# 7. Meta-Analysis and Variance-Known Models EPSY 8268

**Meta-Analysis**

It’s possible to consider a meta-analysis as composed of nested data. In most cases, we can consider study participants as nested within study. There is variation within studies among participants and there is variation between studies.

Multilevel studies can also be conceived as meta-analyses. Each group (e.g., each school) can be conceived of as a study, were we examine correlates of an outcome in each group (each study). Then we pool these results across groups (across studies) and use group information (study characteristics) to explain variation between groups (studies).

In a typical meta-analysis, the participant or individual level data are not available. Only summary statistics are accessible – whatever is reported in the published study. With all the variations in study design, measurement tools, sample characteristics, and others, we can find a set of common effect sizes that capture the result of the study and use these as outcomes at “level 2” to estimate and potentially explain variation between studies.

*Standardized Mean Difference*

One of the most common effect sizes is the standardized mean difference

The difference in the mean outcome for the experimental group and the control group divided by the pooled within-group standard deviation. This puts the difference in terms of standard deviations. Each *dj* is an estimate of the population mean difference in SD units. There is a corresponding population parameter of interest:

The accuracy of *dj* as an estimate of δ*j* is a function of the sample size – as we know that the precision of a statistic (as an estimate of the population parameter) is a function of the sample size.

We can apply the notions of sampling distribution theory to effect sizes, just as we do for other statistics: *dj* | δ*j* ~ **N**(δ*j* , V*j*).

In the case of the standardized mean difference, the sampling variance is

Different effect sizes have different estimates of sampling variance. For example, a correlation is a common effect size. We can consider *rj* to be an estimate of the population correlation and estimated in a series of studies. But we know that the sampling distribution of correlations is not normal, especially as the mean correlation departs from zero. We typically transform correlations to be able to combined them across studies, through a normalizing and variance transformation, the Fishers’ *r* to *Z* transformation:

With the sampling variance of

The first question addressed in a meta-analysis: Is a common effect size parameter estimated by all studies? Do sample effect sizes appear similar across studies? The unknown parameter to be estimated in a meta-analysis is generically denoted by δ in HLM approaches to meta-analysis. In other contexts, the notation is dependent on the specific effect size. Thus, the null hypothesis of equality of effects across *j* studies is

H0: δ1 = δ2 = … = δ*j* = δ

This hypothesis is tested using statistical theory for the distribution of a set of *j* independent, asymptotically normally distributed estimates of study effects δ*j*, each with a large sample variance *Vj*.

Following the procedures of a bare-bones meta-analysis, the weighted mean effect is

where the weight *wj* is the reciprocal of the variance of each effect sixe,  is specifically chosen to obtain an efficient estimator. The variance of the effect size in study *j* is a function of the sampling distribution for the effect size.

The variance of the mean effect size is  with a standard error of the mean . The meta-analyst can quantify uncertainty of the estimate of δ in terms of its standard error. A confidence interval can be computed with the usual formula, with the two-tailed critical *z*-value of the standard normal distribution. The test of H0 above is the *Q*-test of homogeneity

which reflects a ratio of between-study to within-study variance, and is distributed  with *df* = *j*–1 (Hedges wrote a series of articles about this test in 1982). The test of H0 is not without assumptions, primarily we assume that the *dj*s are based on independent samples large enough to satisfy the asymptotic normality assumption.

Regarding independence of effects, a violation occurs when multiple effects from a single study are based on the same sample. Several methods have been introduced to address this issue including weighting or estimation of a single study-level effect.

**Fixed-Effects vs. Random-Effects Models** (From Rodriguez & Maeda, 2006)

Rejection of the null hypothesis of effect homogeneity suggests that there is no common effect—there is significant variation in effects across studies. At this point, it becomes necessary to select the model under which this variation may be explained: a fixed-effects, random-effects, or mixed model. The decision about choice of a model requires judgment on the part of the researcher; however, the decision should be made explicitly and on theoretical and empirical bases. Briefly, the decision regarding which model to employ depends on the universe to which generalizations are made. Hedges (1994, pp. 30-31), Hedges and Vevea (1998), and Hunter and Schmidt (2000) provide guidance for this decision; however, Hunter and Schmidt argued that fixed-effects “models and procedures are rarely, if ever, appropriate for real data in meta-analyses” (p. 284). The question for the random-effects model, as clarified by Hedges (1994), is not *What is true about these studies?* but *What is true about studies like these that could have been done?* Such generalizations can be handled through statistical means by incorporating additional uncertainty to allow for inference to studies not identical to those in the sample (a random-effects model).

[This paragraph was written in the context of doing meta-analysis of estimates of reliability, coefficient alpha.] In a random-effects model, the population coefficient alpha for study *i*, , is not fixed but random with its own distribution—note the parameter has an additional subscript *i*. Total variability of an observed study estimate includes both conditional variation, *vi*, of the estimate around each population  and random variation, , of  around the mean parameter. The additional uncertainty, when considering sample studies to be representative of a larger universe, comes from the fact that study contexts, treatments, and administration procedures differ in many ways that potentially impact results.

In this context, the weights that minimize the variance of the estimated mean parameter are inversely proportional to the sum of the conditional and random-effects variances, . Estimation of the random variance component (the between-studies variance ) is no easy matter. Various estimation procedures result in different estimates with important consequences (see Raudenbush, 1994). Hedges and Vevea (1998, p. 492) provided examples of a method of moments estimate of the random-effects variance component that can be computed with any statistical software program. Raudenbush and Bryk (2002, pp. 205-217) provided a method for computing the maximum likelihood estimate of this variance component through their software program HLM (Hierarchical Linear Models). Other examples are also available (Sheu & Suzuki, 2001). The method of moments random variance component is equal to the maximum likelihood method when conditional variances are equal across effects. The maximum likelihood method would be preferred when possible (maximum likelihood estimates are consistent and asymptotically unbiased and efficient); however, this is more difficult to estimate because of the required iterative estimation procedures (Raudenbush, 1994).

Hedges (1994) describes the differences this way:

1. Fixed effects model – conditional model
	1. Sampling error or uncertainty comes from variation resulting from the sampling of people into studies
	2. The universe is structured; a collection of ensembles of studies where each study is an ensemble corresponds to a study in every other ensemble (identically) with some effect size parameter
	3. The model conditions (holds fixed) the characteristics of students that might be related to the effect size parameter
	4. Inferences are limited to cases in which the ensemble of values of the predictor variables are represented in the sample of studies
2. Random effects model – unconditional model
	1. Study sample is presumed to be a sample from a hypothetical collection (or population) of studies
	2. The universe to which generalizations are made consists of a population of studies from which the study sample comes
	3. Studies in the sample and in the universe differ in study characteristics and in effect size parameter
	4. Studies also differ as a consequence of the sampling of people into the groups of the study
	5. The model does not condition the characteristics of studies that might be related to the effect size parameter
	6. The definition of the universe may be ambiguous

A common method of moments estimate of the random variance component is

 and another is .

These have performed poorly in simulation and replication studies because they are dependent on the effect size estimates themselves in the use of *Q*.

**Level-1 (Within Studies) Model**

The level-1 model for studies *j* = 1, …, J is

*dj* = δ*j* + *ej*

with sampling error *ej* associated with *dj* as an estimate of δ*j*, where *ej* ~ **N**(0, V*j*).

**Level-2 (Between Studies) Model**

Here, the unknown population effect size δ*j* depends on study characteristics and random error

δ*j* = γ0 + γ1*W*1*j* + γ2*W*2*j* +…+ γs*W*s*j* + *uj* where *uj*~ N(0, τ)

The combined model is

*dj* = γ0 + Σγs*W*s*j* + *uj* + *ej*

so that *dj* is normally distributed: *dj* ~ N(γ0 + Σγs*W*s*j*, τ + *Vj*) where Var(*dj*) = τ + *Vj* = Δ*j*

**Estimation**

Estimation is slightly different, since we do not need to estimate *Vj* – it is assumed known. We use ML to estimate τ, and the fixed effects (level-2 coefficients) are estimated using weighted least squares, where the weights are the estimates of precision for each effect size, the inverse sampling variance.

Employing our standard multilevel modeling approach, we then compute empirical Bayes estimators for each study effect

Where , reliability.

References

Hedges, L.V. (1982a). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*, 490-499.

Hedges, L.V. (1982b). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*, 119-137.

Hedges, L.V. (1982c). Fitting continuous models to effect size data. *Journal of Educational Statistics, 7*, 245-270.

Hedges, L.V. (1994). Statistical considerations. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29-38). New York: Russell Sage Foundation.

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. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*, 275-292.

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.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Rodriguez, M.C., & Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods, 11*(3), 306-322.

Sheu, C., & Suzuki, S. (2001). Meta-analysis using linear mixed models. *Behavior Research Methods, Instruments, & Computers, 33*, 102-107.