Realities of the Effect Size Calculation Process: Considerations for Beginning Meta-Analysts

Patricia D. Perry

INTRODUCTION

This brief, nontechnical chapter is intended for individuals who have not conducted a meta-analysis, but may be considering doing so. The goal of the chapter is to alert potential meta-analysts to unanticipated hazards they may encounter during the effect size calculation process. It is hoped that awareness of these hazards may assist individuals to assess the extent to which these hazards apply to their particular field of study, to implement plans to minimize the effect of these hazards, and to document the extent to which these hazards were encountered.

Meta-analysis has become a well-accepted method of conducting a quantitative literature review. The intuitive appeal of meta-analysis comes from the belief that findings from multiple studies may provide a more stable and meaningful measure of the magnitude of a treatment effect than results from a single study. Findings from individual studies can be aggregated if their quantitative results are transformed into a standardized difference between a treatment and a control group (i.e., an effect size). While the concept of an effect size is straightforward, the meta-analyst must rely on data provided by other researchers to calculate an effect size estimate. The quality and quantity of those data can vary as a function of the subject matter for the meta-analysis, editorial practices of specific journals, and individual study methodologies.

Since the ability to calculate an effect size is dependent on the way in which the primary researchers conducted and reported their respective studies, the ideal data from which to calculate an effect size may not be available in all studies. Alternatively, the data may be reported in such detail that the meta-analyst has several options available for calculating an effect size estimate. These two complementary issues and their corresponding hazards are the focus of this chapter. The implications of these issues to the planning and implementation of a meta-analysis are also described.

LACK OF SPECIFIC INFORMATION FROM WHICH TO ESTIMATE AN EFFECT SIZE

Ideally, the posttest means and standard deviations (SDs) on a given outcome measure for a treatment and a control group, along with their corresponding sample sizes, are included within individual research reports. However, frequently one or more of these summary statistics is missing for a particular outcome measure, and a meta-analyst must use alternative data to calculate an effect size estimate. Several examples follow.

No Information About Nonsignificant Results

Outcomes with nonsignificant (NS) findings are frequently described within the narrative of the report, or designated as "NS" within a table. A common practice in meta-analysis is to record an effect size of zero for a nonsignificant outcome. However, the magnitude of nonsignificance could impact the aggregated effect size for that particular outcome measure within a meta-analysis. For example, suppose effect sizes from three studies are used to determine the average effect for a given outcome. Study A and study B reported nonsignificant findings, while study C reported data that yielded an effect size estimate of 0.20. The average effect size from these studies would be 0.07 (assuming all other variables to be equal among the studies). By contrast, suppose that studies A and B reported specific means and SDs for each outcome, even though the findings were nonsignificant within their respective studies. If studies A, B, and C had effect size estimates of 0.07, 0.05, and 0.20 respectively, the average effect size would be 0.11 (e.g., 0.32/3). This simplistic example is intended to demonstrate that estimating values, when precise values are missing, can affect the meta-analytic finding. (See Hedges and Olkin 1985 for a discussion on the consequences of observing only significant effect sizes.)

Rounded Probability Levels

When effect sizes are calculated from levels of significance (i.e., probability of type I error), more precise effect size estimates can be obtained if the actual probability level is reported. An effect size calculated from a significance level of $p < 0.05$ is different from one in which the probability level was reported as $p =$

0.35. The magnitude of difference in the effect size estimate will depend on the degrees of freedom, but in principle, the more specific probability level will produce a better estimate of the population effect size than an estimate calculated from a rounded value.

Calculating Effect Size Estimates From Values Within a Graph

Data tables are often one of the best sources of information for a meta-analysis. By contrast, data contained within a graph are usually imprecise. The purpose of a graph is to show a trend (or trends) within the data. When graphs are small, the individual data points that produce the trend are difficult to quantify, and the meta-analyst is forced to rely on a best guess for the actual data points that comprise the graph. It can be a frustrating experience to ponder a graph, knowing that the information is before one's eyes, yet be unable to reproduce the exact value obtained by the original researcher.

Sample Size Not Reported

The sample sizes of the treatment and control groups are used to calculate the inverse of an effect size variance, which, in turn, can be used to weight the effect size estimates according to their respective sample sizes (Hedges 1986). When precise sample sizes are missing from individual reports, they can be obtained by contacting the original researcher, estimated by simple division of the total sample size by the number of groups in the study, or estimated from the degrees of freedom. It is surprising how many research reports state a total sample size but do not report the sample size for the treatment and comparison groups, or for the subgroups within the treatment and comparison groups (e.g., males and females).

The four examples previously described were encountered by the author of this chapter during the effect size calculation process in Tobler's meta-analysis of adolescent drug prevention programs (Tobler 1993). Table 1 lists the proportion of studies in which all the data were available to calculate an effect size estimate for a single drug use outcome measure[1] (i.e., no estimated values were used) and the proportion of studies in which at least one component of the effect size was estimated.

TABLE 1. *Sources of estimated values for effect size calculations from Tobler's 1993 meta-analysis of 120 studies.*

Source of estimated value	Proportion of studies (N)
Finding reported as nonsignificant	16% (19)
Rounded probability level	8% (10)
Outcome value obtained from a graph	10% (12)
Other (misc.)	3% (4)
None; all values reported	63% (75)

TABLE 2. *Source of derivation of sample size from Tobler's 1993 meta- analysis of 120 studies.*

Source of sample size derivation	Proportion of studies (N)
Sources of estimated sample size	
Total sample size divided by the number of groups	19% (23)
Estimated from degrees of freedom	9% (11)
Sources of exact sample size	
No estimation; all sample sizes given	57% (68)
Primary researcher contacted	15% (18)

lists the proportion of studies in which the sample size was estimated (e.g., by degrees of freedom or by dividing the total sample size by the number of groups) and the proportion of studies for which the actual sample size was known. Data from tables 1 and 2 indicate that estimating a value required for effect size calculation was common. While it is certainly preferable to have actual values for effect size calculation, those values simply may not be available from the individual studies that comprise the meta-analysis.

SURPLUS OF INFORMATION FROM WHICH TO CALCULATE AN EFFECT SIZE ESTIMATE

While the lack of information from which to calculate an effect size can be frustrating for a meta-analyst, inclusive information presents a different set of issues and options that deserve consideration. The meta-analytic literature contains multiple references regarding effect size calculation methods (Hedges and Olkin 1985; Hunter and Schmidt 1990; Glass et al. 1981; Lund 1988; McGaw and Glass 1980; Rosenthal 1991; Seifert 1991; Thomas 1986). Indeed, it is the variety of methods available to the meta-analyst that complicates selection of effect size calculation methods.

For example, within Tobler's meta-analysis, 15 different effect size calculation methods were utilized depending on the summary statistics reported within the individual research studies. These methods included:

- Raw posttest means and standard deviations;
- Between groups independent t-test;
- Two-group F statistic;
- One-way analysis of variance (ANOVA) with three groups (omnibus F statistic);
- Chi-square between two groups;
- Correlated t from gain scores;
- Dependent t (matched pairs);
- Raw gain score;
- Level of statistical significance (probability);
- Repeated measure ANOVA;
- Proportions;
- Probit transformation of percentage (Tobler 1989);
- Probit change scores;
- Probit transformation of posttest percentage rates; and
- Regression coefficients.

The extent to which each of these methods was used depended on their frequency of use within the individual studies of interest. Given the multiple options for effect size calculation available to the meta-analyst, frequently more than one method could be used to calculate a specific effect size from a given study. The meta-analyst should be sure that all of the effect size estimates are estimating the same parameter.

An example of this issue is when an F statistic, derived from analysis of covariance, is used to calculate an effect size from an individual study. The F statistic resulting from analysis of covariance incorporates prior information in the final analysis. In other words, the F statistic represents the difference between two groups adjusting for some preexisting differences between groups. For example, the effectiveness of a specific teaching strategy might be evaluated by giving a pretest and a posttest. The individual researcher may want to

control for differences in the outcome that may have been present at the pretest. The pretest would be considered a covariate and variance in the outcome attributable to the pretest would be accounted for in the summary statistic. An effect size calculated from an F statistic derived from analysis of covariance would be a more precise estimate of the population effect size (e.g., effective- ness of the teaching strategy) when compared to an effect size calculated from an F statistic derived from the more simple ANOVA (in which differences in outcome due to preexisting differences were not considered).

When an analysis of covariance F statistic is reported, the meta-analyst has the option of selecting two methods for calculating an effect size estimate. One method would produce a more accurate estimate of the population effect size for that particular study (i.e., by using the covariate-adjusted F statistic to calculate an effect size estimate), while the other method would modify the F statistic to estimate differences between the groups without such adjustment (Smith et al. 1980). Thus, two different effect sizes could be computed for a specific outcome, each one representing a different effect size concept.

The previous example was included to demonstrate that calculating an effect size estimate may not be a simple, straightforward procedure. There are many options available to the meta-analyst for calculating an effect size, and decisions about the effect size calculation process can affect the final meta-analytic findings. When the subject for meta-analysis contains a set of studies in which the degree of effectiveness of the treatment is computed variously, the meta-analyst must consider the extent to which comparable effect sizes can be derived from this set of studies.

IMPLICATIONS FOR PLANNING AND IMPLEMENTING A META-ANALYSIS

The planning and implementation of a meta-analysis is a sophisticated task. There is a great deal of technical information that must be understood regarding sampling error, sampling bias, aggregation of effect sizes, and so forth (see Cook et al. 1992; Cooper and Hedges 1994; Hunter and Schmidt 1990). It is imperative that the beginning meta-analyst be familiar with the different meta-analytic methods available, and proceed to select a method that is compatible with the study objective and the particular field from which the literature will be reviewed. In order to do this, the meta-analyst must be familiar with two sets of literature: the field of study for the meta-analysis, and the meta-analytic literature.

Several steps can be taken during the planning and implementation of the meta-analysis to monitor the extent to which the hazards described in this chapter may be present. First, the meta-analyst should develop

a set of inclusion criteria to determine whether an individual study will or will not be in the analysis. While many factors need to be considered in developing the set of inclusion criteria (e.g., date of publication, type of outcome measurement, type of research design), it is important that the type of primary data available from which to calculate an effect size be incorporated into this set of criteria. This process requires that the meta-analyst be cognizant of potential effect size calculation methods early in the literature review process. Indeed, inability to calculate an effect size is a reasonable criterion for excluding a study from the meta-analysis.

Second, a meta-analyst must decide how restrictive the inclusion criteria will be. For example, must all studies have the exact sample sizes reported, or will an estimate of the sample size be acceptable? The criteria for inclusion should be stated so other meta-analysts and consumers of the meta-analysis will be informed regarding potential sources of error variance in the effect size estimate.

Third, documentation of effect sizes calculation methods should be built into the meta-analytic process (if more than one procedure was utilized). It is important to document how the effect size estimates were calculated so that the method of calculation can be examined vis-a-vis outcome. For example, if better effect size estimates resulted from one effect size calculation method, the results could be examined to determine whether the effect size calculation method itself introduced an artifact that affected the meta-analytic results. Fourth, consideration should be given to calculating and recording an effect size using alternative methods. Using the analysis of covariance example cited earlier, a meta-analyst could calculate a covariate and a noncovariate adjusted effect size estimate. This would enable the meta-analyst to conduct a general meta-analysis (with all studies represented) and a subanalysis of studies that reported a covariate-adjusted summary statistic. The degree of concordance between the two analyses could be informative regarding the strength of treatment effect.

Finally, the meta-analyst is in a unique position to monitor the methodological state of the art for a particular field of interest. If an abundance of studies within the literature lack scientific rigor, the meta-analyst is well placed to discuss such issues. The meta-analyst can also document the extent to which reporting practices lack specificity. For example, data within tables 1 and 2 suggest that the primary research for Tobler's (1993) meta-analysis contained meaningful reporting deficiencies that affected the meta-analytic process. One can only speculate the extent to which the results of Tobler's meta-analysis might have differed if more precise information had been available. Recommendations for improvement

in research methodology and editorial practices are an important outcome of a meta-analytic investigation.

CONCLUSION

This chapter was designed to alert the beginning meta-analyst to a few potential hazards that could be encountered during the effect size calculation process. The extent to which these hazards will be experienced depends on the subject and scope of the meta-analysis. Indeed, one way of minimizing these hazards is to limit the set of studies in the meta-analysis by creating strict inclusion criteria. However, when the purpose of the meta-analysis is broad in scope, excluding studies from the meta-analysis defeats that goal. There is often a tenuous balance between creating inclusion criteria that enhance the validity of the meta-analysis (a factor that tends to limit the number of studies included) while maintaining the goal of a comprehensive review (a factor that supports including numerous studies).

The beginning meta-analyst has much to consider. The meta-analyst must not only be familiar with the field of study, but also possess sufficient competence in statistical analysis to recognize and address the unique issues that arise from quantitatively combining individual research reports (which can vary in almost every aspect of research design). The intuitive attraction of conducting a meta-analysis (i.e., the attempt to summarize the literature from a quantitative perspective) must be attenuated by an appreciation for the complexity of the process by which the meta-analytic findings are generated. Without such an appreciation and a willingness to conduct an indepth study of the statistical procedures associated with calculating and aggregating effect size estimates, the results of the meta-analysis are likely to be spurious and uninterpretable.

NOTES

1. Many studies had more than one drug use outcome measure. The data in table 1 are derived from a single outcome measure in each study and do not represent the entire set of effect sizes in Tobler's meta-analysis.

REFERENCES

Cook, T.; Cooper, H.; Cordray, D.; Hartmann, H.; Hedges, L.; Light, R.; Louis, T; and Mosteller, F., eds. *Meta-Analysis for Explanation: A Casebook.* New York: Russell Sage Foundation, 1992.

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

Glass, G.V.; McGaw, B.; and Smith, M.L. *Meta-Analysis in Social Research.* Beverly Hills, CA: Sage Publications, 1981.

Hedges, L.V. Advances in statistic methods for meta-analysis. In: Cordray, D.S., and Lipsey, M.W., eds. *Evaluation Studies Review Annual: Vol. 11*. Newbury Park, CA: Sage, 1986. pp. 731-748.

Hedges, L.V., and Olkin, I. *Statistical Methods for Meta-Analysis.* Orlando, FL: Academic Press, Inc., 1985.

Hunter, J.E., and Schmidt, F.L. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage, 1990.

Lund, T. Some metrical issues with meta-analysis of therapy effects. *Scand J Psychol* 29:1-8, 1988.

McGaw, B., and Glass, G.V. Choice of metric for effect size in meta-analysis. *Am Educ Res J* 17(3):325-337, 1980.

Rosenthal, R. Meta-analysis: A review. *Psychosom Med* 53:247-272, 1991.

Seifert, T.L. Determining effect sizes in various experimental designs. *Educ Psychol Meas* 51:341-347, 1991.

Smith, M.L.; Glass G.V.; and Miller, T.I. *The Benefits of Psychotherapy.* Baltimore: Johns Hopkins University Press, 1980.

Thomas, H. Effect size standard errors for the non-normal non-identically distributed case. *J Educ Stat* 11:293-303, 1986.

Tobler, N.S. *Measuring Drug Use Differences from Pretest and Posttest Change Scores.* Unpublished manuscript. Available from author, Box 246, San Lake, NY 12153. 1989.

Tobler, N.S. "Updated Meta-Analysis of Adolescent Drug Prevention Programs." Paper presented at the conference on Evaluating School-linked Prevention Strategies: Alcohol, Tobacco, and Other Drugs, University of California, San Diego, March 17-20, 1993.

ACKNOWLEDGMENTS

This chapter was prepared with support from the National Institute on Drug Abuse grant #R01DA05505-2.

AUTHOR

Patricia D. Perry, Ph.D.
Independent Consultant
P.O. Box 9545
Niskayuna, NY 12309

Click here to go to page 130