Introduction to Meta-Analysis

Victoria Martinez, Ed Connor, and Jacqueline Levy

Approaches to Research Synthesis:

Introduction

Due to the ever-growing number of studies in experimental ecology, methods for summarizing results across a series of studies and reaching general conclusions are needed. The process of statistically synthesizing the results of separate independent experiments is known as meta-analysis. Meta-analysis is a quantitative re-evaluation of the outcomes of two or more studies. Meta-analysis involves combining the results of multiple studies to reach an overall conclusion about the magnitude of a treatment effect or covariate examined in a group of studies. Meta-analysis can be performed whenever two or more studies examine the same conceptual hypothesis (i.e., same null hypothesis). For example, meta-analysis could be used to determine whether an initial study and a replication of that study yield similar or different results, and what overall inferences could be drawn from the combined results. Meta-analysis can be used for synthesizing two studies or 200 studies, or any other number of studies that examine the same conceptual hypothesis. Meta-analysis is used to estimate the average effect of a treatment or covariate among a group of studies; is the effect large or small, is the effect positive or negative; does the overall combined effect differ from zero?

The three major approaches to meta-analysis are: vote counting, combining significance levels, and by combining estimates of effect size. Each approach to research synthesis has its advantages and disadvantages and the techniques used are dependent upon the type of information available for synthesis.

Vote Counting

An approach to meta-analysis commonly found in review articles is the technique called vote counting. Vote counting is a method for synthesizing results across studies by counting the number of instances found in the literature that are consistent or inconsistent with an hypothesis. For example, Schoener (1983), Connell (1983), and Denno *et al*. (1995) examined the ecological literature for experimental studies to determine if

interspecific competition is common or rare in nature. For each instance in each study that found interspecific competition, they tallied one vote. They then reported what proportion of studies show interspecific competition - calculated as the number of positive "votes" divided by the total number of instances examined. Schoener (1983) defined a positive vote to be detection of a negative effect of one species on another, while Connell (1983) defined a positive vote to be detection of a statistically significant negative effect of one species on another ($\alpha = 0.05$). The interpretation that the overall pattern in the data is either consistent with the hypothesis that interspecific competition is common in nature is either based on a subjective assessment that the observed proportion of instances of interspecific competition is sufficiently high, or could possibly be subjected to a binomial test with the binomial parameter equal to 0.5.

The problems with vote counting are that one can define a positive vote in different ways, and more importantly, vote counting treats each study – each vote - as being equal. A vote derived from a study with a sample size of 2 is equivalent to vote from a highly replicated study. Furthermore, a vote from a study in which the magnitude of the observed effect is very small is equal to a study in which the magnitude of the observed effect is very large. Because vote counting does not take into account the sample sizes of the studies, vote counting is biased towards studies with small sample sizes, since studies with large sample sizes and small sample sizes are given the same weight (Cooper and Hedges 1994).

On the other hand, if no other information than the existence and direction of an effect is reported in a series of studies, vote counting is the only means of synthesizing results.

Combining Significance levels

A long-standing approach to meta-analysis involves combining the significance levels derived from multiple tests of the same underlying null hypothesis. When the studies under evaluation provide data that do not meet the assumptions necessary to apply the parametric models described below, or only report the *p*-values, tests based on combining significance values can be used for synthesizing results. Because of the non-parametric nature of the tests of combined significance, they can be applied broadly and are fairly easy to compute. *P*-values from studies in which an *F*, *t*, X^2, or other test statistic was applied can be readily combined to obtain an overall test of significance. While there are a number of tests of combined significance that can be used in synthesizing studies, Fisher's method for combining probabilities is most widely used (Becker 1994, Sokal and Rohlf 1995):

$$X^2 = -2\sum_{i=1}^{k} \log(p_i),$$

where p_i is the significance level obtained from the *i*th study, and $-2 \sum \log (p_i)$ is distributed as a X^2 variate with $2k$ degrees of freedom. In a series of experiments when each individual experiment yields a non-significant hypothesis test, if the treatment

consistently increases of decreases the response variable the combined test of significance is likely to be statistically significant. For an example of Fisher's Method applied to ecological studies see Simberloff and Connor (1981) or McQuate and Connor (1990).

The problem with combining probabilities is that the same probability calculated from different studies could arise if one study had a large sample size and a small effect size, and another study had a large effect size and a small sample size (Becker 1994). Hence, the overall test is for statistical significance and provides no information on the average magnitude of the treatment or covariate effect.

Combining Estimates of Effect Size

The most recent development in meta-analysis are procedures that permit one to combine estimates of "effect sizes" to obtain an overall estimate of the average effect size and its standard error, and to test hypotheses about the effects of covariates on the average effect size observed in a series of studies. Combining effect sizes is superior to combining probabilities because the same probability calculated from different studies could arise if one study had a large sample size and a small effect size, and another study had a large effect size and a small sample size (Becker 1994). However, effect sizes may be combined in an unambiguous way by weighting each effect size in proportion to its respective variance, which is in part a function of sample size (Shadish and Haddock 1994).

The effect size is critical in meta-analysis. The effect size is chosen by the investigator and reflects the differences between experimental and control groups or is utilized to find the degree of relationship between the independent and dependent variables (Gurevitch and Hedges 1999). The outcome of each study is summarized as an index of the effect size and these indices are summarized across studies (Gurevitch and Hedges 1999).

Effect sizes are measures of the effect of some experimental treatment or a covariate on a response variable that is observed in each study. A variety of effect size measures are available to be used with response variable that are either continuous or discrete (Rosenthal 1994). Two families of effect size measures are available for continuous variables the *d* - family and the *r* - family. Effect size measures in the *d* – family are appropriate when effect sizes arise from the assessment of the effect of a discrete covariate such as in an ANOVA or *t*-test. Effect size measures in the *r* – family are appropriate when effect sizes arise from the assessment of the effect of a continuous covariate such as in regression or correlation.

The sample size used in estimating an effect size may differ among studies; thus estimates of effect sizes may vary in precision. Therefore, when combining effect sizes, each effect size must be weighted in proportion to its precision. The precision is a function of study sample size; thus the larger the sample sizes the greater the weight. In some instances it is possible to obtain estimates of effect sizes even when only a *p*-value or test statistic has been reported (See Box below).

Obtaining effect-size when significance levels are given

When an effect-size is not reported, it can be obtained from the significance level, where there is a given *p*-value. Knowing the significance level is useful when an effect size estimate or a test of significance is not accessible. Even so, this information can be used to obtain a lower limit effect size estimate using $r = \phi = Z/\sqrt{N}$ (Cooper and Hedges 1994). A table of the standard normal deviates is needed in order to find *Z*, *t*, *F*, or X^2, depending on what kind of *p*-level you have. Once the *p* values or *t* values are obtained, then *r*, Cohen's *d*, or Hedges *g* can be calculated in order to get the effect size indices.

The *d*- family is the most common method for obtaining effect size from significance levels when using categorical covariates. The *r*- family is also used and in some cases the *d* and *r* families are combined to obtain effect size estimates. The *d*-family effect size estimate and the *r*-family effect size estimate can be inter-converted.

To obtain *d* use Cohen's d equation: $d = \frac{t(n_1 + n_2)}{\sqrt{df}\sqrt{n_1 n_2}}$

To obtain *r* use: $r = \sqrt{\frac{t^2}{t^2 + df}}$

(See Cooper and Hedges (1994), Chapter 16 for computational formulas and calculations for the *d* and *r* family).

Fixed-and Random-effects models

Combining estimates of effect sizes in meta-analysis can be consummated by using one of two models: Fixed-effect models or Random-effect models. For a fixed-effect model, one assumes that the studies under examination share a common true effect size, and that the differences of the actual effect size are from sampling error alone (Scheiner and Gurevitch 1993). Unlike fixed-effect models, in random-effects models one assumes that there is a distribution of effect sizes and that differences in effect sizes between studies are due not only to sampling error, but also to other factors such as measurement error and inherent differences between studies. The computations involved in fitting either model depend upon obtaining an effect size estimate for each study examined.

Fixed-Effects Models

The assumptions of fixed effects meta-analysis are that studies under examination share a common true effect size, the control and experimental groups are normally distributed, and the differences of effect size are assumed to be due to sampling error alone. The variances of the sampling error are known as conditional variances, and will be applied in

the actual synthesis of data. “The unbiased estimate of the population effect would then be the simple average of observed study effects; and its standard error would allow computations of confidence intervals around that average” (Cooper and Hedges 1994).

In the both the Fixed - and Random - effects model there are two null hypotheses that can be examined;

$$H_o : \theta_{..} = 0,$$

the overall grand average effect size $(\theta_{..})$ does not differ from zero; and

$$H_o : \theta_{1.} = \theta_{2.} = ... = \theta_{p.},$$

there is no difference between in average effect sizes among the p levels of the covariate.

The null hypothesis of no covariate effect can be examined for both categorical and continuous covariates. For fixed-effects models, the model fitting with categorical covariates involves a weighted ANOVA and with continuous covariates a weighted regression. Statistical packages, such as SPSS, can be used to perform weighted ANOVA and weighted regression to fit fixed-effects model meta-analyses. The model sum-of-squares is distributed as a X^2 variate with number of covariate levels – 1 dfs in the weighted ANOVA, and number of covariates in the weighted regression.

Random-Effects Models

In a fixed effects meta-analysis, the sample estimates of effect sizes, T_i, from the k studies are viewed as estimates of a common population parameter θ_I that is the underlying population effect size and is a fixed value so that $\theta_1 = \theta_2 = ... = \theta_k = \theta$. T_i values from any particular study differs from θ because of sampling error or conditional variability. Because T_i is based on a random sample of subjects from a population it will differ somewhat from θ for the population.

In a random effects model, θ_i is not a fixed value, rather it is a random variable that follows its own distribution. Hence, the total variability of an observed effect size v_i* is a combination of both the sampling error or conditional variation, v_i, about each population's θ_I, and random variation, σ_θ^2, of each θ_i around the mean population effect size:

Variance of estimated effects	= random effects variance	+ estimation (or conditional) variance
v_i*	= σ_θ^2	+ v_i.

σ_θ^2 is referred to as the random effects variance, the between studies variance, or the variance component, v_i as the within-study variance, estimation variance, or the conditional variance of the T_i (i.e., conditional on θ being fixed at the value θ_i), and v_i* as the unconditional variance. If the between studies variance equals zero, then the equations for random effects models reduce to those of fixed effects models.

When would a random effects model be appropriate? If σ_θ^2 is significantly different from 0, then it might be appropriate to use a random rather than a fixed effects model. However, since the power of this test might be low, the use of a random effects model may be warranted even when such a test is insignificant. If the studies in a synthesis are viewed as a random sample from some larger population of potential studies that have been or could be done, and the researcher wishes to draw inferences about the larger population of potential studies, then *a priori* a random effects model is appropriate.

Things to consider before you begin

When planning a meta-analysis it is important to consider sources of variation in the studies that are being included in the meta-analysis. Osenberg *et al*. (1999) suggest that variation among studies in effect magnitudes may arise from four sources: experimental, parametric, functional, and structural. Experimental variation arises when the procedures under which studies were conducted lead to differences in effect sizes. Parametric variation occurs when systems are governed by the same basic processes, yet differ in effect magnitudes generated by those processes. Functional variation is when systems are so distinct that the functions that describe the interactions between variables assume different shapes. Structural variation occurs when systems differ in their causal processes. In any event, one must be aware of sources of variation in effect sizes, and account for such variation by appropriate selection of an effect size measure or perhaps by conducting a mixed-model analysis.

Publication Bias

Like any study a meta-analysis is only as good as the data used in it. There can be problems with the available data such as: incomplete reporting of data, lack of independence, publication bias, and research bias. Studies that fail to report sample size and variance cannot be included meta-analyses that combine estimates of effect size. If more than one parameter is used in a study then the parameters are not independent. To correct for this lack of independence separate analyses need to be conducted or only one parameter must be examined. Studies performed in the same lab are also an example of a lack of independence that could lead to between study biases. Publication bias may exist when significant studies are published more than non-significant. Begg (1994) outlines approaches to determine if the published literature represents a biased sample of the studies actually conducted. Begg (1994) also describes the file-drawer problem, and a method of estimating how many non-significant, unpublished studies would have to exist to change the conclusion of a meta-analysis. It is also possible that researchers choose to study organism or systems in which it is more likely to detect an effect, this could be a

problem for a meta-analysis which is trying to make generalization about the natural world (Gurevitch and Hedges 1999).

Calculations

Fixed-Effects Models

The steps required to compute a fixed effects model in meta-analysis are similar to those in calculating an ANOVA; the means, sum of the scores, and the variance are calculated for each group. The steps involved include:

1. The calculation of the grand-mean
2. Calculation of means for different categories of explanatory variables
3. Calculation of the confidence intervals around the means
4. Statistical tests are completed to determine the consistency of the effects within and among categories of the studies.

Effect size is calculated for each experiment as the difference between the means of two groups of individuals, divided by their pooled standard deviation to standardize the effect among studies.

(We use an effect size measure from the *d* - family to illustrate the calculations for the Fixed-effects model and one from the *r* - family to illustrate the Random Effects Model).

Notation:

k = total number of independent studies among all groups
m_i = number of studies in each group
p = number of groups (a level of the covariate)
T_i = observed effect size
v_i = conditional variance
w_i = weight $= \frac{1}{v_i}$,
θ = population effect size, under the fixed–effects model, we assume $\theta_1 = ... = \theta_k = \theta$ is the common effect size.

Group Weighted Mean

This is the general formula for the group-weighted mean. The singular dot indicates that the effect size measure has been averaged across all studies within a particular level of the covariate. The group weighted mean effect size estimate for the *ith* group $\overline{T}_{i\cdot}$ is

$$\overline{T_{i\bullet}} = \frac{\sum_{j=1}^{m_i} w_{ij} T_{ij}}{\sum_{j=1}^{m_i} w_{ij}}$$

$i = 1, \ldots, p,$ where the weight w_{ij} is the reciprocal of the variance of T_{ij}, $w_{ij}=1/v_{ij}$.

Grand Weighted Mean

The Grand Weighted Mean, $\overline{T}_{..}$, is obtain by summing the group weighted means among all groups. Two dots indicate the overall grand mean.

$$\overline{T_{\bullet\bullet}} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij} T_{ij}}{\sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij}}$$

Group Mean Conditional Variance

The conditional variance is given by the reciprocal of the sum of the weights in each group.

$$v_{i\bullet} = \frac{1}{\sum_{j=1}^{m_i} w_{ij}}$$

Grand Mean Conditional Variance

The Grand Mean Conditional Variance ($v..$) or sampling variance is obtained by summing the Group Mean Conditional Variance among groups.

$$v_{\bullet\bullet} = \frac{1}{\sum_{i=1}^{p}\sum_{j=1}^{m_i} w_{ij}}$$

Now that you have obtained the grand weighted mean $\overline{T}_{..}$ and the sampling variance $v_{..}$, one can test the null hypothesis that the overall grand mean effect size does not differ from zero.

Reject $H_o : \theta_{i.} = \theta_o$ if the absolute value of

$$Z = \frac{(\overline{T_{i\bullet}}) - \theta_0}{(v_{i\bullet})^{1/2}}$$

Exceeds $c_\alpha = 100(\alpha)$ of the standard normal distribution at $\alpha = 0.05$

Confidence Intervals

Confidence intervals for the Grand mean or Group mean effect sizes can be obtained using the following formula and by inserting the appropriate weighed mean effect sizes and conditional variances

$$\overline{T_{i\bullet}} - c_\alpha (v_{i\bullet})^{1/2} \leq \theta_i \leq \overline{T_{i\bullet}} + c_\alpha (v_{i\bullet})^{1/2}$$

If the confidence interval does not include zero, reject *Ho.*

Test of Heterogeneity of Effect Sizes Between and Within Groups

To test the null hypothesis of no difference between groups (levels of the covariate) in the average effect size, an omnibus test for between group differences is conducted using the following formula:

$$Q_{BET} = \sum_{i=1}^{p} w_{i\bullet} (\overline{T_{i\bullet}} - \overline{T_{\bullet\bullet}})^2$$

$w_{i.}$ is the reciprocal of the variance ($1/v_{i.}$), of $\overline{T}_i$.

Q_{BET} can be considered to be the weighted sum of squares of group mean effect sizes about the grand mean effect size.

The null hypothesis is tested by comparing the observed value of Q_{BET} with the upper-tail critical values of the χ^2 distribution with p-1 degrees of freedom (Cooper and Hedges 1994). If Q_{BET} exceeds C_α, *Ho* is rejected at α - level.

To test for heterogeneity within groups, an omnibus test for within-group variation is conducted using the following formula:

$$Q_w = \sum_{i=1}^{p} \sum_{j=1}^{m_i} w_{ij} \left(\overline{T_{ij}} - \overline{T_{i\bullet}} \right)^2$$

The w_{ij} are the reciprocals of v_{ij}, which is the sampling variance of T_{ij}.

The null hypothesis is tested by comparing the obtained value of Q_w with the upper-tail critical values of the chi-squared distribution with k-p degrees of freedom, where $k = m_1 + m_2 + ... + m_p$ is the total number of studies (Cooper and Hedges 1994). If Q_w exceeds 100(1-α), Ho is rejected. A significant Q_w test would suggest that a Fixed-Effects Model might be inappropriate.

Estimating the random effects variance

Several procedures are available to estimate the random effects variance, σ_θ^2. Shaddish and Haddock (1994) present two approaches that are appropriate when no attempt is being made to determine if study characteristics (covariates) account for variation in effect sizes. Raudenbush (1994) outlines a more general procedure that can be used when covariates are used to model the effects of study characteristics. The Raudenbush (1994) approach will be presented below in the section on fitting random effects models with covariates.

Shaddish and Haddock (1994) Method 1 for computation of σ_θ^2

Step 1 - Compute the unweighted variance of the effect sizes, $T_1,\ldots T_k$.

$$s^2(T) = \sum_{i=1}^{k} \left[\left(T_i - \overline{T}_i \right)^2 / (k-1) \right].$$

Step 2 - Compute an estimate of the random effects variance, $\hat{\sigma}_\theta^2$, as the difference between the total variance in the effect sizes minus the $1/k$ times the sum of the conditional variances, v_i.

$$\hat{\sigma}_\theta^2 = s^2(T) - (1/k) \sum_{i=1}^{k} v_i \ .$$

Step 3 - Compute the unconditional variance of each effect size as

$$v_i^* = \hat{\sigma}_\theta^2 + v_i .$$

We see that the unconditional variance in the random effects model is the sum of two components, and will always be greater than or equal to the unconditional variance estimated for a fixed effects model with the same data. As a result, standard errors and confidence intervals will be larger for the random effects models. Therefore, hypothesis tests will be conservative in the random effects case relative to the fixed effects case.

Combining effects sizes under a random effects model

When one is not examining the contribution of study characteristics (covariates) in a random effects model, then the procedures for combining effect sizes are similar to those used for fixed effect models. Confidence intervals of the average effect size and tests of the null hypothesis that the average effect size equals zero, $H_0 : \theta = 0$, are calculated in a similar fashion, except that the unconditional variance, v_i*, is used in place of v_i as the weight for each effect size.

Fitting Models with Categorical or Continuous Covariates under a Random Effects Model

Suppose we are interested in determining if the effect sizes differ among studies categorized into three groups, and group membership is determined via knowledge about each study. If we have k sample estimates, T_i, of the true effect size for each study θ_i, then

$$T_i = \theta_i + e_i ,$$

where the e_i are the errors of estimation and are assumed to be statistically independent with mean of zero and variance v_i. A model of the true effect sizes can be formulated to depend on study characteristics plus error:

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \mu_i ,$$

where β_0 is the model intercept, $x_{i1},\ldots x_{ip}$ are characteristics of the studies hypothesized to affect the study effect size, θ_i ; $\beta_1,..., \beta_p$ are coefficients measuring the association between study characteristics and effects sizes: and μ_i is the random effect of study i. The random effect of study i is the deviation of study i's true effect size from that predicted by the model. Each μ_i is assumed to be independent with a mean of zero and variance σ_θ^2.

The model depicted above is identical to that for a fixed effects model except that the term μ_i has been added. This model could be viewed as a mixed model with fixed effects $\beta_1,..., \beta_p$ and random effects $\mu_i = 1,\ldots, k$.

Estimation

Raudenbush (1994) outlines a procedure involving maximum likelihood estimates of the random effects variance σ_θ^2. This is an iterative procedure where one refits the model until the estimate of the random effects variance stabilizes, and then fits the final model with the weights given by

$$w_i^* = 1/v_i^* = 1/\left(\hat{\sigma}_\theta^2 + v_i\right).$$

One could write a computer program to iteratively estimate $\hat{\sigma}_\theta^2$ and then use that estimate to fit the final model, or use existing modules in statistics packages in an iterative fashion to do the same. For example, for a meta-analysis in which one is combining correlations and has coded studies into $p = 3$ groups, one would perform the following steps to estimate $\hat{\sigma}_\theta^2$ and fit the final model:

Step 1 - Code the p groups with p - 1 dummy variables taking on values of 0 and 1

Step 2 - Transform the r - values to Z_r using Fisher Z tranformation

$$Z_r = \left[0.5\left(\ln\left(1 + r/1 - r\right)\right)\right]$$

Step 3 - Perform a Ordinary least squares regression (*i.e.*, an unweighted regression) using the regression module in SPSS of the Z_r - values on the dummy variables. In the case of $p = 3$, do a regression of Z_r on x_1 and x_2. From this regression obtain the Mean Square Error (*MSE*).

Step 4 - Calculate the conditional variances for each study. For Z_r - values the condition variance is $v_i = \left[1/\left(n_i - 3\right)\right]$, where n_i is the sample size in the ith study, and calculate the mean conditional variance (*MCV*) using the Descriptive Statistics module under the Summarize option in SPSS.

Step 5 - Calculate an initial estimate of the random effects variance $\hat{\sigma}_\theta^2$ as (*MSE* - *MCV*)

Step 6 - Compute the weights $w_i = 1/\left(v_i + \left(MSE - MCV\right)\right)$ using the transform command in SPSS

Step 7 - Perform the weighted least squares regression (WLS) of Z_r - values on x_1 and x_2. Use the SPSS Weight Estimation Option in the Regression module.

Step 8 - Use the regression coefficients estimated from the WLS regression to calculate residuals using a compute procedure under the Transform menu in SPSS. The residuals are calculated as resid = Z_r - b_0 - b_1x_1 - b_2x_2.

Step 9 - Compute $w_{i2} = 1/\left(v_i + \left(MSE - MCV\right)\right)^2$ using a compute procedure in the Transform menu in SPSS

Step 10 - Compute $pp_i = w_{i2}$ (resid^2 - v_i) also using a compute procedure

Step 11 - Use the descriptive statistic procedure to compute the sum of the w_{i2} and the sum of the pp_i values

Step 12 - Compute by hand a new estimate of the random effects variance, $\hat{\sigma}_\theta^2$, as

$$\sum_{i=1}^{p} pp_i \Big/ \sum_{i=1}^{p} w_{i2}$$

Step 13 - Go back to Step 6 and repeat subsequent steps unless the new value of $\hat{\sigma}_\theta^2$ differs very little from the old value. If the estimated random effects variance is negative set $\hat{\sigma}_\theta^2 = 0$.

Step 14 - Once the final estimate of the random effects variance is obtained, perform a WLS regression of Z_r on the dummy variables x_1 and x_2 with weights as $\left(1/\left(v_i + \hat{\sigma}_{\theta final}^2\right)\right)$.

Step 15 - The results of this regression provides the necessary information for testing the hypothesis that the study characteristics are do not affect the estimated effect sizes. The regression sum-of-squares is distributed as χ^2 with $(p - 1)$ degrees of freedom. Reject the null hypothesis of no effect of study characteristics if the calculated value of χ^2 exceeds the critical value at the specified α.

Connor *et al*. (2000) provide an example of a Random Effects Model meta-analysis using an effect size measure from the *r*-family.

References:

Becker, B.J. 1994. Combining significance levels, p. 215-230. In, H. Cooper and L.V. Hedges (eds.) The Handbook of Research Synthesis. Russell Sage Foundation: New York.

Begg, C.B. 1994. Publication bias, p. 399-409. In, H. Cooper, and L.V. Hedges (eds.). The Handbook of Research Synthesis. Russell Sage Foundation: New York.

Connell, J.H. 1983. On the prevalence and relative importance of interspecific competition: Evidence from field experiments. American Naturalist 122: 661-696.

Connor, E.F., A.C. Courtney, and J.M. Yoder. 2000. Individuals-area relationships: The relationship between animal population density and area. Ecology 81: 734-748.

Cooper H., and L.V. Hedges (eds.). 1994. The Handbook of Research Synthesis. Russell Sage Foundation: New York.

Denno, R.F., M.S. McClure, and J.R. Ott. 1995. Interspecific interactions in phytophagous insects: Competition reexamined and resurrected. Annual Review of Entomology 40: 297-331.

Gurevitch J., and L.V. Hedges 1999. Statistical issues in ecological meta-analysis. Ecology 80: 1142-1149.
Hedges, L.V. 1994. Fixed effects models, p. 285-300. In, H. Cooper and L.V. Hedges (eds.) The Handbook of Research Synthesis. Russell Sage Foundation: New York.
McQuate, G.T. and E.F. Connor. 1990. Insect responses to plant water deficits: II. Effects of water deficits in soybean plants on the growth and survival of Mexican bean beetle larvae. Ecological Entomology 15: 433-445.
Osenberg, G.W., O. Ranelle, and D.E. Goldberg.1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. Ecology 80:1105-1117.
Raudenbush, S.W. 1994. Random effects models, p. 301-322. In, H. Cooper and L.V. Hedges (eds.) The Handbook of Research Synthesis. Russell Sage Foundation: New York.
Rosenthal, R. 1994. Parametric measures of effect size, p. 231-244. In, H. Cooper and L.V. Hedges (eds.) The Handbook of Research Synthesis. Russell Sage Foundation: New York.
Schoener, T.W. 1983. Field experiments on interspecific competition. American Naturalist 122: 240-285.
Shadish, W.R., and C.K. Haddock. 1994. Combining estimates of effect size, p. 261-281. In, H. Cooper and L.V. Hedges (eds.) The Handbook of Research Synthesis. Russell Sage Foundation: New York.
Simberloff, D. and Connor E. 1981. Missing Species Combinations. Ecology 118: 215-239.
Sokal, R.R., and F.J. Rohlf. 1995. Biometry. Third Edition. W.H. Freeman: New York.

Useful Websites on Meta-analysis

Site	Link
MetaWin software	http://www.metawinsoft.com/
BMJ articles on Meta-Analysis	http://bmj.com/collections/ma.htm
Statistics Software for Meta-Analysis	http://www.yorku.ca/faculty/academic/schwarze/meta_e.htm
Ralf Schwarzer: Computer Programs for Meta-Analysis	http://www.fu-berlin.de/gesund/gesu_engl/meta_e.htm
Statistics.com Meta-analysis page	http://www.statistics.com/content/freesoft/mno/meta-ana53.html
Meta-analysis and effect size	http://seamonkey.ed.asu.edu/~alex/teaching/WBI/es.html