Meta-Analysis of Correlations

This handout is based on material from:

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6* (2), 161–180.

Meta-analysis is a statistical technique by which information from independent studies is assimilated. Traditionally, social science literatures were assimilated through discursive reviews. However, such reviews are subjective and prone to 'reviewer-biases' such as the selective inclusion of studies, selective weighting of certain studies, and misrepresentation of findings (see Wolf, 1986). The inability of the human mind to provide accurate, unbiased, reliable and valid summaries of research created the need to develop more objective methods. Meta-analysis *arguably* provides the first step to such objectivity (see Schmidt, 1992), although it too relies on subjective judgements regarding study inclusion (and so is still problematic because of biased selections of studies, and the omission of unpublished data—the file drawer problem). Since the seminal contributions of Glass (1976), Hedges and Olkin (1985), Rosenthal and Rubin (1978) and Hunter, Schmidt and Jackson (1982) there has been a meteoric increase in the use of meta-analysis. Field (2001) reports that over 2200 published articles using or discussing meta-analysis were published between 1981 and 2000. Of these, over 1400 have been published since 1995 and over 400 in the past year. Clearly, the use of meta-analysis is still accelerating.

Basic Principles

To summarise, an effect-size refers to the magnitude of effect observed in a study, be that the size of a relationship between variables or the degree of difference between group means. There are many different metrics that can be used to measure effect size: the Pearson product-moment correlation coefficient, r; the effect-size index, d; as well as odds ratios, risk rates, and risk differences. Of these, the correlation coefficient is used most often (Law, Schmidt & Hunter, 1994) and so is the focus of this study. Although various theorists have proposed variations on these metrics (for example, Glass's Δ, Cohen's d, and Hedges's g are all estimates of δ), conceptually each metric represents the same thing: a standardized form of the size of the observed effect. Whether correlation coefficients or measures of differences are calculated is irrelevant because either metric can be converted into the other, and statistical analysis procedures for different metrics differ only in how the standard errors and bias corrections are calculated (Hedges, 1992).

In meta-analysis, the basic principle is to calculate effect sizes for individual studies, convert them to a common metric, and then combine them to obtain an average effect size. Studies in a meta-analysis are typically weighted by the accuracy of the effect size they provide (i.e. the sampling precision), which is achieved by using the sample size (or a function of it) as a weight. Once the mean effect size has been calculated it can be expressed in terms of standard normal deviations (a Z score) by dividing by the standard error of the mean. A significance value (i.e. the probability, p, of obtaining a Z score of such magnitude by chance) can then be computed. Alternatively, the significance of the average effect size can be inferred from the boundaries of a confidence interval constructed around the mean effect size.

Johnson, Mullen and Salas (1995) point out that meta-analysis is typically used to address three general issues: central tendency, variability and prediction. Central tendency relates to the need to find the expected magnitude of effect across many studies (from which the population effect size can be inferred). This need is met by using some variation on the average effect size, the significance of this average or the confidence interval around the average. The issue of variability pertains to the difference between effect sizes across studies and is generally addressed using tests of the homogeneity of effect sizes. The question of prediction relates to the need to explain the variability in effect sizes across studies in terms of moderator variables. This issue is usually addressed by comparing study outcomes as a function of differences in characteristics that vary over all studies. As an example, differences in effect sizes could be moderated by the fact that some studies were carried out in the USA whereas others were conducted in the UK.

Fixed versus Random Effects Models

So far, we have seen that meta-analysis is used as a way of trying to ascertain the true effect sizes (i.e. the effect sizes in a population) by combining effect sizes from individual studies. There are two ways to conceptualise this process: fixed effects and random effects models[1]. Hedges (1992) and Hedges and Vevea (1998) explain the distinction between these models wonderfully. In essence, in the fixed effect conceptualisation, the effect sizes in the population are fixed but unknown constants. As such, the effect size in the population is assumed to be the same for all studies included in a meta-analysis (Hunter & Schmidt, 2001). This situation is called the *homogenous* case. The alternative possibility is that the population effect sizes vary randomly from study to study. In this case each study in a meta-analysis comes from a population that is likely to have a different effect size to any other study in the meta-analysis. So, population effect sizes can be thought of as being sampled from a universe of possible effects—a 'superpopulation' (Hedges, 1992, Becker, 1996). This situation is called the *heterogeneous* case. To summarise, in the random effects model studies in the meta-analysis are assumed to be only a sample of all possible studies that could be done on a given topic whereas in the fixed effect model the studies in the meta-analysis are assumed to constitute the entire universe of studies (Hunter & Schmidt, 2001).

In statistical terms the main difference between these models is in the calculation of standard errors associated with the combined effect size. Fixed effects models use only within-study variability in their error term because all other 'unknowns' in the model are assumed not to affect the effect size (see Hedges, 1992; Hedges & Vevea, 1998). However, in random effects models it is necessary to account for the errors associated with sampling from populations that themselves have been sampled from a superpopulation. As such the error term contains two components: within-study variability and variability arising from differences between studies (see Hedges & Vevea, 1998). The result is that standard errors in the random-effects model are typically much larger than in the fixed case if effect sizes are heterogeneous and, therefore, significance tests of combined effects are more conservative.

In reality the random effects model is probably more realistic than the fixed effects model on the majority of occasions (especially when the researcher wishes to make general conclusions about the research domain as a whole and not restrict their findings to the studies included in the meta-analysis). Despite this fact, the National Research Council (1992) reports that fixed effects models are the rule rather than the exception. Osburn and Callender (1992) have also noted that real-world data are likely to have heterogeneous population effect sizes even in the absence of known moderator variables (see also Schmidt and Hunter, 1999). Despite these observations, Hunter and Schmidt (2001) reviewed the meta-analytic studies reported in Psychological Bulletin (a major review journal in psychology) and found 21 studies reporting fixed-effects meta-analyses but none using random effects models. In addition, Field (submitted) has demonstrated that using fixed effects models in situations in which the population effect sizes are variable results in error rates ranging from 31% to 72% depending on the sample size used. At best, he concluded, 1 in 3 meta-analyses will make a Type I error

[1] In reality it is possible to combine fixed and random effects conceptualizations to produce a mixed model. For the purpose of this study the mixed model is ignored but the interested reader is referred to Hedges (1992).

(i.e. conclude that there is an effect in the population when none exists) and at worst half to three quarters of studies will make a similar error.

Although fixed-effect models have attracted considerable attention (Hedges, 1992, 1994a,b), as Hedges and Vevea (1998) point out, the choice of model depends largely on the type of inferences that the researcher wishes to make: fixed-effect models are appropriate only for conditional inferences (i.e. inferences that extend only to the studies included in the meta-analysis) whereas random-effects models facilitate unconditional inferences (i.e. inferences that generalise beyond the studies included in the meta-analysis). For real-world data in the social sciences researchers typically wish to make unconditional inferences and so random-effects models are often more appropriate.

Over the last 20 years three methods of meta-analysis have remained popular (see Johnson, Mullen & Salas, 1995): the methods devised by Hedges, Olkin and colleagues, Rosenthal and Rubin (see Rosenthal, 1991), and Hunter and Schmidt (1990)[2]. Hedges and colleagues (Hedges & Olkin, 1985; Hedges, 1992; Hedges & Vevea, 1998) have developed both fixed- and random-effects models for combining effect sizes, whereas Rosenthal (1991) presents only a fixed-effects model, and Hunter and Schmidt present what they have labelled a random-effects model (see Schmidt & Hunter, 1999). Although Johnson et al. (1995) overview these three meta-analytic techniques, they did not use the methods for correlation advocated by Hedges and colleagues (or use the random-effects versions) and Schmidt and Hunter (1999) have made subsequent observations about the correct use of their method. Therefore, an overview of the techniques used in the current study, with reference to the original sources, is included as a pedagogical source for readers unfamiliar with meta-analysis of correlation coefficients.

Hedges-Olkin and Rosenthal-Rubin Method

For combining correlation coefficients, Hedges & Olkin (1985), Hedges and Vevea (1998) and Rosenthal and Rubin (see Rosenthal, 1991) are in agreement about the method used. However, there are two differences between the treatments that Hedges and colleagues and Rosenthal and Rubin have given to the meta-analysis of correlations. First, Rosenthal (1991) does not present a random effects version of the model. Second, to estimate the overall significance of the mean effect size, Rosenthal and Rubin generally advocate that the probabilities of each effect size occurring by chance are combined (see Rosenthal, 1991; Rosenthal & Rubin, 1982).

Fixed-Effects Model

When correlation coefficients are used as the effect-size measure, Hedges and Olkin and Rosenthal and Rubin both advocate converting these effect sizes into a standard normal metric (using Fisher's *r*-to-*Z* transformation) and then calculating a weighted average of these transformed scores. Fisher's *r*-to-*Z* transformation (and the conversion back to *r*) is described in equation (1). The first step, therefore, is to use this equation to convert each correlation coefficient into its corresponding *Z* value (see Field, 1999 for an example).

$$z_{r_i} = \tfrac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right) \qquad\qquad r_i = \frac{e^{(2z_i)}-1}{e^{(2z_i)}+1} \qquad (1)$$

The transformed effect sizes are then used to calculate an average in which each effect size is weighted. Equation (2) shows that the transformed effect size of the *i*th study is weighted by a weight for that particular study (w_i).

[2] Although Hunter, Schmidt & Jackson (1982) originally developed this method, Hunter and Schmidt (1990) provide an updated and more comprehensive exposition of the technique.

$$\bar{z}_r = \frac{\sum_{i=1}^{k} w_i z_{r_i}}{\sum_{i=1}^{k} w_i} \tag{2}$$

Hedges and Vevea (1998) note that effect sizes based on large samples will be more precise than those based on small samples and so the weights should reflect the increased precision of large studies. In fact, the optimal weights that minimise the variance are the inverse variances of each study (see Hedges & Vevea, 1998, equation 2), and for correlation coefficients the individual variance is the inverse of the sample size minus three (see Hedges & Olkin, 1985, p. 227 and p. 231).

$$w_i = \tfrac{1}{v_i} \qquad v_i = \tfrac{1}{n_i - 3} \qquad \therefore\ w_i = n_i - 3$$

As such, the general equation for the average effect size given in equation (2) becomes equation (3) for correlation coefficients (this is equation 4.16 in Rosenthal, 1991, p. 74).

$$\bar{z}_r = \frac{\sum_{i=1}^{k} (n_i - 3) z_{r_i}}{\sum_{i=1}^{k} (n_i - 3)} \tag{3}$$

The sampling variance of this average effect size is simply the reciprocal of the sum of weights (Hedges and Vevea, 1998, equation 4) and the standard error of this average effect size is simply the square root of the sampling variance. So, in its general form the standard error is:

$$SE(\bar{z}_r) = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}} \tag{4}$$

Given that for correlation coefficients the weights are simply n – 3, the standard error becomes:

$$SE(\bar{z}_r) = \sqrt{\frac{1}{\sum_{i=1}^{k} (n_i - 3)}} \tag{5}$$

Hedges and colleagues recommend that a *z*-score of the mean effect size be calculated by simply dividing the mean effect size by its standard error (see equation (6)). The probability of obtaining that value of Z can then be calculated using the standard normal distribution (e.g. Field, 2000, p. 471). However, Rosenthal and Rubin recommend that the probability of obtaining the average effect size be calculated by combining the individual probability values of each correlation coefficient (see Rosenthal, 1991, p. 85-86, equation 4.31). This is the only respect in which the Rosenthal-Rubin and Hedges-Olkin fixed-effects methods differ.

$$Z = \frac{\bar{z}_r}{SE(\bar{z}_r)} \tag{6}$$

Finally, to test the homogeneity of effect sizes across studies, the squared difference between the observed transformed *r* and the mean transformed *r* is used. To create a chi-square statistic some account has to be taken of the variance of each study and as before, for correlation coefficients the variance is just the sample size minus 3 (see Hedges & Vevea, 1998, equation 7). This gives us the statistic *Q* in Equation (7), which has a chi-square distribution (Rosenthal, 1991, equation 4.15, p. 74; Hedges & Olkin, 1985, equation 16, p. 235; Hedges & Vevea, 1998, equation 7, p. 490).

$$Q = \sum_{i=1}^{k} (n_i - 3)\left(z_{r_i} - \bar{z}_r\right)^2 \qquad (7)$$

Random–effects model

Rosenthal (1991) does not present a random effects version of the model previously described. However, Hedges and Olkin (1985) and Hedges and Vevea (1998) clearly elaborate on how a random-effects model can be calculated. The main difference in the random effects model is that the weights are calculated using a variance component that incorporated between-study variance in addition to the within-study variance used in the fixed-effect model. This between-study variance is denoted by τ^2 and is simply added to the within-study variance. As such the weights for the random-effects model $\left(w_i^*\right)$ are (see Hedges & Vevea, 1998, equation 13):

$$w_i^* = \frac{1}{v_i + \tau^2} \qquad v_i = \frac{1}{n_i - 3} \qquad \therefore w_i^* = \left(\frac{1}{n_i - 3} + \tau^2\right)^{-1}$$

These new weights can simply be used in the same way as for the fixed-effects model to calculate the mean effect-size, its standard error and the *z*-score associated with it (by replacing the old weights with the new weights in equations 2, 4 and 6).

The question arises of how the between-study variance might best be estimated. Hedges and Vevea (1998) provide equations for estimating the between-study variance based on the weighted sum of squared errors, *Q* (see equation (7)), the number of studies in the meta-analysis, *k*, and a constant, *c* (see equation (9)).

$$\tau^2 = \frac{Q - (k-1)}{c} \qquad (8)$$

The constant is calculated using the weights from the fixed effects model:

$$c = \sum_{i=1}^{k} w_i - \frac{\sum_{i=1}^{k} (w_i)^2}{\sum_{i=1}^{k} w_i} \qquad (9)$$

When combining correlation coefficients the weights are just *n* − 3 and the constant, therefore, becomes:

$$c = \sum_{i=1}^{k} (n_i - 3) - \frac{\sum_{i=1}^{k} (n_i - 3)^2}{\sum_{i=1}^{k} (n_i - 3)} \qquad (10)$$

If, however, the estimate of between-study variance, τ^2, yields a negative value then it is set at zero (because the variance between-studies cannot be negative).

Finally, the estimate of homogeneity of study effect sizes is calculated in the same way as for the fixed-effect model. In short, the only difference in the random-effects models is that the weights used to calculate the average and its associated standard error now include a between-study component that is estimated using equation (8).

Hunter and Schmidt Method

Hunter and Schmidt advocate a single method (a random-effects method) based on their belief that fixed-effects models are inappropriate for real-world data and the type of inferences that researchers usually want to make (Hunter & Schmidt, 2001)[3]. Hunter and Schmidt's method is thoroughly described by Hunter, Schmidt & Jackson (1982) and Hunter and Schmidt (1990). In its fullest form, this method emphasises the need to isolate and correct for sources of error such as sampling error and reliability of measurement variables. Although there is rarely enough information reported in a study to use the full Hunter and Schmidt technique, even in its simplest form it still differs from the method advocated by Hedges and colleagues and Rosenthal and Rubin. The main difference is in the use of untransformed effect-size estimates in calculating the weighted mean effect size. As such, central tendency is measured using the average correlation coefficient in which untransformed correlations are weighted by the sample size on which they are based. Equation (11) shows how the mean effect size is estimated and it differs from equations (2) and (3) in that the weights used are simply the sample sizes on which each effect size is based, and each correlation coefficient is not transformed.

$$\bar{r} = \frac{\sum_{i=1}^{k} n_i r_i}{\sum_{i=1}^{k} n_i} \tag{11}$$

Like Hedges and colleagues' method, the significance of the mean effect size is obtained by calculating a *Z* score by dividing the mean by its standard error. However, the estimate of the standard error is different in Hunter and Schmidt's method and there has been some confusion in the literature about how the standard error is calculated. Johnson et al. (1995) reported the equation of the variance across studies (the frequency weighted average squared error reported by Hunter and Schmidt 1990, p. 100). The square root of this value should then be used to estimate the standard deviation (as in Equation (12)). The best estimate of the standard error is to divide this standard deviation of the observed correlation coefficients by the square root of the number of studies being compared (Osburn & Callender, 1992; Schmidt et al., 1988). Therefore, as Schmidt and Hunter (1999) have subsequently noted, the equation of the standard deviation used by Johnson et al. should be further divided by the square root of the number of studies being assimilated. Equations (12) and (13) show the correct version (according to Schmidt & Hunter, 1999) of the standard deviation of the mean and the calculation of the standard error. The *Z* score is calculated simply by dividing the mean effect size by the standard error of that mean (Equation (14)).

---

[3] In fact the equation for the mean effect size (see equation 11) implies a fixed-effects model because the use of n_i as a weight assumes homogeneity (and indeed Hunter and Schmidt, 1990, p. 100 assert the homogeneity assumption). However, in more recent work (Schmidt & Hunter, 1999; Hunter & Schmidt, in press) the authors have been quite explicit in labelling their model as random-effect.

$$SD_r = \sqrt{\frac{\sum_{i=1}^{k} n_i (r_i - \bar{r})^2}{\sum_{i=1}^{k} n_i}} \qquad (12)$$

$$SE_{\bar{r}} = \frac{SD_r}{\sqrt{k}} \qquad (13)$$

$$Z = \frac{\bar{r}}{SE_{\bar{r}}} \qquad (14)$$

In terms of homogeneity of effect sizes, again a chi-square statistic is calculated based on the sum of squared errors of the mean effect size (see p. 110-112 of Hunter and Schmidt, 1990). Equation (15) shows how the chi-square statistic is calculated from the sample size on which the correlation is based (*n*), the squared errors between each effect size and the mean, and the variance.

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - 1)(r_i - \bar{r})^2}{\left(1 - \bar{r}^2\right)^2} \qquad (15)$$

Comparison of the Methods

There are two major differences between the methods described. The first difference is the use of transformed or untransformed correlation coefficients. The Fisher transformation is typically used to eliminate a slight bias in the untransformed correlation coefficient: the transformation corrects for a skew in the sampling distribution of *r*s that occurs as the population value of *r* becomes further from zero (see Fisher, 1928). Despite the theoretical basis for this transformation Hunter and Schmidt (1990) have long advocated the use of untransformed correlation coefficients using theoretical arguments to demonstrate biases arising from Fisher's transformation (see Hunter, Schmidt & Coggin, 1996). Hunter and Schmidt (1990) note that 'the Fisher *Z* replaces a small underestimation or negative bias by a typically small overestimation, or positive bias, a bias that is always greater in absolute value than the bias in the untransformed correlation' (p. 102, see also Hunter et al., 1996; Schmidt, Gast-Rosenberg and Hunter, 1980; Schmidt, Hunter & Raju, 1988; Field, 1999).

Some empirical evidence does suggest that transforming the correlation coefficient can be beneficial. Silver and Dunlap (1987) claimed that meta-analysis based on Fisher transformed correlations is always less biased than when untransformed correlations are used. However, Strube (1988) noted that Silver and Dunlap had incorrectly ignored the effect of the number of studies in the analysis and so had based their findings on only a small number of studies. Strube (1988) showed that as the number of studies increased the overestimation of effect sizes based on Fisher transformed correlations was almost exactly equal in absolute terms to the underestimation of effect sizes found when untransformed *r*s were used. Strube's data indicated that the bias in effect size estimates based on transformed correlations was less than the bias in those based on untransformed correlations only when 3 or less studies were included in the meta-analysis (and even then only when these studies had sample sizes of 20 or less). It would be the exception that actual meta-analytic reviews would be based on such a small number of studies. As a final point, Hunter et al. (1996) have argued that when population correlations are the same for studies in the meta-analysis (the homogenous case) then results based on transformed correlations should be within rounding error of those based on untransformed values.

The second difference is in the equations used to estimate the standard error. If we compare the random-effects model described by Hedges and Vevea (1998) to Hunter and Schmidt's, the estimates of standard error are quite different. Hedges and Vevea (1998) have suggested that Hunter and Schmidt 'advocate the use of suboptimal weights that correspond to the fixed-effects weights, presumably because they assume that τ^2 [the between-study variance] is small' (p. 493, parentheses added). Therefore, if the between-study variance is not small, the Hunter and Schmidt method will underestimate the standard error and hence overestimate the *z*-score associated with the mean (Hedges & Vevea, 1998). However, Hedges and Vevea's (1998) estimate of the between-study variance is truncated (because negative values lead to the assumption that $\tau^2 = 0$), and so when there are only a small number of studies in the meta-analysis the estimate of between-study variance will be biased and the weights used to calculate the average effect size (and its significance) will be biased also.

Johnson et al. (1995) used a single database to compare the Hedges-Olkin (fixed-effect), Rosenthal-Rubin and Hunter-Schmidt meta-analytic methods. By manipulating the characteristics of this database Johnson et al. looked at the effects of the number of studies compared, the mean effect size of studies, the mean number of participants per study and the range of effect sizes within the database. In terms of the outcomes of each meta-analysis, they looked at the resulting mean effect size, the significance of this effect size, homogeneity of effect sizes, and prediction of effect sizes by a moderator variable. Their results showed convergence of the methods in terms of the mean effect size and estimates of the heterogeneity of effect sizes. However, the significance of the mean effect size differed substantially across meta-analytic methods. Specifically, the Hunter and Schmidt method seemed to reach more conservative estimates of significance (and hence wider confidence intervals) than the other two methods. Johnson et al. concluded that Hunter and Schmidt's method should be used only with caution.

Johnson et al.'s study provides some of the only comparative evidence to suggest that some meta-analytic methods for combining correlations should be preferred over others (although Overton, 1998, has investigated moderator variable effects across methods); however, although their study clearly provided an excellent starting point at which to compare methods, there were some limitations. First, Schmidt and Hunter (1999) have criticised Johnson et al.'s work at a theoretical level claiming that the wrong estimate of the standard error of the mean effect size was used in their calculation of its significance. Schmidt and Hunter went on to show that when a corrected estimate was used, estimates of the significance of the mean effect size should be comparable to the Hedges and Olkin and Rosenthal and Rubin methods. Therefore, *theoretically* the methods should yield comparable results. Second, Johnson et al. applied Hedges and Olkin's method for *d* (by first converting each correlation coefficient from *r* to *d*). Hedges and Olkin (and Hedges & Vevea, 1998) provide methods for directly combining *r*s (without converting to *d*) and so this procedure did not represent what researchers would actually do. Finally, the circumstances under which the three procedures were compared were limited to a single database that was manipulated to achieve the desired changes in the independent variables of interest. This creates two concerns: (1) the conclusions drawn might be a product of the properties of the data set used (because, for example, adding or subtracting a fixed integer from each effect size allowed Johnson et al. to look at situations in which the mean effect size was higher or lower than in the original database; however, the relative strength of each effect size remained constant throughout); and (2) the data set assumed a fixed population effect size and so no comparisons were made between random-effects models. A follow-up study is needed in which Monte Carlo data simulations are used to expand Johnson et al.'s work.

Field (2001) did such a follow up comparing the methods of meta-analysis in both the fixed and homogenous case. Field's results (based on 200,000 Monte Carlo trials) showed that the Hunter-Schmidt method tended to provide the most accurate estimates of the mean population effect size when effect sizes were heterogeneous, which is the most common case in meta-analytic practice. In the heterogeneous case, Hedges and colleagues' method tended to overestimate effect sizes by about 15-45%, whereas the Hunter-Schmidt method tended to underestimate it by a smaller amount (about 5-10%), and then only when the population average correlation exceeded 0.5. In terms of the Type I error rate for the significance tests associated with these estimates Hedges and colleagues' method did control this error rate in the homogenous case. The most surprising finding is that neither random-effects method controlled the Type I error rate in the heterogeneous case (except when a large number of studies were included in the meta-analysis) — although Hedges and colleagues' method inflates the Type I error rate less than the Hunter-Schmidt method. Given that the National Research Council (1992) and others have suggested that the heterogeneous case is the rule rather than the exception, this implies that estimates and significance tests from meta-analytic studies containing less than 30 samples should be interpreted very cautiously. Even then, random-effects methods seem poor at detecting small population effect sizes.

Further Reading

Field, A. P. (2001). Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6* (2), 161–180.

Field, A. P. (submitted). The problems in using Fixed-effects models of meta-analysis on real-world data: a primer. Submitted.

Field, A. P. (1999). A bluffer's guide to meta-analysis I: Correlations. *Newsletter of the Mathematical, Statistical and computing section of the British Psychological Society*, *7* (1), 16-25. [http://www.cogs.susx.ac.uk/users/andyf/research/articles/meta1.pdf]

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.

Johnson, B. T., Mullen, B., & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology*, *80* (1), 94-106.

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

Wolf, F. M. (1986). *Meta-Analysis*. Sage university paper series on quantitative applications in the social sciences, 07–061. Newbury Park, CA: Sage.