$$\frac{\sum N r_i}{N_i}$$

Meta-Analysis: Methods of Accumulating Results Across Research Domains

Larry C. Lyons, Manassas VA

Abstract

This paper describes the Hunter-Schmidt method of conducting a Meta-Analysis. Meta-analysis is a set of statistical procedures designed to accumulate experimental and correlational results across independent studies that address a related set of research questions. The paper gives a brief description of meta-analysis methods based on procedures suggested by Hunter, Schmidt, and Jackson (1982) and Hunter and Schmidt (1990). It also presents the formulas and procedures needed for converting study statistics to a common metric, calculating the sample weighted mean r and d, and correcting for range restriction and sampling and measurement error.

1. Introduction.

Meta-analysis is a set of statistical procedures designed to accumulate experimental and correlational results across independent studies that address a related set of research questions. Unlike traditional research methods, meta-analysis uses the summary statistics from individual studies as the data points. A key assumption of this analysis is that each study provides a differing estimate of the underlying relationship within the population (rho). By accumulating results across studies, one can gain a more accurate representation of the population relationship than is provided by the individual study estimators.

While there are a variety of other meta-analysis techniques, this paper will focus on the methods developed by Hunter and Schmidt (Hunter, Schmidt and Jackson, 1982, Hunter and Schmidt, 1990).

Glass and colleagues (e.g., Glass, 1976; 1977; Glass &;Smith, 1977; McGaw &;Glass, 1980; Smith &;Glass, 1977; and Smith, Glass &;Miller, 1980) coined the term meta-analysis, and introduced most of the currently used procedures to psychology.

> *Meta-analysis refers to the analysis of analyses . . . the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.*
>
> (Glass, 1976, p 3).

There are two types of Quantitative Review procedures. One type involves the combination of probability values or Z scores. The procedure for this method was developed in parallel during the 30's by Cochran (1937), Fisher (1932), Pearson (1933) and Tippett (1931). These procedures were developed to address the need in agricultural research to combine the results of a number of independent tests, all of which were planned to test a common hypothesis. An alternative approach was also developed by Fisher in 1932, the r to Z transformation.

The demands of World War II served to assist in the development of combinatorial procedures. In their landmark study on the American soldier, Stouffer and colleagues during the 1940's developed a probability combination method. A more recent version of the combinatorial procedure is Winer's (1971) method of combining independent t tests. The other type of meta-analysis is the accumulation of effect sizes, correlation coefficients or to Cohen's d statistic. Thorndike (1933) was among the earlier researchers to accumulate results across studies using an average correlation. He also corrected the observed variance of results across studies for sampling error (unreliability). The intent of this procedure was to integrate differing research on intelligence.

While procedures for averaging correlations were available since the 1930's, as noted above, and were discussed in various behavioral statistics texts (cf. McNemar, 1969), these procedures generally involved the use of Fisher's r to Z transformation, or were generally not used. Unfortunately no guidelines existed that allowed for a "dimensionless" statistic which could be used as a rubric or common statistic which would be independent of any specific measurement unit. Cohen (1977) developed one such statistic now in common use, the effect size statistic, or d. it was originally developed for use in statistical power analysis and to estimate the optimal sample size for a study.

In the 1970's Glass and colleagues coined the term meta-analysis, and also introduced most of the currently used procedures to psychology. Concurrently, Rosenthal was further developing the combinatorial procedures previously mentioned. Meanwhile, Schmidt and Hunter developed what is commonly termed validity generalization procedures. These involve correcting the effect sizes in the meta-analysis for sampling, and measurement error and range restriction.

Since the late 1970's the use of the quantitative review method had grown almost geometrically (Rosenthal, 1991). To give you an idea of how phenomenal the growth has been, I went to the library and accessed the Psych-Lit CD-ROM. I had the database look up journal references that used the term meta-analysis. Since 1978 there has been 909 references to this term. Before 1983, there were 51 references, while after 1982 there was 858 references.

Figure 1 shows the number of meta-analysis references published from 1975 to 1990. As you can see there is nearly a geometric increase in articles for the last 15 years.

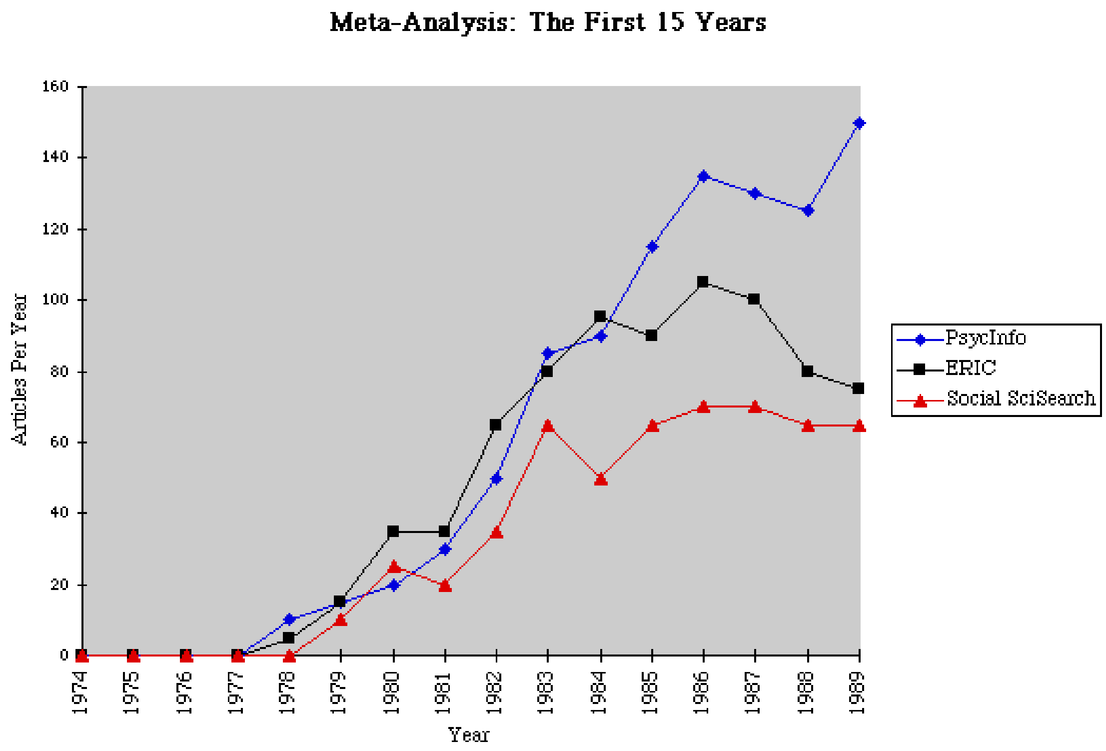

Figure 1: The First 15 years of Meta-Analysis

There are a variety of different procedures for conducting a meta-analysis involving the accumulation of correlations (r), standardized differences between mean scores (d), p values, or Z-scores (Glass, 1976, 1977; Hunter et al., 1982; Hunter and Schmidt, 1990; Rosenthal, 1991; Smith and Glass, 1977; Smith, Glass and Miller, 1980; Wolf, 1986).

Schmidt, Hunter, and their colleagues (Schmidt and Hunter, 1977; Hunter, et. al., 1982; Hunter and Schmidt, 1990) developed one method of meta-analysis that does not rely on the combination of Z-scores or probability values as the common metric. This procedure uses either r or d as the combinatorial statistic. It progressively corrects the mean r or d and their obtained variances for sampling error and then measurement error and range restriction.

In a meta-analysis the literature base is thoroughly searched for experimental and correlational studies that are relevant to the investigation. These studies become the database for the subsequent analysis. Studies reporting on the reliability of the measures used in the various studies and their standard deviations (for range deviation adjustments) in either or both variables are included in the database.

2. Converting Study Statistics to Effect Sizes.

Once the database is assembled, one converts the individual study statistic to a common metric for later accumulation (either r or d). Tables 1 and 2 show several of the more common methods of converting the individual study statistic to either r or d.

Table 1.

Formulas and Procedures for Converting Study Statistics to r.

Statistic to be Converted	Formula for Transformation to r	Notes
t	$r = \sqrt{\frac{t^2}{t^2 + df}}$	Can use with either paired or unpaired t tests
F	$r = \sqrt{\frac{F}{F + df(e)}}$	Use only with one way ANOVAS.
Two-Way ANOVA	$r = \sqrt{\frac{(Fa * dfa)}{(Fa * dfa) + (Fb * dfb) + (Fab * dfab) + df(e)}}$	Fa = Main Effect of Interest dfa = df for A Fb = Second Main Effect dfb = df for B Fab = Interaction effect s dfab = Interaction df df (e) = error df
X 2	$r = \sqrt{\frac{X^2}{N}}$	n = sample size Use only when df = 1
d	$r = \frac{d}{\sqrt{d^2 + \frac{4(N-2)}{N}}}$	d = Cohen's d; N = combined sample sizes.
p	1) Convert the 2 tailed p value into a one tailed p (i.e., p/2). 2) Look up the associated Z in a normal probability table.	Can use for either exact p values or when the author reports an approximate p (e.g., p < .05).

Table 2.

Formulas and Procedures for Converting Study Statistics to d.

Statistic to be Converted	Formula for Transformation to d	Notes
Means and Standard Deviations	$d = \frac{Xe - Xc}{s_p}$	Xe Experimental Group Mean Xc Control Group Mean Sp Pooled (Within Subjects) Standard Deviation
Pooled Within Subjects Variance	$S_p^2 = \frac{(N_e - 1)\ S_e^2 + (N_c - 1)\ S_c^2}{(N_e + N_c - 2)}$	Ne Experimental Group N Nc Control Group N S2e Experimental Group Variance S2c Control Group Variance
t	$d = \frac{2t}{\sqrt{df}}$	Can use with either paired or unpaired t tests
F	$d = \frac{2\sqrt{F}}{\sqrt{df\ (error)}}$	Use only with one way ANOVAS.
r	$d = \frac{2\ r}{\sqrt{1 - r^2}}$	

3. Accumulating the Effect Size and Correcting for Sample Variation.

After the data are converted to a common statistic, reliabilities and range departure information are accumulated. If all the studies include reliability estimates, or range departure data, then the effect size can be corrected at the individual study level. However, most studies do not provide this information. Fortunately Hunter and Schmidt (1990) and Hunter et al (1982) provide procedures for estimating the corrections for reliability and range departures by constructing distributions for the independent and dependent variables.

When the literature base is assembled, and reliability and range information are collected, the next step is to eliminate the downward bias caused by sampling error. This refers to the random variation due to sample size. The sample weighted mean correlation is

$$\bar{r} = \frac{\Sigma\,[N_i\, r_i\,]}{\Sigma\, N_i}$$

where **Ni** is the number of subjects in the study, and **ri** is the effect size for the individual study. The sample weighted mean correlation is defined by the following variance formula:

$$S_r^2 = \frac{\Sigma\,[N_i\,(\,r_i - \bar{r}\,)^2\,]}{\Sigma\, N_i}$$

4. Calculating and Correcting Error Variance.

While the sample weighted mean correlation is not affected by sampling error, its variance is greatly increased. A two-stage procedure is used to correct the variance of the sample weighted mean correlation. The first stage calculates the sampling error variance:

$$S^2_{e_r} = \frac{K\,(1-\bar{r}^2)^2}{\Sigma N_i}$$

where **K** is the number of studies in the analysis.

To estimate the biased population variance, uncorrected for measurement or range departure, the sampling error variance is subtracted from s^2r.

$$S^2_{p_{xy}} = S^2_r - S^2_{e_r}$$

5. Correcting for Unreliability.

So far this meta-analysis technique has corrected for one source of error, sampling error. There are two other forms of error. One is ***measurement error***, which is assessed by the two reliabilities which apply to the study: r_{xx} and r_{yy}. To make the reliability corrections, a distribution of reliability coefficients is constructed. This distribution has the mean of:

$$\bar{r}_{xx} = \frac{\Sigma\sqrt{r_{xx}}}{K}$$

where $\mathbf{r_{xx}}$ is the reliability for the individual study, and **K** is the number of reliability studies. The variance for this distribution is defined as:

$$S^2_{r_{xx}} = \frac{\Sigma\left(\sqrt{r_{xx}} - \bar{r}_{xx}\right)^2}{K}$$

The mean reliability and the variance for the dependent variable use the same formulas.

6. Correcting for Range Departure.

The other source of error is ***range departure***. This refers to the random deviation from *rho* because of variation due to the restriction (when selecting a decreased range of scores) or inflation (when selecting extreme scores only) of the range of possible scores on any measure. To correct for range departure, first compute the ratio of the standard deviation of the individual study to the standard deviation of some reference population, "**u**". This ratio is used to construct "**c**", an estimate of the range departure for the individual study that presents the standard deviation information. The formula for calculating **c** is presented below:

$$c = \sqrt{u^2 + (1 - u^2)\,r^2}$$

where u is the ratio of the study standard deviation to the reference population standard deviation, and r is the effect size found in that study. Since this information is infrequently reported, a distribution of range departure elements is constructed, with a mean and variance of:

$$\overline{C} = \frac{\Sigma C_i}{\Sigma N_i} \qquad\qquad S_c^2 = \frac{\Sigma\,(C - \overline{C})^2}{K}$$

7. Estimating the Relationship within the Population.

To simplify matters from here on, a notation system will be used in which the mean $\mathbf{r_{xx}}$ and $\mathbf{r_{yy}}$ will be denoted as **a** and **b** respectively and the variance of the two reliabilities will be denoted as $\mathbf{s^2a}$ and $\mathbf{s^2b}$.

Given these statistics, the sample weighted mean r, corrected error variance, a, b, c, and the variances for the mean reliabilities and the range departures, the relationship within the population (*rho*) can be estimated. First, correct the sample weighted mean r for measurement error and range departure using the following formula:

$$\overline{r}_{TU} = \frac{\overline{r}}{a\,b\,c}$$

Second, correct the variance of the relationship within the population for measurement error and range departure using the means and variances of the reliability and range correction factors a, b and c:

$$S^2_{r_{TU}} = \frac{S^2_{p_{xy}} - \overline{r}^2_{TU}(b^2\,c^2\,s_a^2 + a^2\,c^2\,s_b^2 + a^2\,b^2\,s_c^2\,)}{a^2\,b^2\,c^2}$$

This is an estimate of the variance of rho. If there is no reason to expect a serious amount of range variation across studies, as in the case of most psychological research, then the correction procedure for the range departure may be omitted (Hunter et al., 1982).

8. Moderator Variables.

When conducting a meta-analysis, look for ***moderating variables*** (third factors that may influence the relationship of interest). Hunter et al. (1982) present a Chi-Square test for systematic variation, which is useful in determining whether there is a moderator variable present.

$$\chi^2_{K-1} = \frac{N}{(1 - \overline{r}^2)^2}\,S_r^2$$

If the Chi-square is not significant, then no moderator variable is present. Statistically this is a very powerful test, given a large enough N, it will reject the null hypothesis even if there is only trivial variation among studies. Alternatively Hunter et al (1982) give a rule of thumb, in which s^2r and s^2er are compared. If error variance accounts for less than 75% of the uncorrected variance, then a moderator variable may be present.

9. Conclusions.

The overall goal of this paper was to acquaint the reader with the procedures and assumptions involved with a Hunter and Schmidt meta-analysis. Meta-Analysis provides a strong alternative to the more traditional review methods. Over the last 15 to 20 years there has been an increased criticism of the social sciences because of the confusing state of the research literature. While one reviewer could find a set of studies which supported his viewpoint, a second reviewer commonly found several which did not. A common conclusion in reviews was "Conflicting Results In The Literature, More Research Is Needed To Resolve This Issue." Which typically resulted in more studies which did nothing to clarify the issue. Meta-analysis offers a way out of this quagmire. By using carefully constructed and comprehensive coding and accumulation procedures, questions which cannot be easily answered with a single study and be resolved using meta-analysis.

10. References.

1. Abramowitz, M. and Stegun, I. (1964). Handbook of mathematical functions with formulas, graphs, and mathematical tables. Washington, DC: United States Government Printing Office.
2. Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd Ed). Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Glass, G. (1976). Primary, secondary and meta-analysis of research. Educational Researcher, 5, 3-8.
4. Glass, G. (1977). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.
5. Haase, R. (1983). Classical and partial eta square in multifactor ANOVA designs. Educational and Psychological Measurement, 43, 35-39.
6. Huffcutt A., Arthur, W., and Bennett W. (1993). Conducting meta-analysis using the PROC MEANS procedure in SAS. Educational and Psychological Measurement, 53, 119-131.
7. Hunter, J. and Schmidt, F. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Beverly Hills CA: Sage.
8. Hunter, J., Schmidt, F., and Jackson, G. (1982). Meta-Analysis: Cumulating research findings across studies. Beverly Hills CA: Sage.
9. McDaniel, M. (1986). Computer programs for calculating meta-analysis statistics. Educational and Psychological Measurement, 46, 175-177.
10. Mullen, B. and Rosenthal R. (1986). BASIC meta-analysis: Procedures and programs. Hillsdale, NJ: Lawrence Erlbaum Associates.
11. Mullen, B. (1990). Advanced BASIC meta-analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.
12. Rosenthal, R. (1991). Meta-analytic procedures for social research (rev ed). Beverly Hills CA: Sage.
13. Schmidt, F. L. and Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.
14. Smith, M., and Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32,752-760.
15. Smith, M. Glass, G. and Miller, T. (1980). The benefits of psychotherapy. Baltimore: The Johns Hopkins University Press.
16. Wolf, F. (1986). Meta-Analysis: Quantitative methods for research synthesis. Beverly Hills, CA: Sage.

11. Steps Involved in Conducting a Meta-Analysis.

1. <u>Define the domain of research</u>

- By independent variable
- By commonly researched variables.
- By causes and consequences of important variables.

2. <u>Establish criteria for including studies in the review</u>

- Published vs. unpublished study.
- The time period covered in the review.
- Operational definitions of the variables.
- The quality of a study.
- etc.

3. <u>Determine type of effect size to use.</u>

- Cohen's d
- Pearson's Product Moment or Point Biserial Correlation.

4. <u>Search for relevant studies.</u>

- Computer search.
- Manual search.
- Conference and Technical Symposium Presentations
- Letters to researchers in the area to be studied.

5. <u>Select the final set of studies.</u>

- Do individually.
- Do by more than one individual.

6. <u>Extract data on variables of interest, sample sizes, effect sizes, reliability of measurement and other noteworthy characteristics of each study.</u>

- Use all the data when multiple measures are reported.
- Use a subset of the data.
- Average multiple study measures to one outcome measure.

7. <u>Code each study for characteristics that might be related to the effect size reported in the study.</u>

- Research design factors.
- Sample Characteristics.
- Type of dependent variable.
- etc.

8. <u>Conduct Reliability checks on the coding procedures.</u>

- With a subset of the data, using 1 to 4 other coders.
- With all the data, using 1 to 4 other coders.

9. <u>When there are multiple measures of independent &/or;dependent variables, decide whether to group them a priori or not.</u>

- Theoretical diversity among variables.
- Operational measurement diversity among variables.

10. <u>Determine the mean and variance of effect sizes across studies.</u>

- Mean effect size weighted by sample size.
- Calculate Chi Square test for homogeneity.
- Calculate Fail Safe N.
- Between-studies variance in effect size for determining moderator variables.
- Estimation of artifactual sources of between studies variance (sampling error, attenuation due to measurement error, and/or range restriction)
- Estimation of true between-studies variance.
- Estimation of true mean effect size corrected for measurement and sampling error, and range restriction.

11. <u>Decide whether to search for moderator variables.</u>

- Significance Test (Chi Square test)
- Amount of between-studies variation that is artifactual.
- Rule of thumb: if the variance accounted for by the error variance is less than 75% of the variance of the sample weighted correlations than there may be a moderator variable otherwise the variation is mainly due to random error (e.g., range restriction, sampling error, or measurement error.

12. <u>Select Potential Moderators (if warranted).</u>

- Theoretical considerations.
- Operational measurement considerations.

13. <u>Determine the mean and variance of effect sizes within moderator subgroups --</u>

Procedure similar to Step 10.