Educational and Psychological Measurement

http://epm.sagepub.com

Weighting by Inverse Variance or by Sample Size in Random-Effects Meta-Analysis

Fulgencio Marín-Martínez and Julio Sánchez-Meca

Educational and Psychological Measurement 2010; 70; 56 originally published online Aug 31, 2009;
DOI: 10.1177/0013164409344534

Educational and Psychological
Measurement
70(1) 56–73

DOI: 10.1177/0013164409344534
http://epm.sagepub.com

Weighting by Inverse Variance or by Sample Size in Random-Effects Meta-Analysis

Fulgencio Marín-Martínez[1] and Julio Sánchez-Meca[1]

Abstract

Most of the statistical procedures in meta-analysis are based on the estimation of average effect sizes from a set of primary studies. The optimal weight for averaging a set of independent effect sizes is the inverse variance of each effect size, but in practice these weights have to be estimated, being affected by sampling error. When assuming a random-effects model, there are two alternative procedures for averaging independent effect sizes: Hunter and Schmidt's estimator, which consists of weighting by sample size as an approximation to the optimal weights; and Hedges and Vevea's estimator, which consists of weighting by an estimation of the inverse variance of each effect size. In this article, the bias and mean squared error of the two estimators were assessed via Monte Carlo simulation of meta-analyses with the standardized mean difference as the effect-size index. Hedges and Vevea's estimator, although slightly biased, achieved the best performance in terms of the mean squared error. As the differences between the values of both estimators could be of practical relevance, Hedges and Vevea's estimator should be selected rather than that of Hunter and Schmidt when the effect-size index is the standardized mean difference.

Keywords

meta-analysis, random-effects model, effect size, standardized mean difference

Meta-analysis is currently a consolidated methodology in the behavioral, educational, social, and health sciences for quantitatively integrating the results of a set of empirical studies on a common topic. To accomplish this objective, the results must be

[1]University of Murcia, Spain

Corresponding Author:
Fulgencio Marín-Martínez, Department of Basic Psychology and Methodology, Faculty of Psychology, Espinardo Campus, University of Murcia, P.O. Box 4021, 30100 Murcia, Spain
Email: fulmarin@um.es

converted into a common metric by means of an effect-size index, the standardized mean difference or a correlation coefficient being the most commonly used in the fields of education and psychology. Most of the statistical procedures in meta-analysis are usually based on the estimation of average effect sizes from the set of primary studies, with the research about the best procedure for averaging effect sizes being a point of remarkable methodological interest (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Petticrew & Roberts, 2006; Rosenthal, 1991).

In recent years, there has been some debate on the appropriateness of using fixed- versus random-effects models in meta-analysis, with a wide consensus supporting the use of the random-effects model as the most effective in achieving the scientific aims of a meta-analysis (Brockwell & Gordon, 2001; Erez, Bloom, & Wells, 1996; Field, 2001, 2003, 2005; Hedges & Vevea, 1998; Hunter & Schmidt, 2000; Kisamore & Brannick, 2008; National Research Council, 1992; Overton, 1998; Raudenbush, 1994; Sánchez-Meca & Marín-Martínez, 2008; Schmidt, 2008; Schmidt, Oh, & Hayes, 2009; Schulze, 2007). Basically, in the random-effects model the population effect size, δ, is assumed to be a continuous random variable with variance τ^2 and mean μ_δ, each i study estimating a different δ_i value from the distribution of δs. This means that there are two sources of variability in the effect sizes of the primary studies: within-study variance, σ_i^2, due to sampling error in the selection of the sample of each study; and between-studies variance, τ^2, due to sampling error in the selection of the studies in the meta-analysis. The fixed-effects model can be considered a particular case of the more general random-effects model, where the between-studies variance is zero, $\tau^2 = 0$, and the within-study variance, σ_i^2, is the only source of variability in the sample effects sizes of the studies.

The purpose of this work was to compare the statistical performance, by means of Monte Carlo simulation, of three procedures for estimating the average effect size of a set of independent effect sizes in a meta-analysis, through different weighting schemes proposed in the approaches of Hedges and Olkin (1985) and Hunter and Schmidt (2004). Here, we focus on the standardized mean difference as the effect-size index in the meta-analysis extending the results of our previous work in Sánchez-Meca and Marín-Martínez (1998), where the performance of weighting by the inverse variance and by the sample size was compared by assuming a fixed-effects model. In this article, we compare the procedures for estimating an average effect size by assuming a random-effects model, which is considered to be a more realistic model in meta-analysis.

The most accurate estimation of the mean parametric effect size in a random-effects model, μ_δ, should take into account the precision of the effect estimates in the individual studies, d_i, giving larger weights to the most accurate or least variable ones. Hedges (1983) showed that the optimal weight for averaging a set of k independent effect sizes is the inverse variance of each effect size, the uniformly minimum variance unbiased estimator of μ_δ being

$$\hat{\mu}_{\text{UMVU}} = \frac{\sum_i w_i d_i}{\sum_i w_i}, \tag{1}$$

with w_i being the optimal or true weights $w_i = 1/(\tau^2 + \sigma_i^2)$. However, in practice, the true variability of each effect size, $\tau^2 + \sigma_i^2$, is unknown and must be estimated using the empirical data in the meta-analysis. Thus, the weights applied in Equation 1 will be estimates of the optimal weights and will be affected by sampling error, giving an average effect size that is less accurate than the optimal estimator. The three procedures for averaging a set of independent effect sizes that we compare here are based on different proposals for estimating or approximating the optimal weights to be substituted in Equation 1.

Three Procedures for Averaging *d*s

Conceptually, we assume a meta-analysis with the standardized mean difference, d, as the effect-size index, where a set of k independent studies estimates a normal distribution of population effect sizes, $\delta \sim N(\mu_\delta, \tau^2)$. Each individual study consists of a two-group design (usually experimental vs. control) with a continuous outcome as the comparison criterion of the groups. Let us also assume that for each study a standardized mean difference estimating the magnitude of the difference between the groups is obtained by (see Hedges, 1981)

$$d_i = \left[1 - \frac{3}{4(n_{\text{E}i} + n_{\text{C}i}) - 9}\right] \frac{\bar{y}_{\text{E}i} - \bar{y}_{\text{C}i}}{S_i}, \tag{2}$$

where d_i is an approximately unbiased estimator of the population standardized mean difference corresponding to the ith study, δ_i; $\bar{y}_{\text{E}i}$ and $\bar{y}_{\text{C}i}$ are the experimental and control group means, respectively, of the ith study; $n_{\text{E}i}$ and $n_{\text{C}i}$ are the sample sizes; and S_i is the ith pooled within-group standard deviation computed through

$$S_i = \sqrt{\frac{(n_{\text{E}i} - 1)S_{\text{E}i}^2 + (n_{\text{C}i} - 1)S_{\text{C}i}^2}{n_{\text{E}i} + n_{\text{C}i} - 2}}, \tag{3}$$

$S_{\text{E}i}^2$ and $S_{\text{C}i}^2$ being the unbiased variances of the two groups of the ith study.

The within-study variance of d_i is given by

$$\sigma_i^2 = \frac{n_{\text{E}i} + n_{\text{C}i}}{n_{\text{E}i} n_{\text{C}i}} + \frac{\delta_i^2}{2(n_{\text{E}i} + n_{\text{C}i})}. \tag{4}$$

In practice, as the population effect size of the ith study, δ_i, is unknown, the within-study variance of each d_i is usually estimated by substituting the sample d_i for δ_i in Equation 4 (Hedges & Olkin, 1985):

$$\hat{\sigma}_i^2 = \frac{n_{\text{E}i} + n_{\text{C}i}}{n_{\text{E}i} n_{\text{C}i}} + \frac{d_i^2}{2(n_{\text{E}i} + n_{\text{C}i})}, \tag{5}$$

where $\hat{\sigma}_i^2$ is the estimated within-study variance of d_i.

The between-studies variance estimator of d_i most usually applied in the meta-analytic literature is the one proposed by DerSimonian and Laird (1986), which is based on the moments method and consists of estimating the population between-studies variance by

$$\hat{\tau}^2_{\mathrm{DL}} = \frac{Q - (k-1)}{c}, \tag{6}$$

where k is the number of studies in the meta-analysis, Q is the heterogeneity statistic usually applied in meta-analysis to test the homogeneity hypothesis of the effect sizes (Hedges & Olkin, 1985), given by

$$Q = \sum_i \hat{w}_i^{\mathrm{FE}} (d_i - \hat{\mu}_{\mathrm{FE}})^2, \tag{7}$$

with $\hat{w}_i^{\mathrm{FE}} = 1/\hat{\sigma}_i^2$ being the estimated inverse variance of the ith study assuming a fixed-effects model, $\hat{\sigma}_i^2$ being computed with Equation 5; $\hat{\mu}_{\mathrm{FE}}$ is the estimated mean effect size also assuming a fixed-effects model, given by Equation 11; and c is given by

$$c = \sum_i \hat{w}_i^{\mathrm{FE}} - \frac{\sum_i (\hat{w}_i^{\mathrm{FE}})^2}{\sum_i \hat{w}_i^{\mathrm{FE}}}. \tag{8}$$

When the estimated between-studies variance, $\hat{\tau}^2_{\mathrm{DL}}$, computed through Equation 6, gives a negative value, it must be truncated to zero.

Another alternative between-studies variance estimator[1] is based on maximum likelihood estimation and consists of an iterative procedure given by (Viechtbauer, 2005),

$$\hat{\tau}^2_{\mathrm{ML}} = \frac{\sum_i \hat{w}_i^2 [(d_i - \hat{\mu}_{\mathrm{ML}})^2 - \hat{\sigma}_i^2]}{\sum_i \hat{w}_i^2}, \tag{9}$$

with $\hat{w}_i = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$, where $\hat{\sigma}_i^2$ is given by Equation 5, $\hat{\tau}^2$ is initially estimated by Equation 6 or setting $\hat{\tau}^2 = 0$ and $\hat{\mu}_{ML}$ is given by

$$\hat{\mu}_{\mathrm{ML}} = \frac{\sum_i \hat{w}_i d_i}{\sum_i \hat{w}_i}. \tag{10}$$

In each iteration of Equations 9 and 10, the estimate of τ^2 has to be checked to avoid having negative values truncating it to zero. Convergence is usually achieved within fewer than 10 iterations.

Bearing in mind the general Equation 1 for averaging a set of k independent effect sizes d_i, the three procedures to be compared in this article are based on three different estimations of the optimal weights, $w_i = 1/(\tau^2 + \sigma_i^2)$. The first of the procedures (FE average) was proposed by Hedges and colleagues (Hedges, 1983; Hedges & Olkin, 1985; Hedges & Vevea, 1998) and was devised for meta-analyses that assume a fixed-effects model. In these conditions, the average of k independent standardized mean differences is given by

$$\hat{\mu}_{FE} = \frac{\sum_i \hat{w}_i^{FE} d_i}{\sum_i \hat{w}_i^{FE}}, \qquad (11)$$

where the weight for estimating the inverse variance of each d value under a fixed-effects model, $\hat{w}_i^{FE}$, is a function of the within-study variance only (the between-studies variance is assumed to be null) and is computed through $\hat{w}_i^{FE} = 1/\hat{\sigma}_i^2$, with $\hat{\sigma}_i^2$ given by Equation 5.

Although, as we mentioned above, it is considered unrealistic to assume a fixed-effects model for most meta-analyses, the random-effects model being the most appropriate choice, we have included this first FE procedure for several reasons. On the one hand, some authors such as Hedges and Vevea (1998) contend that the fixed-effects model can be used when the meta-analyst wishes to generalize the results to a population of studies with the same characteristics as those included in the meta-analysis, the objective of the meta-analysis being a conditional inference. On the other hand, in practice many of the meta-analyses published to date assume a fixed-effects model and compute the FE average as in Equation 11. Thus, including the FE procedure in our simulation study may help assess the accuracy of the results in previous meta-analyses, as it is a point of reference for better assessing and comparing the performance of the other two procedures.

The second of the procedures or averages (HV procedure) was also proposed by Hedges and colleagues (Hedges & Olkin, 1985; Hedges & Vevea, 1998), although in this case assuming a random-effects model. In meta-analyses where a random-effects model is assumed, the average of k independent standardized mean differences is given by

$$\hat{\mu}_{HV} = \frac{\sum_i \hat{w}_i^{HV} d_i}{\sum_i \hat{w}_i^{HV}}, \qquad (12)$$

where the weight estimating the inverse variance of each d value, $\hat{w}_i^{HV}$, is a function of both the within-study and the between-studies variances, being computed through $\hat{w}_i^{HV} = 1/(\hat{\sigma}_i^2 + \hat{\tau}_{DL}^2)$, with $\hat{\sigma}_i^2$ and $\hat{\tau}_{DL}^2$ given by Equations 5 and 6, respectively. We chose DerSimonian and Laird's between-studies variance estimator, $\hat{\tau}_{DL}^2$, as it is the one that is most usually applied in meta-analysis. However, we also included in our Monte Carlo simulation a variant of the HV procedure, that of HVML, where the between-studies variance estimator was the maximum likelihood estimator, $\hat{\tau}_{ML}^2$, in Equation 9.

Finally, the third procedure (HS average) was proposed by Hunter and Schmidt (1990, 2004) and is a simpler procedure than the other two, which consists of averaging the set of independent effect sizes and weighting by the sample size of each study. This average is given by

$$\hat{\mu}_{HS} = \frac{\sum_i N_i d_i}{\sum_i N_i}, \qquad (13)$$

where N_i is the global sample size of the ith study: $N_i = n_{Ei} + n_{Ci}$. Assuming a random-effects model, these authors consider that although the sample sizes are not the optimal weights for the standardized mean difference, these N_i closely approximate the inverse variances of the sample *d*s, being less affected by sampling error than the estimated weights proposed in the Hedges approach: $\hat{w}_i^{FE}$ and $\hat{w}_i^{HV}$ in Equations 11 and 12, respectively. Furthermore, although weighting by sample size gives a slightly less efficient estimator of the mean than the two Hedges procedures, the HS estimator is practically unbiased in comparison with the slight negative bias in the FE and HV averages (Hedges, 1983; Sánchez-Meca & Marín-Martínez, 1998; Schmidt et al., 2009).

Review of Previous Simulation Studies

The three procedures that we have just presented can be considered as two general approaches to average a set of independent effect sizes: the approach by Hedges, which consists of weighting by an estimation of the inverse variance of each effect size (FE and HV procedures); and the Hunter and Schmidt approach, which consists of weighting by the sample size of each study (HS procedure). On the one hand, according to statistical theory, it is expected that the Hedges approach will be the most accurate, as it is clearly stated that the optimal weight is the inverse variance of each effect size, $w_i = 1/(\tau^2 + \sigma_i^2)$, but in practice the sampling error in the estimated weights, $\hat{w}_i^{FE}$ and $\hat{w}_i^{HV}$, affects the accuracy of the Hedges averages. On the other hand, weighting by the sample size (the HS approach) is a reasonable alternative to the approach by Hedges, as the sample sizes are directly proportional to the optimal weights and are less affected by sampling error than the estimators $\hat{w}_i^{FE}$ or $\hat{w}_i^{HV}$.

Several Monte Carlo studies have empirically compared the performance of the two weighting strategies: based on the sample size or on the optimal weights as estimated from the Hedges approach. Most often, by simulating a random-effects model, the Monte Carlo studies compared the random-effects version of the Hedges procedure with that of Hunter and Schmidt. However, most of these studies were based on meta-analyses where the effect-size index was the Pearson correlation coefficient, *r*, and their results were obscured by the use of the nonlinear Fisher's *z* transformation of correlations in the approach by Hedges. The general conclusion of such studies was that weighting by sample size gave the most accurate estimates of the average correlation coefficient, although the differences between the different weighting procedures can be considered irrelevant in practice (Field, 2001, 2005; Hall & Brannick, 2002; Schulze, 2004). Taking into account that most of the simulation studies gave their results (i.e., bias and efficiency) in terms of averages based on thousands of replies, differences between the procedures that may seem irrelevant could have important practical implications at the level of a single meta-analysis.

To the best of our knowledge, Sánchez-Meca and Marín-Martínez (1998) is the only Monte Carlo study that systematically compared the statistical performance between the averages resulting from weighting by sample size or by the inverse variance, when the effect-size index is the standardized mean difference. One of the main limitations of our study was that the simulation was run assuming a fixed-effects model that represents just one particular condition ($\tau^2 = 0$) of the more realistic random-effects model. In the aforementioned article, we compared only the fixed-effects version of the Hedges procedure (FE procedure in Equation 11) with Hunter and

Schmidt's proposal of weighting by sample size (HS procedure in Equation 13) without including the HV procedure in Equation 12. Our main conclusion was that although the FE average was slightly more efficient than the HS average, the FE estimator also showed a slight negative bias in all the simulated conditions, whereas the HS average was kept practically unbiased. Anyway, the differences between both estimators (either in terms of bias or efficiency) seemed to be small and of negligible practical relevance.

In the current article, we extend the findings of our previous study by simulating the more realistic conditions of a random-effects model and adding the random-effects version of the Hedges procedure (the HV procedure in Equation 12). We expected that under a random-effects model, the larger the between-studies variance, the more pronounced the differences between the averaging procedures. In these circumstances, the selection of the most accurate procedure could be of practical interest.

Method

A Monte Carlo simulation study of meta-analyses was programmed in GAUSS (Aptech Systems, 2001) with the standardized mean difference as the effect-size index. To simulate each individual study in a meta-analysis, we defined a two-group design (e.g., experimental vs. control) and a continuous outcome. Assuming homogeneous variances in the experimental and control populations of each study, the parametric standardized mean difference for the ith study, δ_i, was defined as

$$\delta_i = \frac{\mu_{\mathrm{E}i} - \mu_{\mathrm{C}i}}{\sigma_{w_i}}, \qquad (14)$$

where $\mu_{\mathrm{E}i}$ and $\mu_{\mathrm{C}i}$ were the population means for the experimental and the control groups in the ith study, and σ_{w_i} was the common population standard deviation of the ith study. For each study, normal distributions in the experimental and control groups were assumed for the continuous outcome.

Under a random-effects model, the population standardized mean differences, δ_i, were normally distributed with mean μ_δ and variance τ^2, that is, $\delta_i \sim N(\mu_\delta, \tau^2)$. From the normal distribution of δ_i values, collections of k independent studies were randomly generated to simulate a meta-analysis. Once a δ_i value was randomly selected, the ith study was simulated by generating two normal distributions (for the experimental and control groups) with means of $\mu_{\mathrm{E}i} = \delta_i$ and $\mu_{\mathrm{C}i} = 0$ and common standard deviation $\sigma_{w_i} = 1$. Next, pairs of independent samples (experimental and control) were randomly selected from the two distributions of the continuous outcome with sample sizes $n_{\mathrm{E}i} = n_{\mathrm{C}i}$. For each of the k studies in a meta-analysis, the standardized mean difference, d_i, and its estimated within-study variance, $\hat{\sigma}_i^2$, were computed by applying Equations 2 and 5, respectively. Then, with the data of each simulated meta-analysis, the between-studies variance, τ^2, was estimated through Equations 6 and 9, and the three averages $\hat{\mu}_{\mathrm{FE}}$, $\hat{\mu}_{\mathrm{HV}}$, and $\hat{\mu}_{\mathrm{HS}}$ were computed by applying Equations 11, 12, and 13, respectively. Furthermore, we also computed a variant of the HV procedure, the average $\hat{\mu}_{HVML}$, where the between-studies variance in the weights was estimated through the maximum likelihood estimator in Equation 9. Also, taking advantage of the fact that in our simulations the optimal weights, $w_i = 1/(\tau^2 + \sigma_i^2)$, are known values, the uniformly minimum variance unbiased estimator of μ_δ in Equation 1, $\hat{\mu}_{UMVU}$, was also computed for comparison purposes.

To examine and compare the performance of the FE, HV, HS, HVML, and UMVU averages, we manipulated the following factors in the simulation. First, the average parametric standardized mean difference, μ_δ, was manipulated with values 0.2, 0.5, and 0.8, which can be considered to be effects of low, medium, and high magnitude, respectively (Cohen, 1988). Second, the population between-studies variance, τ^2, was manipulated with values 0, 0.04, 0.08, 0.16, and 0.32, the value $\tau^2 = 0$ simulating a fixed-effects model, and the other values simulating the more realistic random-effects model. Second, the number of studies, k, in each meta-analysis was manipulated, with values 5, 10, 20, 40, and 100. Finally, the average sample size of the studies included in the meta-analyses was manipulated with values 30, 50, 80, and 100.

The sample size distribution used in our simulations was obtained from a review of 30 real meta-analyses published in 18 international journals in the fields of education and psychology, where the Pearson skewness index of the sample size distribution gave a value of +1.464 (for more details, see Sánchez-Meca & Marín-Martínez, 1998). Thus, four vectors of five sample sizes each were selected, averaging 30, 50, 80, or 100, using the skewness index given above, with the following values for N_i: (12, 16, 18, 20, 84), (32, 36, 38, 40, 104), (62, 66, 68, 70, 134), and (82, 86, 88, 90, 154). Each vector of five samples was then replicated either 2, 4, 8, or 20 times to generate meta-analyses of $k = 5$, 10, 20, 40, and 100 studies, respectively. For each simulated study, the sample sizes for the experimental and control groups were equal ($n_E = n_C$), with $N = n_E + n_C$. For example, the sample size vector (12, 16, 18, 20, 84) meant that the experimental and control groups had sample sizes of $n_E = n_C =$ 6, 8, 9, 10, and 42, respectively.

The values shown above for the manipulated factors μ_δ (the average population standardized mean difference), τ^2 (the population between-studies variance), k (the number of studies), and $\overline{N}$ (the average sample size) were selected in an attempt to approximate those usually found in real meta-analyses with the d index. In total, 300 conditions were manipulated [3(μ_δ values) $\times$ 5(τ^2 values) $\times$ 5(k values) $\times$ 4(N values)], and for each of them, 10,000 replicates (meta-analyses) were performed. The FE, HV, HS, HVML, and UMVU averages were computed in each of these replications. On the one hand, the bias of each of the five averages was assessed as the difference between the mean of the 10,000 empirical values of each estimator and the average parametric standardized mean difference μ_δ. On the other hand, the variability of the averages was assessed by calculating the mean squared difference (mean squared error [MSE]) for each of the five estimators with respect to the true value μ_δ, across the 10,000 replications of one single condition. The MSE is the sum of the variance of the estimator in the 10,000 replications and its squared bias, being a global index of the accuracy of each estimator that includes both its variability and its bias. The mathematical computations of both the bias and the MSE of each estimator are presented below.

Define $\hat{\mu}^j$ as any of the averages or estimators of the mean effect size, and $\mathrm{AVE}(\hat{\mu}^j)$ and $\mathrm{VAR}(\hat{\mu}^j)$ as the mean and variance, respectively, of the 10,000 empirical values of $\hat{\mu}^j$ in a particular condition, that is,

$$\mathrm{AVE}\,(\hat{\mu}^j) = \frac{\sum_i \hat{\mu}_i^j}{10,000} \tag{15}$$

and

$$\mathrm{VAR}\,(\hat{\mu}^j) = \frac{\sum_i [\hat{\mu}_i^j - \mathrm{AVE}(\hat{\mu}^j)]^2}{10,000}. \tag{16}$$

Thus, on the one hand, the bias of the estimator $\hat{\mu}^j$ was given by

$$\mathrm{BIAS}\,(\hat{\mu}^j) = \mathrm{AVE}\,(\hat{\mu}^j) - \mu_\delta, \tag{17}$$

where $\mathrm{AVE}\,(\hat{\mu}^j)$ was given by Equation 15 and μ_δ was the population average standardized mean difference. On the other hand, the MSE of the estimator $\hat{\mu}^j$ around the parameter μ_δ was given by

$$\mathrm{MSE}\,(\hat{\mu}^j) = \frac{\sum_i (\hat{\mu}_i^j - \mu_\delta)^2}{10,000}, \tag{18}$$

the relationship among the MSE and the bias of each estimator being

$$\mathrm{MSE}\,(\hat{\mu}^j) = \mathrm{VAR}(\hat{\mu}^j) + [\mathrm{BIAS}(\hat{\mu}^j)]^2, \tag{19}$$

where $\mathrm{MSE}\,(\hat{\mu}^j)$, $\mathrm{VAR}(\hat{\mu}^j)$, and $\mathrm{BIAS}(\hat{\mu}^j)$ are given by Equations 18, 16, and 17, respectively.

Results

Tables 1 and 2 show the bias and MSE, respectively, of the three estimators, $\hat{\mu}^{FE}$ (FE average proposed by Hedges in meta-analyses where a fixed-effects model is assumed), $\hat{\mu}^{HV}$ (HV average proposed by Hedges in meta-analyses where a random-effects model is assumed), and $\hat{\mu}^{HS}$ (HS average proposed by Hunter and Schmidt for all meta-analyses), through the 100 simulated conditions in the meta-analyses with $\mu_\delta = 0.5$.

Due to space limitations we only included the data of the three estimators when the average parametric standardized mean difference was of a medium magnitude[2], as the pattern of results was similar to that with $\mu_\delta = 0.2$ and $\mu_\delta = 0.8$. Through all the 300 simulated conditions, Figures 1 and 2 show the average bias and MSE values, respectively, of the FE, HV, and HS procedures, as a function of the average population effect size, the between-studies variance, the number of studies, and the average sample size. Furthermore, we included in Figures 1 and 2 the HVML version of the HV procedure, where the between-studies variance in the weights was estimated through the maximum likelihood estimator in Equation 9, and the UMVU procedure in Equation 1, where the optimal weights were applied.

As expected, the bias in the two Hedges procedures, FE and HV averages, was negative in all the conditions, the average bias values being $-.0197$ and $-.0099$, respectively. The HVML version of HV, with an average bias of $-.0105$, showed very similar bias values to those in the HV procedure. In contrast, the bias in the HS procedure was always lower than that in the FE and HV (or HVML) procedures, with a practically null average value of +.0000 through all the conditions. Furthermore, none of the manipulated parameters in the simulation, the average population effect size (μ_δ), the between-studies variance (τ^2), the number of studies (k), or the average sample size ($\overline{N}$) seemed to influence the negligible values of the bias in the HS estimator, which remained practically unbiased in all the conditions. The UMVU procedure, which was also expected to be practically unbiased, showed an average bias value of $-.0039$, slightly larger than that in the HS estimator (see Figure 1 and Table 1).

The negative bias in the FE and HV (or HVML) estimators was more pronounced as the magnitude of the average population effect size increased (see Figure 1A). Furthermore, the negative bias in the FE estimator was systematically larger than that of the HV (or HVML) estimator, the difference between the estimators being larger as the between-studies variance increased (see Figure 1B). Note that whereas the FE procedure was proposed for meta-analyses that assume a fixed-effects model or a null between-studies variance ($\tau^2 = 0$), most of the conditions in the simulation assumed a random-effects model with $\tau^2 > 0$. Consequently, as the between-studies variance increased, the negative bias in the FE procedure clearly increased. However, the influence of τ^2 on the bias in the HV (or HVML) estimator was less pronounced and it was modulated by the sample size: With small sample sizes the bias tended to slightly increase as τ^2 also increased, whereas with larger sample sizes this trend tended to disappear. Moreover, the sample size affected the bias of the FE and HV (or HVML) estimators in the expected direction: the larger the sample size, the lower the bias, this trend being more pronounced in the HV (or HVML) estimator. Finally, the number of studies scarcely affected the bias of any of the estimators (see Figure 1).

The average MSE values of the HV, FE, and HS procedures, computed through all the conditions, were .0162, .0177, and .0186, respectively. The HVML version of HV, with an average MSE of .0161, showed very similar MSE values to those in the HV procedure. As a consequence, the HV (or HVML) estimator was the most efficient or accurate overall, followed by the FE and HS procedures. A similar trend was observed in most of the particular simulated conditions, the HV (or HVML) estimator being usually more efficient than the HS one. For example, in meta-analyses with $\tau^2 = .08$, $k = 20$, and $\overline{N} = 30$, the MSE values were .0130, .0142 (relative increase of 9.23% in respect to .0130), and .0147 (relative increase of 13.08% in respect to .0130) for the HV, FE, and HS estimators, respectively (see Table 2). Moreover, the manipulated factors in the simulation affected the MSE in the expected direction: the MSE values of the three estimators increased as the number of studies (k) and sample sizes ($\overline{N}$) decreased, and as the between-studies variance (τ^2) increased, the differences among the three estimators being more pronounced, with the minimum MSE for the HV (or HVML) procedure and the largest MSE for the HS procedure (see Figure 2).

Table 1. Bias of the FE, HV, and HS Estimators as a Function of the Between-Studies Variance, the Number of Studies, and the Average Sample Size ($\mu_\delta = 0.5$)

		$\tau^2 = 0.0$			$\tau^2 = 0.04$			$\tau^2 = 0.08$			$\tau^2 = 0.16$			$\tau^2 = 0.32$		
k	$\overline{N}$	FE	HV	HS	FE	HV	HS	FE	HV	HS	FE	HV	HS	FE	HV	HS
5	30	−.0135	−.0133	−.0004	−.0141	−.0113	.0021	−.0150	−.0113	.0036	−.0179	−.0114	.0059	−.0292	−.0180	.0044
	50	−.0079	−.0057	.0001	−.0113	−.0073	.0001	−.0148	−.0080	−.0001	−.0233	−.0087	−.0011	−.0305	−.0071	.0020
	80	−.0051	−.0039	−.0001	−.0106	−.0066	−.0020	−.0120	−.0051	.0001	−.0229	−.0097	−.0045	−.0317	−.0060	−.0009
	100	−.0046	−.0034	−.0006	−.0074	−.0036	.0003	−.0106	−.0032	.0007	−.0189	−.0046	−.0009	−.0314	−.0070	−.0021
10	30	−.0125	−.0117	.0024	−.0163	−.0147	.0023	−.0227	−.0201	−.0007	−.0338	−.0240	−.0055	−.0359	−.0202	.0047
	50	−.0075	−.0064	.0015	−.0123	−.0086	.0008	−.0157	−.0088	.0009	−.0254	−.0126	−.0015	−.0359	−.0118	.0010
	80	−.0058	−.0049	−.0002	−.0111	−.0073	−.0015	−.0129	−.0054	.0007	−.0216	−.0060	−.0004	−.0337	−.0057	.0005
	100	−.0051	−.0044	−.0006	−.0085	−.0043	.0002	−.0138	−.0058	−.0011	−.0203	−.0056	−.0002	−.0321	−.0034	.0018
20	30	−.0166	−.0163	−.0010	−.0205	−.0188	−.0010	−.0237	−.0200	−.0005	−.0313	−.0236	−.0011	−.0441	−.0238	−.0009
	50	−.0094	−.0087	.0000	−.0129	−.0096	.0007	−.0180	−.0116	−.0004	−.0257	−.0118	−.0005	−.0386	−.0120	.0004
	80	−.0051	−.0045	.0008	−.0093	−.0052	.0009	−.0151	−.0069	−.0008	−.0232	−.0076	−.0010	−.0368	−.0075	−.0007
	100	−.0051	−.0045	−.0003	−.0083	−.0041	.0008	−.0125	−.0041	.0007	−.0219	−.0057	−.0008	−.0334	−.0037	.0021
40	30	−.0158	−.0158	.0001	−.0197	−.0185	.0004	−.0248	−.0209	−.0008	−.0311	−.0223	.0006	−.0447	−.0249	.0005
	50	−.0099	−.0095	−.0003	−.0127	−.0093	.0014	−.0186	−.0117	−.0005	−.0250	−.0105	.0011	−.0399	−.0116	.0004
	80	−.0057	−.0053	.0003	−.0106	−.0066	.0000	−.0155	−.0072	−.0007	−.0220	−.0061	.0008	−.0364	−.0059	.0011
	100	−.0047	−.0043	.0001	−.0095	−.0051	−.0001	−.0142	−.0056	−.0005	−.0230	−.0064	−.0012	−.0365	−.0056	−.0003
100	30	−.0162	−.0161	.0001	−.0212	−.0198	−.0007	−.0245	−.0207	.0000	−.0316	−.0220	.0007	−.0467	−.0257	−.0007
	50	−.0098	−.0095	−.0001	−.0142	−.0108	.0000	−.0186	−.0115	−.0001	−.0273	−.0124	−.0009	−.0404	−.0119	.0005
	80	−.0062	−.0059	.0000	−.0104	−.0062	.0003	−.0153	−.0070	−.0003	−.0226	−.0064	.0006	−.0380	−.0074	−.0002
	100	−.0054	−.0051	−.0004	−.0097	−.0054	−.0002	−.0140	−.0055	−.0002	−.0221	−.0054	.0000	−.0368	−.0055	−.0001

Note: k = number of studies; $\overline{N}$ = average sample size; τ^2 = between-studies variance; FE = fixed-effects version of the Hedges estimator; HV = random-effects version of the Hedges estimator; HS = Hunter and Schmidt estimator.

Table 2. Mean Squared Error of the FE, HV, and HS Estimators as a Function of the Between-Studies Variance, the Number of Studies, and the Average Sample Size ($\mu_\delta = 0.5$)

		$\tau^2 = 0.0$			$\tau^2 = 0.04$			$\tau^2 = 0.08$			$\tau^2 = 0.16$			$\tau^2 = 0.32$		
k	$\overline{N}$	FE	HV	HS	FE	HV	HS	FE	HV	HS	FE	HV	HS	FE	HV	HS
5	30	.0270	.0306	.0284	.0424	.0432	.0447	.0564	.0531	.0594	.0815	.0700	.0866	.1341	.1010	.1464
	50	.0163	.0170	.0168	.0266	.0265	.0276	.0359	.0342	.0375	.0568	.0508	.0603	.0924	.0821	.1016
	80	.0104	.0106	.0106	.0187	.0186	.0192	.0268	.0262	.0279	.0438	.0425	.0462	.0743	.0739	.0819
	100	.0083	.0085	.0085	.0168	.0167	.0172	.0251	.0250	.0260	.0411	.0407	.0435	.0733	.0741	.0802
10	30	.0132	.0140	.0140	.0201	.0200	.0213	.0279	.0261	.0295	.0405	.0340	.0430	.0657	.0501	.0731
	50	.0081	.0083	.0084	.0129	.0127	.0135	.0178	.0170	.0186	.0283	.0256	.0300	.0456	.0409	.0503
	80	.0051	.0052	.0052	.0092	.0091	.0095	.0137	.0134	.0143	.0216	.0211	.0229	.0375	.0370	.0412
	100	.0040	.0041	.0041	.0082	.0082	.0084	.0122	.0122	.0126	.0195	.0194	.0206	.0352	.0356	.0386
20	30	.0068	.0070	.0070	.0106	.0103	.0109	.0142	.0130	.0147	.0208	.0171	.0218	.0341	.0254	.0370
	50	.0041	.0041	.0042	.0067	.0066	.0069	.0092	.0087	.0095	.0141	.0126	.0147	.0236	.0206	.0254
	80	.0026	.0026	.0026	.0046	.0045	.0047	.0070	.0067	.0071	.0113	.0108	.0117	.0190	.0181	.0202
	100	.0021	.0021	.0021	.0042	.0041	.0042	.0062	.0060	.0063	.0103	.0099	.0106	.0177	.0175	.0190
40	30	.0036	.0036	.0036	.0054	.0052	.0054	.0072	.0064	.0072	.0107	.0087	.0108	.0175	.0127	.0180
	50	.0021	.0021	.0020	.0034	.0032	.0034	.0048	.0043	.0047	.0072	.0062	.0073	.0124	.0102	.0124
	80	.0013	.0013	.0013	.0025	.0024	.0025	.0036	.0033	.0035	.0057	.0053	.0057	.0101	.0091	.0101
	100	.0010	.0010	.0010	.0022	.0021	.0022	.0033	.0031	.0032	.0054	.0050	.0053	.0099	.0092	.0098
100	30	.0016	.0016	.0014	.0024	.0023	.0021	.0034	.0029	.0030	.0049	.0037	.0043	.0083	.0054	.0071
	50	.0009	.0009	.0008	.0015	.0014	.0014	.0021	.0018	.0018	.0033	.0026	.0029	.0060	.0042	.0051
	80	.0005	.0005	.0005	.0010	.0010	.0010	.0016	.0014	.0014	.0026	.0021	.0023	.0051	.0038	.0042
	100	.0004	.0004	.0004	.0009	.0009	.0009	.0014	.0012	.0013	.0025	.0020	.0021	.0048	.0036	.0039

Note: k = number of studies; $\overline{N}$ = average sample size; τ^2 = between-studies variance; FE = fixed-effects version of the Hedges estimator; HV = random-effects version of the Hedges estimator; HS = Hunter and Schmidt estimator.

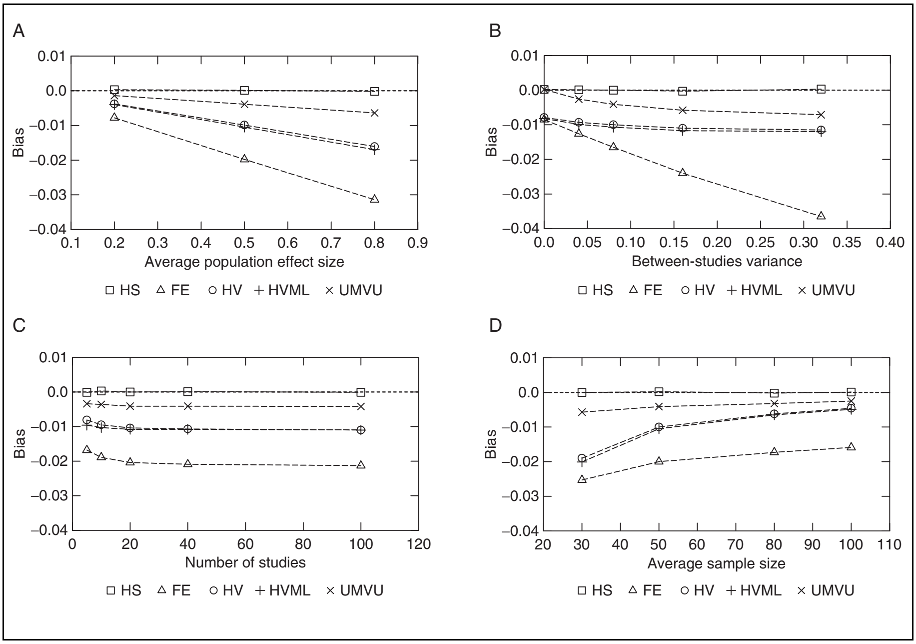

Figure 1. Average bias values of the estimators, as a function of the average population effect size, the between-studies variance, the number of studies, and the average sample size

Note: FE = fixed-effects version of the Hedges estimator; HV = random-effects version of the Hedges estimator; HVML = random-effects version of the Hedges estimator with τ^2 estimated through maximum likelihood; HS = Hunter and Schmidt estimator; UMVU = uniformly minimum variance unbiased estimator.

Therefore, in meta-analyses with large between-studies variance, small sample sizes, and a low number of studies, the differences in accuracy among the HV, FE, and HS estimators were more evident. The optimal UMVU procedure, which was expected to show the minimum MSE values, presented an average MSE value of .0163, very similar to that in the HV (or HVML) estimator (see Figure 2).

Discussion

At this point, the main question in this work remains: Which of the estimators or procedures for averaging a set of independent standardized mean differences is the most suitable in a meta-analysis? Although the nature of the between-studies variance estimator, $\hat{\tau}^2_{DL}$ or $\hat{\tau}^2_{ML}$, scarcely affected the performance of the Hedges and Vevea procedure (weighting by the inverse variance assuming a random-effects model), the HV and HVML averages showing very similar bias and MSE values in all the conditions, more remarkable differences were found among the three main procedures FE, HV,

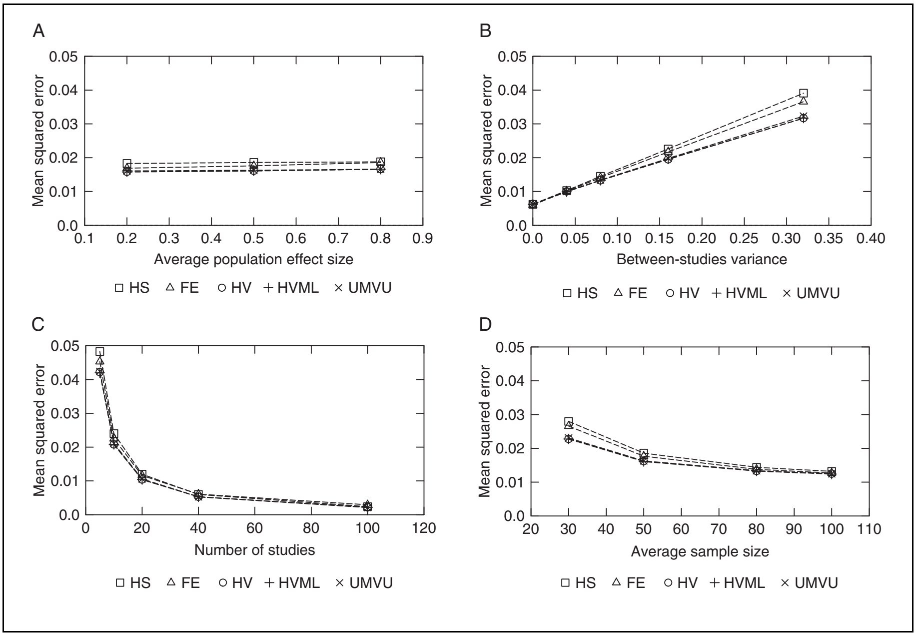

Figure 2. Average mean squared error values of the estimators, as a function of the average population effect size, the between-studies variance, the number of studies, and the average sample size

Note: FE = fixed-effects version of the Hedges estimator; HV = random-effects version of the Hedges estimator; HVML = random-effects version of the Hedges estimator with τ^2 estimated through maximum likelihood; HS = Hunter and Schmidt estimator; UMVU = uniformly minimum variance unbiased estimator.

and HS. In respect to the FE estimator, which was specifically designed for meta-analyses assuming a fixed-effects model, the results showed a generally poor performance in most of the conditions. On the one hand, the FE estimator systematically showed the largest negative bias, which increased as τ^2 increased and $\overline{N}$ decreased in a more pronounced way than the other estimators. On the other hand, the MSE values of the FE estimator were usually larger than those of the HV estimator. Only in meta-analyses with $\tau^2 = 0$, where the real underlying model is the fixed-effects one, the FE estimator showed MSE values similar to and even lower than those of the HV and HS procedures. However, as in practice most of the meta-analyses will have a between-studies variance that is greater than zero, the FE estimator will not usually be the best option. Therefore, we advise against the use of the FE estimator in meta-analysis, the HV estimator being a better option than the FE one, even in the extreme conditions of a null between-studies variance, where the HV procedure was systematically less biased and with a similar MSE than the FE procedure.

In respect to the comparison between the HV and HS estimators, an advantageous property of the latter was that it proved to be a practically unbiased estimator in all the conditions, whereas the HV estimator showed a systematic slight negative bias. However, the HS estimator also showed MSE values usually larger than those of the HV estimator, the former being a more variable estimator around the parameter than the latter. Taking into account that the MSE is the average of the squared distances between the estimator and the parameter (see Equation 18), showing the global error of an estimator in respect to its parameter, the best estimator should be the one with the minimum MSE. Furthermore, the MSE is the most global measure of accuracy of an estimator, which includes the bias as a component, as the MSE could be also computed as the sum of the squared bias of the estimator and its empirical variability (see Equation 19). Then the choice between the HV and the HS estimators should primarily be guided by their MSE values and subsequently by their bias values.

A careful review of all the particular conditions with $\mu_\delta = 0.5$ in our simulations (see Table 2) revealed that the only conditions where the HS estimator showed a lower MSE than the HV estimator were in meta-analyses with $\tau^2 = 0$, $k = 5$, and $\overline{N} = 30$; $\tau^2 = 0$, $k = 5$, and $\overline{N} = 50$; $\tau^2 = 0$, $k = 40$, and $\overline{N} = 50$; $\tau^2 = 0$, $k = 100$, and $\overline{N} = 30$; $\tau^2 = 0$, $k = 100$, and $\overline{N} = 50$; and $\tau^2 = .04$, $k = 100$, and $\overline{N} = 30$. Note that these are very unusual conditions in practice, as the between-studies variance is usually larger than 0, and in the condition where $\tau^2 = .04$, $k = 100$ is a very large number of studies.

It should also be noted that, as the HS average is a practically unbiased estimator, it was expected that changing the conditions of the meta-analyses to diminish the variability of the estimators, the HS estimator should become the one with the minimum MSE in the long run. Thus, in additional simulations with $\mu_\delta = 0.5$ we found that, for example, in meta-analyses with $\tau^2 = .16$ and $\overline{N} = 30$ a minimum of $k = 600$ studies was needed to make the MSE of the HS estimator lower than that of the HV estimator. Also in meta-analyses with the same $\tau^2 = .16$ and a lower average sample size, $\overline{N} = 12$, a lower minimum of $k = 200$ studies was needed to achieve the same objective. However, neither 600 nor 200 studies with an average sample size of 12 units are usual conditions in real meta-analyses. Thus, we conclude that in the most usual conditions of real meta-analyses, the HV estimator, although slightly biased, shows a better performance than the HS estimator, as the MSE values of the HV estimator were lower than those of the HS estimator.

We also included in our simulations the uniformly minimum variance unbiased (UMVU) estimator of μ_δ (Equation 1), where the optimal weights were applied to average the sample effect sizes. Although in practice the optimal weights are unknown, the UMVU average was included purely for comparison purposes. An unexpected result was that the UMVU average did not systematically show the minimum bias and MSE values in comparison with the HV, HS, and FE averages, where the optimal weights were estimated. The UMVU estimator was approximately unbiased, although with bias values slightly larger than those of the HS estimator (see Figure 1), and generally showed the minimum MSE values, although with very similar (sometimes slightly larger) MSE values to those of the HV average (see Figure 2). This slight deviation of the UMVU procedure in respect to its expected results could be explained by the fact that in practice the Hedges *d* index in Equation 2 follows an approximate (not exact) normal distribution (Hedges, 1981). Anyway, it was relevant to our study that the HV procedure got the best performance in terms of the MSE, showing very similar MSE values to those in the optimal UMVU procedure. Thus, the more direct estimation of the optimal weights in the HV procedure was more efficient than the same estimation in the HS procedure through the sample sizes.

The results of the current study further develop those of our previous article in Sánchez-Meca and Marín-Martínez (1998) on the more realistic conditions of the random-effects model ($\tau^2 \neq 0$), and confirm that as the between-studies variance increased, the performance of the HV estimator in terms of the MSE was better than that of the HS estimator, the differences between the MSE values being more pronounced (see Figure 2B). Contrary to our expectations, the increase of the between-studies variance scarcely affected the negative bias in the HV estimator, which slightly increased only in meta-analyses with small sample sizes (see Figure 1B and Table 1).

Although the estimated weights in the HV average are more affected by sampling error than the weights (samples sizes) in the HS average, even in meta-analyses with small sample sizes and a small number of studies, the HV estimator showed lower MSE values than those in the HS estimator. As expected, as the sample size and the number of studies increased, the differences between the MSE values of the two procedures decreased, but the HV estimator was still the one with the minimum MSE (see Figure 2 and Table 2).

In summary, our results support that in meta-analyses with the standardized mean difference as the effect-size index, weighting by the inverse variance of each *d* value (the Hedges and Vevea procedure) yields more accurate results than weighting by sample size (the Hunter and Schmidt procedure) in the computation of the average effect size. The theoretical expectations in Hedges (1983) that the HV average, despite having a slight negative bias, is more efficient than the HS average were confirmed by our empirical data, showing that in the more usual conditions of real meta-analyses, the HV procedure showed lower MSE values than those in the HS procedure. Other Monte Carlo studies concluded that the best performance was that of weighting by sample size (Brannick, Yang, & Cafri, 2008; Field, 2001, 2005; Hall & Brannick, 2002; Schulze, 2004), but these are based on meta-analyses with the Pearson correlation coefficient as the effect-size index and their results are not comparable with the current study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/orpublication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: Fondo de Investigación Sanitaria of the Spanish Government (Project No. PI07/90384).

Notes

1. See Viechtbauer (2005, 2007) and Sánchez-Meca and Marín-Martínez (2008) for a more extensive list of between-studies variance estimators.
2. Complete data are available from the authors on request.

References

Aptech Systems. (2001). The GAUSS Program (Version 3.6) [Computer software]. Kent, WA: Author.

Brannick, M. T., Yang, L., & Cafri, G. (2008, April). *Comparison of weights in meta-analysis under realistic conditions*. Poster presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, *20*, 825-840.

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

Cooper, H. M., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

DerSimonian, R., & Laird, N. (1986). Meta-analysis of 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.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6*, 161-180.

Field, A. P. (2003). The problems of using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics*, *2*, 77-96.

Field, A. P. (2005). Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychological Methods*, *10*, 444-467.

Hall, S. M., & Brannick, M. T. (2002). Comparison of two random-effects methods of meta-analysis. *Journal of Applied Psychology*, *87*, 377-389.

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., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and 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.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research synthesis* (2nd ed.). Thousand Oaks, CA: Sage.

Kisamore, J. L., & Brannick, M. T. (2008). An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods*, *11*, 35-53.

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

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.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.

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

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Sánchez-Meca, J., & Marín-Martínez, F. (1998). Weighting by inverse variance or by sample size in meta-analysis: A simulation study. *Educational and Psychological Measurement*, *58*, 211-220.

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*, 31-48.

Schmidt, F. (2008). Meta-analysis. A constantly evolving research integration tool. *Organizational Research Methods*, *11*, 96-113.

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of difference in results. *British Journal of Mathematical and Statistical Psychology*, *62*, 97-128.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber.

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Journal of Psychology*, *215*, 90-103.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*, 261-293.

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, *26*, 37-52.