Reference: DeCoster, J. (2004). Meta-analysis. In Kempf-Leonard, K. (Ed.), The Encyclopedia of Social Measurement. San Diego, CA: Academic Press.

Jamie DeCoster

Free University Amsterdam

I. Introduction

II. Steps to conduct a meta-analysis

A. Define the meta-analytic research question

B. Locate the relevant literature

C. Calculate effect sizes and code moderating variables

D. Analyze the meta-analytic database

E. Report and interpret the results

III. Critically evaluating a meta-analysis

Glossary

effect. A research finding examined in a meta-analysis.

effect size. A statistical estimate of the strength of an effect.

heterogeneity. The variability observed in a sample of effect sizes.

meta-analytic summary. Using meta-analysis to answer a specific research question.

moderator. A study characteristic that influences the effect size.

primary research. Research directly conducted on subjects (in contrast to meta-analytic research, which is conducted on the results of other studies).

quantitative literature review. Using meta-analysis to review the behavior of an effect within a research literature.

Meta-analysis, also known as quantitative synthesis, refers to a set of statistical procedures that are used to combine the results from multiple studies in a single analysis. Traditionally, when reviewers wanted to make inferences about the results of a set of studies, they had to use vaguely defined, qualitative methods. To introduce the same scientific rigor to the review process that is typically required in primary research, meta-analysis provides statistically valid rules for aggregating and discriminating the results found within a literature. Its procedures allow researchers to estimate the average strength of a finding observed across several studies, determine whether the finding is consistent or varies from study to study, and test whether variability in the finding relates to study characteristics.

I. Introduction

Since its introduction less then 30 years ago (Glass, 1976), meta-analysis has strongly influenced research in many areas of the social sciences. It has become the accepted standard for summarizing research in many fields because of its emphasis on objective observation and its openness to critical evaluation. Meta-analysis has also provided the means to investigate new types of research questions, such as how the strength of an effect relates to the setting in which it was investigated. Its growing popularity has even impacted those who only perform primary research, as the guidelines for writing journal articles in many fields have changed to facilitate meta-analysis. It has therefore become important for all social scientists, even those who never actually plan to perform a meta-analysis, to have at least a basic understanding of these procedures. Knowing the ideas behind meta-analysis is necessary to be able to read and critically evaluate meta-analyses performed by others. It is also important to understand how to best present your own

research so that it can be used by meta-analysts who wish to review your literature. Without this knowledge you risk isolation from important sources of information and opportunities for your own research to guide future theorizing.

By far the most common use of meta-analysis has been in *quantitative literature reviews*. These are review articles where the authors select an effect that has been investigated in primary research under a large number of different circumstances. They then use meta-analysis to describe the overall strength of the effect, as well as the circumstances that affect its magnitude and direction. As knowledge of meta-analytic techniques has become widespread, it is more common to see researchers using simple *meta-analytic summaries* within primary research papers. In this case, meta-analysis is used to provide support for a specific theoretical statement, usually about the overall strength or consistency of a relationship within the studies being conducted. As might be expected, calculating a meta-analytic summary is typically a much simpler procedure than performing a quantitative literature review.

II. Steps to conduct a meta-analysis

There are five basic steps to a meta-analysis. First, you define the research question that will be addressed by your analysis. Second, you locate a population of primary research studies that can be examined to answer your question. Third, you calculate an effect size from each study and code its characteristics into "moderating variables." Fourth, you statistically analyze the database of effect sizes. Finally, you present and interpret the results of your analyses. This section provides basic overview of how to conduct a meta-analysis, focusing on these five steps. It will also consider any

differences between the procedures for quantitative literature reviews and those for meta-analytic summaries.

A. Define the meta-analytic research question.

The first thing you do when performing a meta-analysis is determine what information you want it to provide. Forming a clear meta-analytic research question requires three decisions. First, you must choose an effect that you want to examine. Second, you must define the population in which you want to examine the effect. Finally, you must decide what information you want the analysis to provide about the behavior of the effect.

An effect can almost always be thought of as a relationship between a pair of variables, so defining your effect primarily consists of choosing appropriate definitions for the variables whose relationship interests you. You must also choose an effect size metric to be the basis for your analysis. This is a specific statistic that you will calculate from the results of each study so that they may all be compared and aggregated. Effect size statistics must provide a measure of the strength of the relationship that is independent of the sample size. Note that the most commonly used statistics (such as t, F, and p-values) are dependent on sample size, and so cannot be used as effect sizes. Descriptions and formulas for a number of acceptable metrics can be found in Lipsey and Wilson (2000). While the choice of effect size metric will not influence the validity or significance of any of your meta-analytic tests, a given effect size is easier to calculate from certain research designs than from others. The effect size metrics most commonly used in the social sciences are the standardized mean difference (g) and the correlation coefficient (r). The standardized mean difference is typically used when most of the

studies in the analysis reported t or F statistics, while the correlation coefficient is typically used when most of the studies reported correlation or regression coefficients.

To define your population you must decide what types of people you want to examine, and in what situations you want to examine them. The more broadly you define the population of your analysis, the more generally you can apply the conclusions drawn from your results. However, you must always make sure that your definition is restrictive enough so that the included studies all concern the same phenomenon. For meta-analytic summaries, it is appropriate to limit your consideration to the studies performed in a specific paper or set of papers. Although you are then limited to drawing conclusions about a very small population, this is usually appropriate for the purpose of the analysis. Quantitative literature reviews, on the other hand, require broader definitions that encompass the different ways researchers have examined the effect of interest.

Meta-analysis can provide three levels of information about your effect size, with each level requiring somewhat more work to obtain than the last. First, it can provide you with an estimate of the average strength of an effect. Second, it can tell you whether a set of effect sizes come from a single distribution, or if there appear to be systematic differences among them. Third, it can test whether the values of the effect sizes are related to other measured study characteristics. When performing a quantitative literature review you will generally want to examine all of these questions. For a meta-analytic summary, you will want to choose the lowest level of analysis that provides you with the information that you require.

B. Locate the relevant literature

Once you have defined the boundaries of your meta-analysis, you next need to find every study within that area that provides an estimate of the effect of interest. You will typically know ahead of time exactly what studies you will include when performing a meta-analytic summary. This is not the case for a quantitative review, which usually requires a comprehensive literature search to locate the relevant articles. Most literature searches start with queries into one or more computerized indices. Indices are available for most topic areas, allowing you to search for specific words and phrases within the titles and abstracts of relevant articles. One important index that should be examined in almost any search is *Dissertation Abstracts*. This index contains a reference to every doctoral dissertation completed in North America, making it an excellent way to locate research that might not have made it into a published form.

While computerized indices are fairly easy to use, some care must be taken when selecting search terms to obtain complete results. You should first identify the theoretical constructs that define the area of your analysis. Next, you should create lists of the different words or phrases that have been used in the literatures you are searching to identify each of the constructs. If the index you are using assigns keywords to its entries, you should make sure that your lists contain any that are related to your constructs. You should also make use of "wildcards" if they are supported by the index. These are special characters (such as "*" or "?") that you can add to your search terms to indicate that matching entries might have additional characters before or after the ones you specify. This can prevent you from missing studies that happen to use the term you want in a different grammatical form than the one you specified. Once you have completed your

lists, you query the index for studies that include at least one term related to each construct. The best way to do this is to first perform a conjunctive ("or") search for each construct, finding studies related to each of your lists. You then perform a disjunctive ("and") search on these results to locate studies that are related to all of the constructs.

While computerized indices can quickly provide you with a large list of studies related to your topic, a thorough literature search does not rely on them alone. Once you have obtained an initial set of studies, you can perform an "ancestor search" by checking their references to locate older articles that might be related to your topic. If there is a seminal study that was of theoretical or methodological importance to the field, you can perform a "descendent search" by using the *Social Science Citation Index* or the *Science Citation Index* to locate later articles that list the seminal study in their references. Sometimes external organizations will establish "research registers" for a field that has significant societal impact, where they actively maintain a list of ongoing research projects. If there are any existing reviews of your topic (whether meta-analytic or not), you should examine the articles that they have already identified as relevant. You might consider a hand-search of any journals that commonly publish research on the phenomenon of interest. Finally, you should write to active researchers in the field to see if they know of any recent projects that might not be indexed in the other sources mentioned above. One good way to locate such authors is through the use of Internet listservers and newsgroups.

Once you have used these different methods to perform a comprehensive search, you must combine their results into a master list lacking redundancies. You then review each study on this list to determine whether it actually fits the area of the meta-analysis.

Sometimes you can determine whether a study fits simply by reading the abstract, while other times you will need to closely examine the study's method and results. This process can be made easier by developing a set of specific criteria that studies must satisfy to be included in your analysis. When you decide to exclude a study from your analysis, you should note the criterion on which you based this decision. It is not uncommon to change aspects of your research question as your knowledge of the topic area becomes more complete through the course of performing the review. Should you decide to redefine your population (and consequently your inclusion criteria), knowing why you decided to exclude each study lets you limit the number of studies that you need to reexamine.

For a quantitative review to draw accurate conclusions, its literature search must be as complete as possible. If you wish to limit the number of studies in your analysis, you should do so by using more restrictive inclusion criteria rather than by performing a less comprehensive search. Sometimes this will not be possible because you intentionally want to review a very general area. In this case you should still perform a comprehensive search to locate the entire population of studies that examined the effect of interest. From this population you can then randomly sample the number of studies you wish to analyze.

C. Calculate effect sizes and code moderating variables

Once you have obtained a set of relevant studies, you must use them to create a quantitative data set for analysis. The first thing you will need to do is calculate an effect size from each study. Every effect size metric has a computational formula that will allow you to calculate it from raw data. However, you will not always have access to this information, especially if you are conducting a quantitative review. When calculating an

effect size from someone else's research, you must rely on whatever statistics they provide in their report (although you may be able to contact the authors directly for additional information). See DeCoster (2002), Johnson and Eagly (2000), or Lipsey and Wilson (2000) for more specific details about how to calculate effect sizes from different test statistics. You will almost always be able to calculate an estimate of the effect size from a study, although it may be very imprecise for statistically impoverished reports. If there is wide variability in the precision of their estimates, meta-analysts will sometimes perform separate analyses only including the more precise estimates to see if the inclusion of imprecise estimates affects their results.

If you want to determine if a study characteristic influences your effect, you must first code it as a moderating variable. Meta-analysts commonly consider moderators related to the study setting, sample, design, and methodology. For each moderating variable, you should write down an explicit coding scheme detailing how to translate study characteristics into values of the moderator. If a moderator has continuous scale you must specify its unit of measurement. For categorical moderators, you must specify the different groups that compose the variable and what study characteristics are associated with each group. You should be sure to pilot the coding scheme to make sure that it is well understood by the coders and that it provides reliable data.

Your final meta-analytic dataset should contain variables for the effect size, the sample size, and any moderators you want to examine. It is essential that these values are accurate for your meta-analysis to produce valid results, so it is standard practice to have at least two different people calculate and code each effect size. This provides an

additional check on the accuracy of the calculations and allows you to report the reliability of the moderator codes.

D. Analyze the meta-analytic database

As with data from primary research, you should always examine the basic distribution of your effect sizes before performing your analyses. It is especially important to determine if the distribution has multiple modes or outliers. A multi-modal distribution can indicate that your studies are coming from several different subpopulations. In this case you should try to identify a moderating variable that can explain the different groups of effects. If the modes are highly distinct, you might also consider analyzing each of the groups separately. Studies that produce outlying effect sizes should be examined closely to make sure that they truly fit within the boundaries of the analysis. You should consider excluding any outlier that is particularly extreme, whether its characteristics fit the defined population or not, as it will have an inappropriately large influence on your results.

You can use descriptive analyses to obtain information about the central tendency and consistency of your effect sizes. The most commonly used estimate of central tendency is the weighted mean effect size, where each effect size is weighted by the inverse of its variance. Some other measures that have been used are the mean effect size weighted by the sample size, the unweighted mean effect size, and the median effect size. Estimates of the mean effect size combined with its standard error are often used to calculate confidence intervals around the population effect size (Shadish & Haddock, 1994).

After estimating the central tendency, you can determine how well this single value describes the full pattern of effect sizes by computing its heterogeneity (Q). The heterogeneity statistic tells you how likely you would be to observe the amount of variability present in your sample if the effect sizes were all drawn from a single distribution (Hedges, 1994). A significant amount of heterogeneity indicates that your effect sizes are likely not all from the same distribution. In this case you will often want to explain the variability through moderator analyses.

Moderator analyses are used to determine if the effect sizes are related to study characteristics. Before you can test any moderators, you need to decide whether you wish to treat the individual studies as fixed or random factors in your statistical models (although this question may be circumvented by using Bayesian analysis, as described in Louis & Zelterman, 1994). This decision has implications for the formulas that you use to analyze your data, as well as for the inferences you may draw from your results. Fixed-effects models limit you to drawing conclusions about the specific studies and conditions present in your analyses, while random-effects models allow you to draw conclusions about a wider theoretical population of which the studies in your analysis are considered to be a random sample. Fixed models are used much more often than random models, although this is more due to the fact that fixed-effects analyses are simpler and typically more powerful, rather than because they are more appropriate. See Hedges and Vevea (1998) for a more thorough discussion of the implications of choosing fixed- or random-effects models in meta-analysis.

The way that you test a moderator depends on how it was coded. If you want to examine a categorical moderator, you can use heterogeneity statistics to estimate the

between-group and within-group variability. The amount of between-group heterogeneity (Qb) can be used to test whether the moderator can account for a significant amount of the variability in the effect sizes, while the amount of within-group heterogeneity (Qw) can be used to test whether there is a significant amount of variability that cannot be accounted for by the moderator. It is possible for both of these statistics to be significant. This would indicate that the moderator has a significant influence on the effect size but that there are other important sources of variability.

If you want to test a continuous moderator, you can use a weighted regression analysis to estimate the strength of the linear relationship between the moderator and the effect size. The weight for a given study is usually set to be the inverse of the effect size variance, although sometimes meta-analysts choose to use the sample size instead. Parameter estimates for the weighted regression can be obtained using any standard statistical program, although the standard errors of the estimates will need to be corrected for the unique characteristics of meta-analytic data. See DeCoster (2002, p. 39) for more specific information about this procedure.

Recently researchers have become interested in using more complicated models to predict effect sizes. You can test models containing multiple predictors, including interaction and higher-order polynomial models, using a weighted multiple regression analysis. You can include categorical predictors in these models, but they must be coded by a set of indicator variables. The parameters from a multiple predictor model can be estimated using a standard statistical program, but you will need to correct the standard errors of the estimates as when testing a continuous moderator. When interpreting the results of a multiple predictor model, you should keep in mind that the precision of your

estimates will be reduced if there are correlations among your predictor variables. The significance of the parameters in the model will be determined by the independent ability of each moderator to account for variability in the effect sizes. Moderators that are significant when tested individually may be nonsignificant in the multiple predictor model if they are collinear with other included variables.

E. Report and interpret the results

There are many similarities between performing a meta-analysis and conducting a standard research study. In both cases you pose a research question, collect data, analyze the results, and draw conclusions about your findings. The presentation of a meta-analysis consequently parallels the presentation of primary research. When writing a quantitative literature review you use the standard sections found in empirical articles. In the introduction you explain your theoretical motivation for conducting the analysis, and describe how you defined both your effect and the population of studies. In the method section you explain how you selected studies for inclusion, how you calculated the effect sizes, and how you coded any moderating variables. Here you should also report the reliability of the moderator codes. In the results section you describe the analyses you performed and report your findings. Finally, the discussion section should present the theoretical implications of your results and suggest new directions for research.

You need to provide this same information when reporting a meta-analytic summary, though in a compressed format. If your summary includes moderator analyses, you should present it as a separate study in your paper using the guidelines for quantitative reviews just described. If you are only presenting descriptive analyses, however, your meta-analysis will likely be simple to incorporate it directly into your

introduction or discussion. In this case you should describe the purpose and method of your meta-analysis in one paragraph, with the results and discussion in a second.

The presentation of your meta-analysis should provide enough detail so that another researcher could replicate the analyses you performed. Even though there are standard guidelines for how to perform a meta-analysis, there are also a number of subjective decisions that are up to the discretion of the analyst. For example, the moderator analyses can be strongly influenced by the author's personal beliefs. Different researchers might have different opinions on what moderators should be examined and how they should be defined. You should therefore clearly describe and justify each subjective decision that you make so that others can evaluate your choices.

While performing a meta-analysis necessarily requires a large number of calculations, it is important that you do not allow your statistics to dominate your presentation. You should always imbed your analyses in a theoretical framework that provides them with meaning. Your report should describe how the theoretical perspectives in the field influenced the way you defined your constructs and conducted your analysis. You should also be sure to describe any implications your findings may have for existing models of the phenomenon of interest. You should put specific effort into helping your audience understand the meaning of any effect sizes you report. The easiest way to do this is to compare them to the strengths of other effects that would already be known by your intended audience. Alternatively, Cohen (1977) and Rosenthal (1991) provide some statistical tools you can use to help readers understand the practical significance of your effect sizes.

V. Critically evaluating a meta-analysis

While one of the main purposes of meta-analysis is to provide a scientifically rigorous method for summarizing research, you should not automatically assume that every meta-analysis meets this goal. Just as there is variability in the quality of primary research, there can be great differences in the quality of meta-analyses. When judging a meta-analysis, you should consider whether it uses appropriate methods of data collection and analysis (its internal validity), properly represents the phenomenon of interest (its external validity), and makes a distinct theoretical contribution (its scientific importance).

To evaluate the internal validity of a meta-analysis, you should start by considering the validity of the studies in its sample. A meta-analysis can be no more valid than the primary research that it aggregates. If there are methodological problems with the studies in the sample, then the validity of the meta-analysis must also be called into question. The meta-analysis should include enough studies so that it has sufficient power to test the proposed research question. If a meta-analysis examines moderating variables, then you should consider both the reliability of the moderator codes and how the studies are distributed across the levels of each moderator. Unreliable codes and unbalanced designs will reduce the power of moderator analyses. Finally, you should consider whether confounding variables could provide alternative explanations for any significant moderators.

Possibly the most important factor affecting the external validity of a meta-analysis is the extent that the sample represents the theoretical population of interest. To draw firm conclusions about a given population, the sample should be an unbiased representation of every study that has been conducted within that population. Meta-

analytic summaries are typically applied to very specific populations, making external validity less of an issue. The search procedures used for quantitative literature reviews, however, should be closely examined to determine whether there are any relevant subpopulations that were overlooked. For example, some quantitative reviews exclude unpublished or foreign studies because they are difficult to obtain. If there is reason to suspect that these studies might be different than the ones included in the analysis, then valid conclusions cannot be drawn about the literature as a whole. Finally, you should consider the way the author defined the target population. If it includes highly divergent studies, then the analysis may in fact be combining different phenomena. In this case the results may have little meaning.

If you determine that the conclusions drawn from a meta-analysis are valid, you should then evaluate the importance of its findings. Sometimes a meta-analysis will simply provide an estimate of an effect size within a set of studies. While this may be valuable, it does not make as great a scientific contribution as an analysis that additionally uses moderator analyses to determine the conditions under which the effect is stronger, weaker, or even reversed. Finally, you should consider the extent to which the results of the analysis have theoretical implications. The more implications a meta-analysis has for existing theories, the more likely it will influence future research. The best meta-analyses provide statistically valid information about an internally consistent population, which is then used to provide novel insights about the phenomenon under study.

Bibliography

Cohen, J. (1977). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). New York: Academic Press.

DeCoster, J. (2002). Notes on meta-analysis. Retrieved from the web June 12, 2002. http://www.stat-help.com/notes.html

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, 3-8.

Hedges, L. V. (1994). Fixed effects models. In Cooper, H., & Hedges, L. V. (Eds.), The Handbook of Research Synthesis (pp. 285-300). New York: Sage.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. Psychological Methods, 3, 486-504.

Johnson, B. T., & Eagly, A. H. (2000). Quantitative synthesis in social psychological research. In H. T. Reis & C. M. Judd (Eds.), Handbook of Research Methods in Social Psychology (pp. 496-528). London: Cambridge University Press.

Lipsey, M. W., & Wilson, D. B. (2000). Practical Meta-Analysis. Thousand Oaks, CA: Sage.

Louis, T. A., & Zelterman, D. (1994). Bayesian approaches to research synthesis. In Cooper, H., & Hedges, L. V. (Eds.), The Handbook of Research Synthesis (pp. 411-422). New York: Sage.

Rosenthal, R. (1991). Meta-analytic Procedures for Social Research (rev. ed.). Beverly Hills, CA: Sage.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In Cooper, H, & Hedges, L. V. (Eds.), The Handbook of Research Synthesis. New York: Sage.