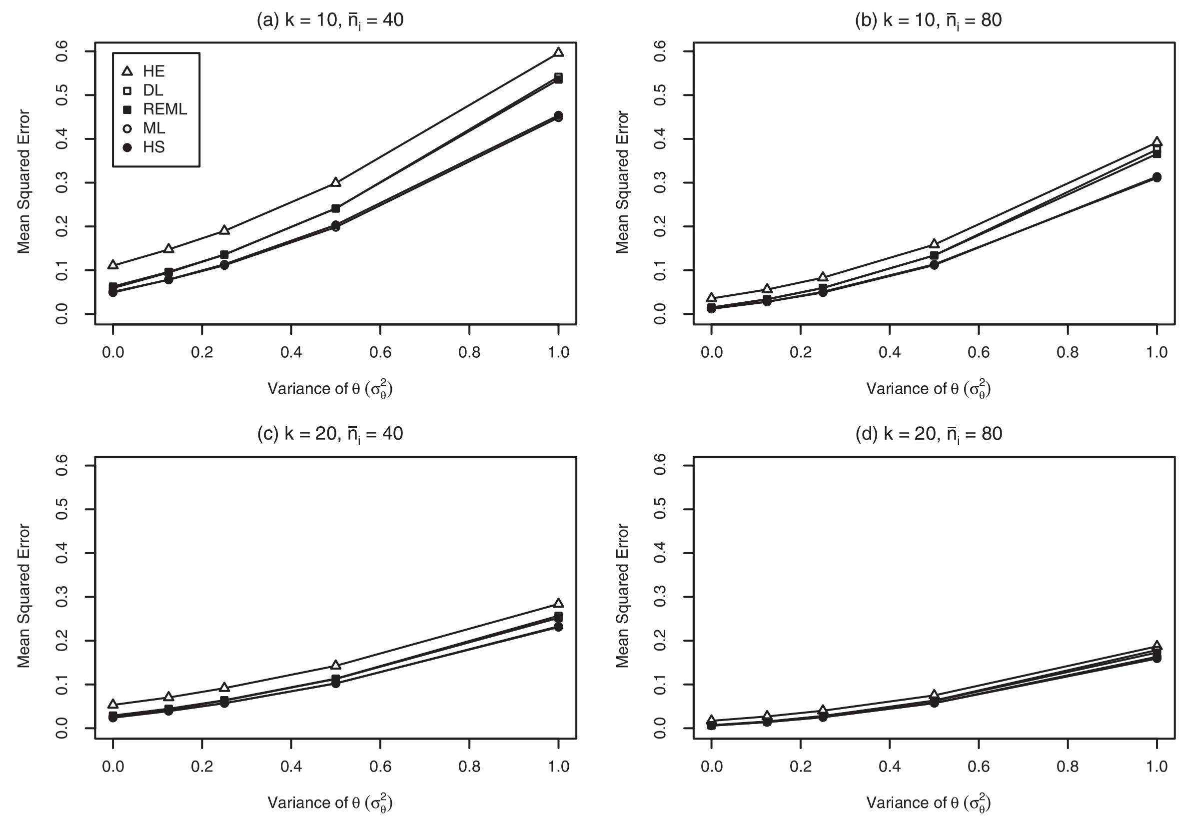

FIGURE 3. *Mean squared error of the Hunter–Schmidt (HS), Hedges (HE), DerSimonian–Laird (DL), maximum likelihood (ML), and restricted maximum likelihood (REML) estimators of* σ^2_θ *when using the unstandardized mean difference (UMD) effect size.*

TABLE 4

Relative Efficiency and MSE of the ML Versus REML Estimators of σ_θ^2

		Number of Effect Sizes (k)				
Parameter	Effect Size	5	10	20	40	80
Relative efficiency	UMD	0.63	0.81	0.90	0.95	0.97
Relative MSE	UMD	0.70	0.84	0.91	0.94	0.96
Relative efficiency	SMD	0.63	0.80	0.90	0.95	0.97
Relative MSE	SMD	0.73	0.88	0.95	0.99	1.01

Note. UMD = unstandardized mean difference; SMD = standardized mean difference.

only depended on k, and not on $\bar{n}_i$ or σ_θ^2. Table 4 shows the relative efficiency and MSE of these estimators and demonstrates that both approached 1 as k increased. In fact, a comparison with Table 3 reveals that the analytic results from section 5.1 closely match the empirical findings. Finally, the third group consists of the HE estimator, which fell below the DL and REML estimators in efficiency and MSE (except when $\sigma_\theta^2 = 1$ and $\bar{n}_i \geq 160$). The relative efficiency of these estimators depended on the values of k, $\bar{n}_i$, and σ_θ^2 and is illustrated for the DL and HE estimators in Figure 4 for $k = 20$. The relative MSE was essentially identical to the relative efficiency.

Estimates of $\overline{ES}$ were unbiased in all cases, but the Type I error for the test of H_0: $\mu_\theta = 0$ was only controlled adequately at $\alpha = .05$ when σ_θ^2 was close to zero and/or k was large. For small k, the probability of falsely rejecting H_0 became increasingly inflated for all methods as $\bar{n}_i$ and σ_θ^2 increased. The Type I errors when using the HS and ML estimators were approximately the same. Similarly, using the HE, DL, and REML estimators resulted in common Type I error rates and overall provided better control of the Type I error, but only marginally so. Table 5 illustrates this for the HS and HE estimators when $\sigma_\theta^2 = 1$ and $\bar{n}_i \leq 80$. There was little to no change for values of $\bar{n}_i$ greater than 80. Note that the coverage probability of a 95% confidence interval for μ_θ would be given by one minus the values in Table 5.

6.2. Standardized Mean Difference

6.2.1. Methods

When $X_{ij}^C \sim N(\mu_i^C, \sigma_i^2)$ and $X_{ij}^E \sim N(\mu_i^E, \sigma_i^2)$ as for the UMD but the measurement scales are not commensurable across studies, then the standardized mean difference (SMD) is usually chosen as an effect size measure. Now, $\theta_i = (\mu_i^E - \mu_i^C)/\sigma_i$, which can be estimated unbiasedly by $ES_i = c(m_i)(\bar{X}_i^E - \bar{X}_i^C)/s_i$, where

$$c(m_i) = \frac{\Gamma\left(\frac{m_i}{2}\right)}{\left(\frac{m_i}{2}\right)^{1/2}\Gamma\left(\frac{m_i - 1}{2}\right)}$$

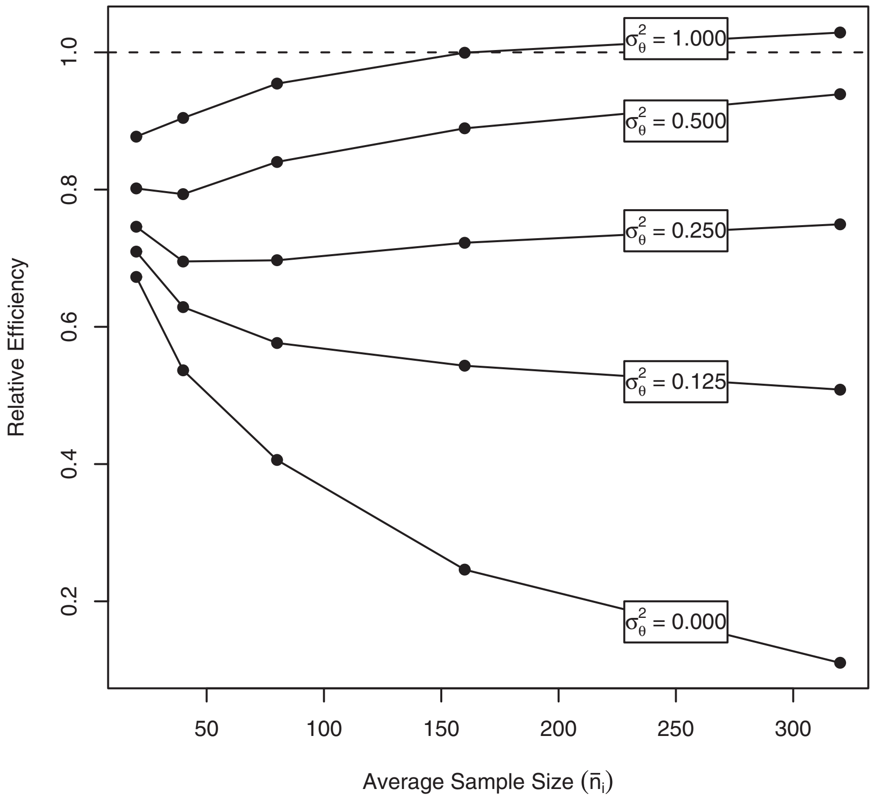

FIGURE 4. *Relative efficiency of the DerSimonian–Laird (DL) versus the Hedges (HE) estimator of* σ^2_θ *when using the unstandardized mean difference (UMD) effect size* (k = 20).

and $m_i = n_i^E + n_i^C - 2$ (Hedges, 1981). An unbiased estimate of $\sigma^2_{\epsilon_i}$ is given by

$$\hat{\sigma}^2_{\epsilon_i} = \frac{1}{\tilde{n}_i} + \left(1 - \frac{(m_i - 2)}{m_i[c(m_i)]^2}\right) ES_i^2,$$

where $\tilde{n}_i = (n_i^E n_i^C)/(n_i^E + n_i^C)$ (Hedges, 1983). The distribution of ES_i is asymptotically normal and is closely related to a noncentral t-distribution. In fact, $c(m_i)^{-1}(\tilde{n}_i)^{1/2} ES_i$ is distributed noncentral t, with m_i degrees of freedom and noncentrality parameter $\theta_i(\tilde{n}_i)^{1/2}$. The exact sampling variance of ES_i is equal to

$$\sigma^2_{\epsilon_i} = \frac{[c(m_i)]^2 m_i \left[1 + \tilde{n}_i \theta_i^2\right]}{(m_i - 2)\tilde{n}_i} - \theta_i^2,$$

which is no longer independent of θ_i, that is, τ_i.

TABLE 5

Type I Error for the Test of H_0: $\mu_\theta = 0$ Based on the HS and HE Estimators of σ_θ^2 Using the Unstandardized Mean Difference Effect Size ($\sigma_\theta^2 = 1$ and $\alpha = .05$)

	Using HS Estimator			Using HE Estimator		
k	$\bar{n}_i = 20$	$\bar{n}_i = 40$	$\bar{n}_i = 80$	$\bar{n}_i = 20$	$\bar{n}_i = 40$	$\bar{n}_i = 80$
5	0.11	0.13	0.15	0.10	0.11	0.12
10	0.09	0.10	0.10	0.08	0.09	0.09
20	0.07	0.07	0.07	0.07	0.07	0.07
40	0.06	0.06	0.06	0.07	0.06	0.06
80	0.06	0.06	0.06	0.06	0.06	0.05

Again, it was assumed that $n_i = n_i^C = n_i^E$. The following factors were manipulated in the simulations: $k = (5, 10, 20, 40, 80)$, $\mu_\theta = (0, 0.2, 0.5, 0.8)$ (following Cohen's [1988] conventional definitions of small, medium, and large SMDs), $\sigma_\theta^2 = (0, 0.01, 0.025, 0.05, 0.1)$, and $\bar{n}_i = (20, 40, 80, 160, 320)$, and consequently, $\sigma_{\epsilon_i}^2$ was between 0.006 and 0.008 for $n_i = 320$ and between 0.101 and 0.110 for $n_i = 20$.

6.2.2. Results

The estimates of σ_θ^2 were again insensitive to the different values of μ_θ. This is somewhat surprising, because the distribution of ES_i becomes increasingly skewed as θ_i increases. Apparently, within the range of θ_i values studied, the extent of skew in the distribution of the ES_i values was not substantial enough to introduce any noticeable dependencies between ϵ_i and τ_i.

The HE estimator was unbiased across all conditions. The DL and REML estimators were slightly negatively biased when $\bar{n}_i = 20$, but the bias never exceeded –0.01. The HS and ML estimators again revealed a negative bias that decreased with larger k and $\bar{n}_i$ and increased with larger σ_θ^2 values. In Figure 5, the bias in the HS estimator is plotted for various values of σ_θ^2 and k when $\bar{n}_i = 40$.

In Figure 6, the MSE of the five estimators is plotted against σ_θ^2 for various values of k and $\bar{n}_i$. Although not quite as equivocal, we can again distinguish three groups based on efficiency and MSE, the first consisting of the HS and ML estimators, the second of the DL and REML estimators, and the third of the HE estimator (Figure 6b demonstrates this hierarchy most clearly). The relative efficiency of the HS and ML estimators compared to the DL and REML estimators again depended only on k. Table 4 shows that the ML estimator was substantially more efficient than the REML estimator for small to moderate values of k with the relative efficiency approaching 1 as k increased. The results for the MSE were similar. Finally, the DL and REML estimators were always more efficient than the HE estimator, often substantially so. Figure 7 illustrates this by showing the relative efficiency of the DL compared to the HE estimator for $k = 20$.

Estimates of μ_θ were slightly negatively biased (by –0.01 to –0.02) when μ_θ was large and $\bar{n}_i$ small. The bias was the same regardless of which estimator of σ_θ^2 was

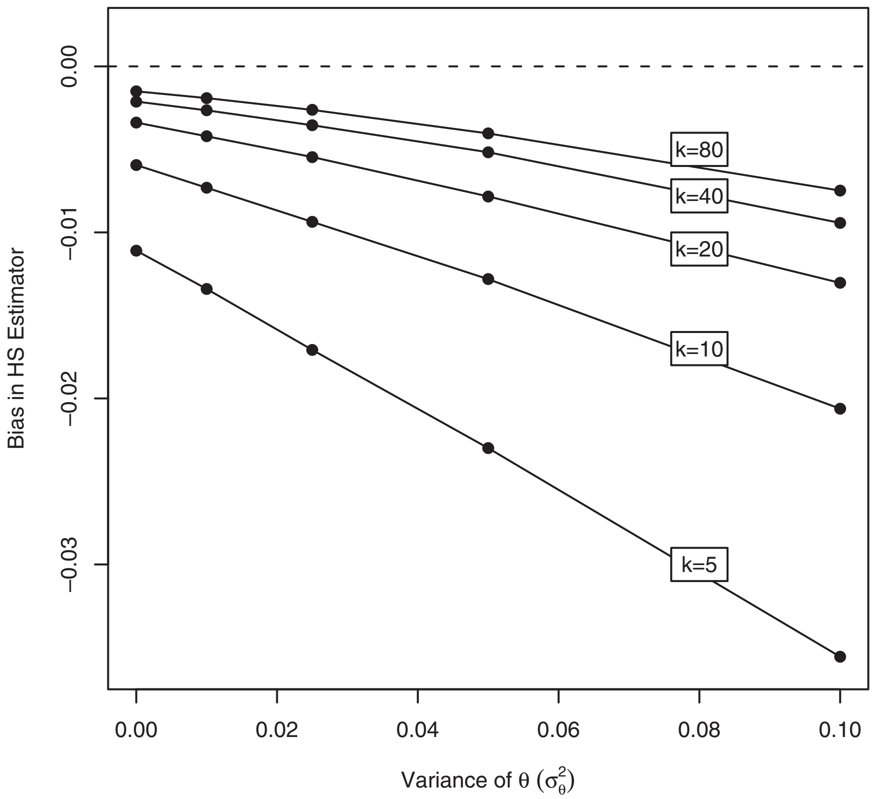

FIGURE 5. *Bias in the Hunter–Schmidt (HS) estimator of* σ^2_θ *when using the standardized mean difference (SMD) effect size* ($\bar{n}_i = 40$).

used to calculate $\overline{ES}$ and did not depend on the value of σ^2_θ. The Type I error for the test of H_0: $\mu_\theta = 0$ was only controlled adequately at $\alpha = .05$ when σ^2_θ was close to zero and/or *k* was large. Again, the Type I error when using the HS and ML estimators and when using the HE, DL, and REML estimators were approximately the same. The latter three resulted in Type I error rates slightly closer to the nominal $\alpha = .05$ value than the HS and ML estimators, as shown in Table 6.

6.3. General Conclusions About the Simulations

For both effect size measures studied, the five variance estimators were not influenced by the value of μ_θ. However, the statistical properties of the various estimators depended on the number of effect sizes, the sample sizes, and the amount of heterogeneity in the population effects.

The DL and REML estimators were slightly biased for small $\bar{n}_i$ values with *k*, the number of effect sizes, having no influence on the bias. The analytic results, on the

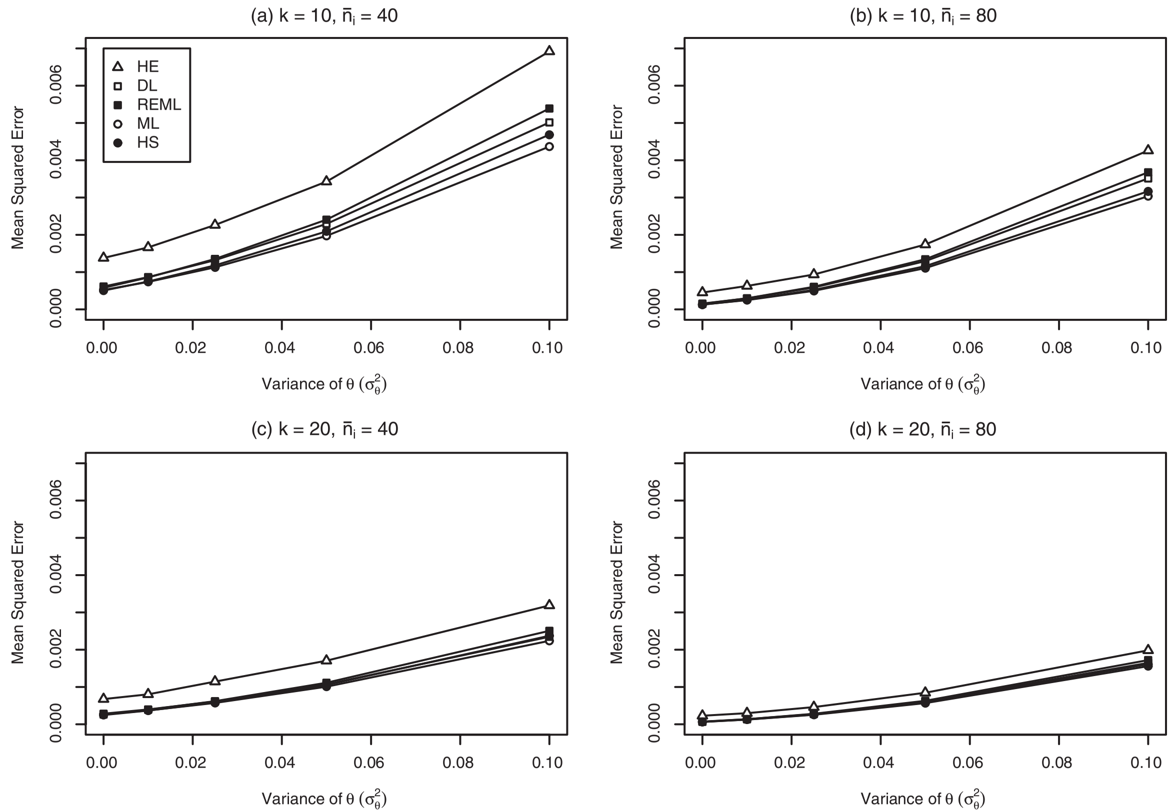

FIGURE 6. *Mean squared error of the Hunter–Schmidt (HS), Hedges (HE), DerSimonian–Laird (DL), maximum likelihood (ML), and restricted maximum likelihood (REML) estimators of* σ_θ^2 *when using the standardized mean difference (SMD) effect size.*

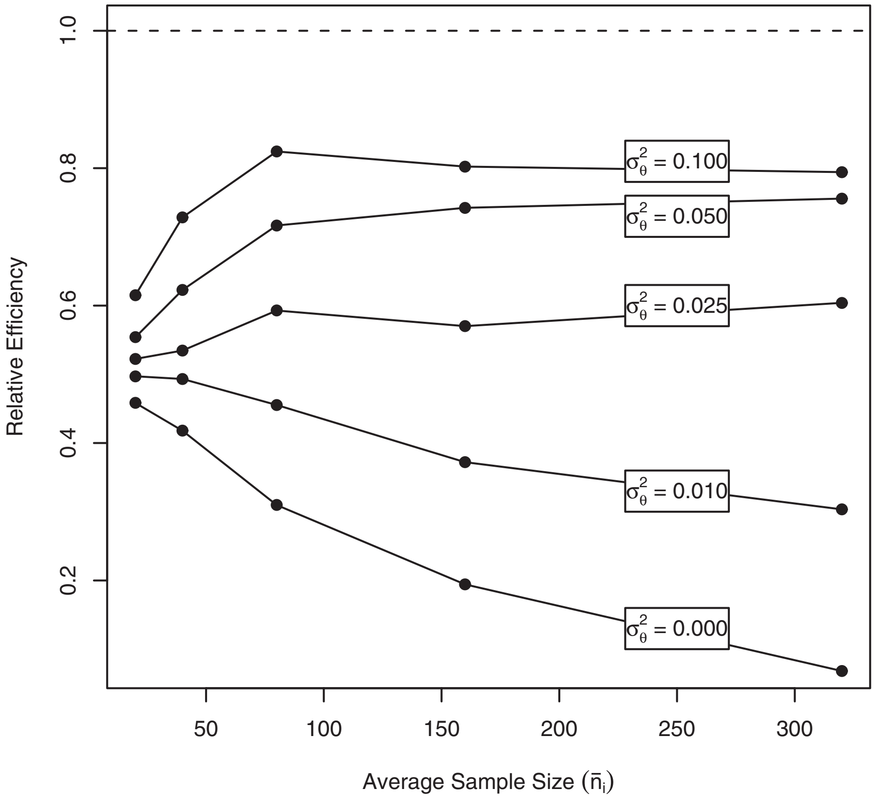

FIGURE 7. *Relative efficiency of the DerSimonian–Laird (DL) versus the Hedges (HE) estimator of* σ_θ^2 *when using the standardized mean difference (SMD) effect size (*k = *20).*

TABLE 6
Type I Error for the Test of H_0*:* $\mu_\theta = 0$ *Based on the HS and HE Estimators of* σ_θ^2 *Using the Standardized Mean Difference Effect Size (*$\sigma_\theta^2 = 0.1$ *and* $\alpha = .05$*)*

	Using HS Estimator			Using HE Estimator		
k	$\bar{n}_i = 20$	$\bar{n}_i = 40$	$\bar{n}_i = 80$	$\bar{n}_i = 20$	$\bar{n}_i = 40$	$\bar{n}_i = 80$
5	0.10	0.13	0.15	0.08	0.10	0.12
10	0.08	0.09	0.10	0.07	0.08	0.08
20	0.07	0.07	0.07	0.06	0.06	0.06
40	0.06	0.06	0.06	0.05	0.06	0.06
80	0.05	0.06	0.06	0.05	0.05	0.05

other hand, indicated that the DL estimator is unbiased when the sampling variances are exactly known. When $\bar{n}_i$ is large, estimates of $\sigma^2_{\epsilon_i}$ are more efficient. In fact, assuming $n_i = n_i^C = n_i^E$, we can show for the unstandardized mean difference that

$$\mathrm{Var}\left[\hat{\sigma}^2_{\epsilon_i}\right] = \frac{4\sigma_i^4}{n_i^2(n_i - 1)},$$

and for the standardized mean difference, relying on some asymptotic results,

$$\mathrm{Var}\left[\hat{\sigma}^2_{\epsilon_i}\right] = \frac{\left(8 + \theta_i^2\right)^2}{128 n_i^4} + \frac{\theta_i^2\left(8 + \theta_i^2\right)}{16 n_i^3}.$$

Clearly, the estimates of the sampling variances quickly become quite accurate, and, therefore, the assumption of known $\sigma^2_{\epsilon_i}$ values may not be that unreasonable in practice. In fact, for $\bar{n}_i \geq 40$, the DL and REML estimators were essentially unbiased.

For the *HS* and *ML* estimators, the bias depended on k, $\bar{n}_i$, and σ^2_θ as we would expect based on Equation 7 and could be quite substantial. Only the HE estimator consistently provided exactly unbiased estimates of σ^2_θ regardless of the conditions. However, the unbiasedness of this estimator came at a price because $\hat{\sigma}^{2\,(HE)}_\theta$ was also the least efficient estimator, with only a few exceptions. The analytic results suggested that for large enough σ^2_θ, the HE estimator should surpass the DL and even the *HS* estimator in efficiency. Although Figures 4 and 7 support this result, we also see that σ^2_θ was essentially not large enough to make $\hat{\sigma}^{2\,(HE)}_\theta$ the more efficient estimator (except for $\bar{n}_i = 320$ and $\sigma^2_\theta = 1$ in Figure 4). With respect to the other estimators, the analytic results for homogeneous sampling variances given by Equations 22 and 23 generally held. In other words, within the conditions studied and ignoring some exceptions, we conclude that, approximately,

$$\mathrm{Var}\left[\hat{\sigma}^{2\,(HS/ML)}_\theta\right] < \mathrm{Var}\left[\hat{\sigma}^{2\,(DL/REML)}_\theta\right] < \mathrm{Var}\left[\hat{\sigma}^{2\,(HE)}_\theta\right]$$

and

$$MSE\left[\hat{\sigma}^{2\,(HS/ML)}_\theta\right] < MSE\left[\hat{\sigma}^{2\,(DL/REML)}_\theta\right] < MSE\left[\hat{\sigma}^{2\,(HE)}_\theta\right].$$

With respect to the Type I error of the test of H_0: $\mu_\theta = 0$, we can make the following observations.

1. Equation 5 shows that the sampling variance of $\overline{ES}$ is underestimated by ignoring the variability of $\hat{\sigma}^2_\theta$ (and by considering the $\sigma^2_{\epsilon_i}$ values as known). This finding suggests that Type I errors should be inflated above the nominal $\alpha = .05$.
2. As discussed in section 5.5, the truncation of negative σ^2_θ estimates leads to a positive bias in the estimators. This bias will make Equation 2 too large on average and should lead to Type I errors that fall below the nominal $\alpha = .05$.

Therefore, the inflation in Type I error rates due to number 1 and the positive bias in σ_θ^2 estimates due to number 2 work against each other. For values of σ_θ^2 close to 0 (where truncation occurs more frequently and the positive bias due to number 2 is stronger), these two effects appear to cancel each other out, as Type I error rates were essentially nominal. On the other hand, for large σ_θ^2 values (where truncation is less prevalent), the effect due to number 2 disappears. Now, as shown in Tables 5 and 6, the effect of number 1 takes over and leads to inflated Type I error rates. Finally, we can add two additional conclusions.

3. As discussed earlier, the HS and ML estimators are negatively biased, which should lead to inflated Type I error rates over and beyond the inflation due to number 1. This result is apparent in Tables 5 and 6, where we note slightly more inflation in Type I error rates for the HS estimator (Type I error rates for the ML estimator were essentially identical).

4. As k increases, estimates of σ_θ^2 become more efficient, which has two effects: truncation becomes less prevalent (see section 5.5) and at the same time, the inflation due to number 1 decreases. Therefore, for large k, the two counteracting effects disappear, and we obtain nominal Type I error rates, even when σ_θ^2 is large (see again Tables 5 and 6).

7. Conclusion

Having to make general recommendations at this point requires a consideration of the trade-off between the bias, efficiency, and MSE of the estimators. The analytic results and simulations indicate that the Hunter–Schmidt and maximum likelihood estimators generally have lower MSE than the three (approximately) unbiased estimators. The results in section 5.4 suggest, however, that none of the estimators actually dominates the others in terms of MSE. Even Hedges' estimator, which usually has the highest MSE of all the estimators considered, will eventually surpass even the Hunter–Schmidt and maximum likelihood estimators if σ_θ^2 becomes large enough (although this seems to be unlikely to happen under realistic conditions). Moreover, as demonstrated in section 5.1, the Hunter–Schmidt and the maximum likelihood estimators are dominated by yet another estimator in terms of MSE when the sampling variances are homogeneous.

When considering bias and efficiency separately, then one should probably avoid the biased Hunter–Schmidt and maximum likelihood estimators because they can potentially provide quite misleading results. Their negative bias might lead researchers to (a) ignore possible heterogeneity in the effect sizes resulting from either random population effect sizes or moderator effects, and (b) to overstate the precision of the estimate of μ_θ. The remaining three estimators are all approximately unbiased, but the results in section 5.2 indicate some problems with the DerSimonian–Laird (and Hunter–Schmidt) estimator when the sampling variances of the effect sizes become small (i.e., when the sample sizes become large). In particular, it seems that the DerSimonian–Laird and Hunter–Schmidt estimators cannot reach the Cramer–Rao lower bound of σ_θ^2 estimators, regardless of how large k becomes, unless the sampling variances are homogeneous, which is unlikely to occur in practice. The restricted maximum likelihood estimator does not suffer from these problems and generally is substantially more efficient than Hedges' estimator. Also, the sampling variance of the restricted maximum likelihood estimator quickly approaches that of the regular maximum likelihood estimator, which is known to be asymptotically fully efficient. Therefore, the restricted maximum likelihood estimator strikes a good balance between unbiasedness and efficiency and, therefore, could be generally recommended.

However, even more problematic than using a suboptimal estimator for σ_θ^2 is the still all-too-common practice of simply presuming that the fixed-effects model, which assumes that $\sigma_\theta^2 = 0$, is appropriate when conducting a meta-analysis. In fact, it seems rather unlikely that all of the variability in a set of effect sizes could simply be accounted for by sampling error alone. Moderator variables can introduce systematic differences into the population effect sizes that can potentially be identified through appropriate statistical techniques (Hedges, 1994; Overton, 1998; Raudenbush, 1994). Alternatively, we can imagine a large number of miniscule moderator effects operating at the population level, where the effect of each moderator variable taken by itself is essentially indiscernible, but as a whole, these disturbances result in population effect sizes that follow an approximate normal distribution.

Regardless of the mechanism that introduces heterogeneity into the population effect sizes, it is important that such variability is properly identified. Claiming that the "true effect size" is equal to some value θ is hardly an accurate or useful statement when σ_θ^2 is large. Moreover, the National Research Council (1992) concluded that "the current practice of assuming a fixed effects model (. . .) [unless] a significance test of the nonhomogeneity of information sources rejects the hypothesis of homogeneity, is inefficient and can lead to understatement of uncertainty about the underlying effect of interest" (p. 186). Adopting a random- or mixed-effects model by default might be the better alternative. This would help to emphasize that there is potentially much more information to be gained from a meta-analysis than simply a single overall effect size. However, these models require estimation of the population heterogeneity. Within the context of the random-effects model, the results of this article should help researchers make a more informed decision regarding their choice of a population heterogeneity estimator.

References

Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations. *American Psychologist, 58,* 15–23.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin, 99,* 388–399.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine, 14,* 395–411.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing, 16,* 1190–1208.

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the Health Professions, 25,* 12–37.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Corbeil, R. R., & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics, 18,* 31–38.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7,* 177–188.

Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology, 49,* 275–306.

Friedman, L. (2000). Estimators of random effects variance components in meta-analysis. *Journal of Educational & Behavioral Statistics, 25,* 1–12.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72,* 320–338.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93,* 388–395.

Hedges, L. V. (1989). An unbiased correction for sampling error in validity generalization studies. *Journal of Applied Psychology, 74,* 469–477.

Hedges, L. V. (1994). Fixed effects models. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Sage.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- versus random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed-effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275–292.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.

Li, Y., Shi, L., & Roth, H. D. (1994). The bias of the commonly used estimate of variance in meta-analysis. *Communications in Statistics, Theory and Methods, 23,* 1063–1085.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and practice (with discussion). *Journal of the American Statistical Association, 78,* 47–65.

National Research Council (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academic Press.

Nelder, J. A., & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal, 7,* 308–313.

Normand, S. T. (1999). Tutorial in biostatistics. Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine, 18,* 321–359.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

Patterson, H. D., & Thompson, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometrics Conference,* 197–207.

Raudenbush, S. W. (1994). Random effects models. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Sage.

Searle, S. R. (1971). *Linear models.* New York: Wiley.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York: Wiley.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18,* 2693–2708.

Viechtbauer, W. (2002). *Bias of certain variance estimators in meta-analysis.* Unpublished manuscript, University of Illinois, Urbana-Champaign.

Author

WOLFGANG VIECHTBAUER is Assistant Professor, Department of Methodology and Statistics, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands; wolfgang.viechtbauer@stat.unimass.nl. His areas of interest are mixed-effects models, multilevel modeling, longitudinal data analysis, and meta-analysis.

Manuscript received May 5, 2003
Revision received September 15, 2003
Accepted April 20, 2004