Meta-analysis: In Practice

An article for the

International Encyclopedia of the Social and Behavioral Sciences

Abstract

The practice of meta-analysis is concerned with the details of implementation of a research synthesis that ensure the validity and robustness of the results from that synthesis. In this article, selected topics are discussed that represent current intellectual themes in the practice of meta-analysis, such as, (i) the role of decisions and judgments, particularly judgments about similarity of studies; (ii) the importance of sensitivity analysis to investigate the robustness of those decisions; and (iii) the role research synthesis plays in the process of scientific discovery. Brief illustrations of the role meta-analysis plays in explanation, program evaluation, and in informing policy decisions are presented.

Meta-analysis: In Practice

1. Introduction

If meta-analysis describes the quantitative approach to combining information from a number of different studies, then the practice of meta-analysis is concerned with the details of implementation of that approach. In effect, the practice of meta-analysis is concerned with ensuring the validity and robustness of the entire enterprise of research synthesis, including, problem formulation, design and data collection, data quality and data base management, and analysis and interpretation. The practice of meta-analysis is really about the practice of data analysis.

The scope of this article, by necessity, is modest. However, a comprehensive guide to the art and science of research synthesis is available in *The Handbook of Research Synthesis* (Cooper and Hedges (eds.) 1994a) and is recommended to the interested reader. Here, selected topics are presented that represent current intellectual themes in the practice of meta-analysis, such as, (i) the role of decisions and judgments, particularly judgments about similarity of studies; (ii) the importance of sensitivity analysis to investigate the robustness of those decisions; and (iii) the role research synthesis plays in the process of scientific discovery, with brief illustrations of how meta-analysis contributes to that process.

2. Selected Issues in Practice

2.1 Similarity Considerations and Combining Information

Implicit in the decision to combine information across studies is the assumption that the studies are similar or exchangeable. The dimensions on which studies might be similar are quite broad and include, for instance, design features, subject pool, quality, measurements, and even context. This assessment about the similarity of studies is central in the practice of meta-analysis because it is at the heart of deciding what information to combine and how to combine that information (Gaver et al. 1992). Underlying the choice of

statistical models for combining information in a meta-analysis, for example, are explicit assumptions about the similarity of the studies. Thus, specifying a fixed-effects model for a meta-analysis is essentially equivalent to the assumption that the studies are homogeneous, measuring the same underlying parameter, and any observed differences among studies is due to sampling variability. Alternatively, a random-effects models, formally incorporates study heterogeneity through the specification of a between-study component of variance.

It will be assumed here that each study in a meta-analysis provides summary information that is common across studies, including measures of an outcome of interest and standard errors of those measures. Often there is little control over the choice of the summary measure and different summary measures may be reported across the studies. In practice, the meta-analyst will need to decide upon a single meaningful measure, frequently a standardized measure such as an effect-size or risk ratio, and subsequently transform the study-specific summaries to the chosen measure. For example, in a study of the effect of teacher expectancy on pupil IQ, the summary measure used for each study was the effect-size, a standardized mean difference calculated as the mean IQ score of the experimental group minus the mean of the control group divided by a pooled standard deviation (see *Meta-analysis: Overview*). Effect-sizes are most useful when study outcomes are assessed with different instruments and are measured on a continuous rather than dichotomous scale. In the case when all the studies to be combined report the same summary measure, as in the IQ study, an important practical decision for the analyst is whether different studies are more usefully compared by their effect sizes or by the simple unscaled differences in means, a quantity often of more direct scientific relevance and interpretability.

Assessments about what studies to combine requires subjective judgments on the part of the investigator and will depend on the motivation and context of the research synthesis. In addition, inherent in the specification of a statistical model to combine information in a meta-analysis are explicit formal assumptions about the similarity of the studies and the data to be combined (Draper et al. 1993). Assume that there are k independent studies

available for a meta-analysis, and let T_i denote the study-specific effect from study i ($i = 1, \ldots, k$). If all the studies are measuring the same thing, for example, the same "true" change in IQ scores, the k studies are assumed to be similar in that each study is assumed to provide an estimate of the true effect θ ($= \theta_1 = \ldots = \theta_k$). This assumption of similarity implies the following fixed-effects model for combining the study-specific effects,

$$T_i = \theta + \epsilon_i \tag{1}$$

where ϵ_i is typically assumed normal with mean 0 and variance V_i, and the ϵ_i are regarded as independent. (For simplicity, assume the V_i are known.) An optimal estimator of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}, \tag{2}$$

a weighted average of the T_i in which the weights w_i are proportional to $\frac{1}{V_i}$. (See also *Meta-analysis: Overview*, and *Meta-analysis: Tools*.)

Alternatively, the meta-analyst might conclude that the studies are heterogeneous, implying that differences in the T_i come both from experimental error and from actual differences between the studies. If explicit information on the source of heterogeneity is available, it may be possible to account for study-level differences in the modeling, but in the absence of such information a useful and simple way to express study-level heterogeneity is with a random-effects or <u>hierarchical model</u>:

$$\begin{aligned} &\text{Stage I:} \quad T_i = \theta_i + \epsilon_i \\ &\text{Stage II:} \quad \theta_i = \theta + \zeta_i \end{aligned} \tag{3}$$

where the ϵ_i are, for example, taken to be normal with mean 0, and variance V_i (the within-study component of variation), and the ζ_i are regarded as normal with mean 0 and variance τ^2 (the between-study component of variation). All the ϵ_i and ζ_i are considered independent. When a probability distribution is specified for the second stage parameters, θ and τ, this class of models is known as a Bayesian hierarchical model (see *Bayesian Statistics*, and *Hierarchical Models*).

A key implication of the second stage model in (3) is that there is information about a particular θ_i from all of the other study-specific effects, $\{\theta_1, \ldots, \theta_k\}$. Specifically, an

estimate of θ_i is

$$\hat{\theta}_i(\tau) = (1 - B_i(\tau))T_i + B_i(\tau)\hat{\theta}(\tau) \tag{4}$$

a weighted average of the observed ith study-specific effect, T_i, and the estimate of the population mean effect, $\hat{\theta}(\tau)$, where the weights are $B_i(\tau) = \frac{V_i}{V_i+\tau^2}$. Although most meta-analyses are interested in learning about θ and τ^2, it is informative to examine the behavior of (4) as a function of τ. The weighting-factor $B(\tau)$ controls how much the estimate of $\hat{\theta}_i$ "shrinks" towards the population mean effect. When τ is taken to be 0, $B(\tau)$ is 1 and the random-effects model is equivalent to the fixed-effects model. Small values of τ describe situations in which studies can "borrow strength" from each other and information is meaningfully shared across studies. Very large values of τ, on the other hand, imply that $B(\tau)$ is close to 0, so that not much is gained in combining the studies. The Bayesian approach to meta-analysis particularly through the specification of a probability distribution for τ provides a formalism for investigating how similar the studies are and for sharing information among the studies when they are not identical. Many practical issues in the implementation of Bayesian meta-analysis are discussed and illustrated in DuMouchel and Normand (2000).

2.2 Sensitivity Analysis

At every step in a research synthesis decisions and judgments are made that have an impact upon the conclusions and inferences drawn from that meta-analysis. Sometimes, a decision may be easy to defend: for example, the omission of a very poor quality study from the meta-analysis, or the use of a random-effects model instead of a fixed-effects model. At other times, a decision may be much less convincing: for example, the use of only the published literature, the omission of a study that appears to yield an unusual effect size estimate, or the use of one regression model instead of another. When the basis for a decision is tenuous, it is important to check whether plausible departures from that basis appreciably affect the conclusions; in other words, to check the sensitivity of the conclusions to small changes in the analysis.

The objective of sensitivity analysis is to evaluate the impact of the various decisions and judgments made at each step of a research synthesis. The following examples are

illustrative of the types of issues, posed as questions, that can arise at different steps in a meta-analysis:

1. How reliable is the information obtained by a single individual who identifies, reviews and abstracts the relevant literature? Should there be at least two individuals who conduct independent searches of the literature and assess the relevant results from the studies? Should the abstracters of the literature be blind to the authorship of the studies?

2. How do the conclusions change if we assume different statistical models for combining study-specific effects? Do the conclusions change if we use a weighted versus an unweighted summary measure of effect-size? How sensitive are the inferences to distributional assumptions, including the assumption of independence of studies? How sensitive are the estimates of θ_i and θ to τ^2?

3. Which studies are outliers? How sensitive is the combined estimate of effect size to any one particular study? In other words, which studies are influential in the sense that the conclusions would change if that study was not included? What features of a study, for example, randomized versus not randomized, blind versus not blind, characterize influential studies?

4. If only statistically significant studies are published in the literature, then the studies that have appeared in the published literature are not a representative sample of all studies addressing the specific research question, a problem known as publication bias. What are the effects of publication bias on the conclusions of a research synthesis? Can the magnitude and the impact of publication bias be assessed in a research synthesis? What is the effect on the results of a research synthesis if the retrieval of the literature is not exhaustive?

5. How robust are statistically significant associations found in a meta-analysis? Meta-analysis is an observational study, and as such, is subject to the influence of uncontrolled variables. How sensitive are the conclusions of a meta-analysis to the

effects of confounding variables, either measured or unmeasured? Could alternative hypotheses explain the observed associations?

Little has been written about methods for sensitivity analysis in research synthesis (see Greenhouse and Iyengar, 1991). The decisions made over the course of a meta-analysis have either a substantive or statistical character. By and large, the latter are easier to deal with than the former, because computer packages have made it possible to re-analyze data fairly easily, while some substantive decisions are irrevocable. For instance, early decisions regarding the focus of the primary and secondary research questions determine the nature of the literature search, the kind of information that is extracted from the retrieved literature, and the way that information is processed for later synthesis. In this context, changing the goals in mid-stream calls for an entirely new meta-analysis. Despite that, it may still be useful to ponder what might have happened if, for example, the search for primary studies were more complete.

3. Meta-Analysis for Explanation

Research synthesis plays a central role in the process of scientific discovery with meta-analysis providing a formal methodology for the systematic accumulation and evaluation of scientific evidence. As Cooper and Hedges (1994b) note "Research syntheses attempt to integrate empirical research for the purpose of creating generalizations. Implicit in this definition is the notion that seeking generalizations also involves seeking the limits and modifiers of generalizations." Perhaps the single most important strength of research synthesis over individual studies, is the ability to investigate the robustness of a relationship and the conditions moderating its magnitude and direction. For example, in 1955, Beecher used a meta-analysis of 15 studies to demonstrate the consistency of the placebo-response and the effectiveness of placebo in the relief of pain across a number of different conditions, including severe postoperative pain, angina, and headache among others (Mosteller and Chalmers 1992). By comparing results across studies it is possible to investigate systematic variation in effect size due to person characteristics, design features, outcomes, and time. The contemporary practice of meta-analysis goes beyond the

simple description and estimation of effects to address the problem of explaining why, or under what conditions a given effect can be observed.

The achievements of meta-analysis have been considerable and its use in many disciplines, especially in the social, behavioral and medical sciences, continues to grow. This growth is facilitated, in part, by the existence of electronic bibliographic databases such as ERIC, PsychINFO and MEDLINE, and initiatives such as the Cochrane Collaboration, an international organization that prepares, maintains and ensures the accessibility of systematic reviews of the effects of health care interventions (Chalmers et al. 1997). Meta-analysis has become a standard methodology in the researcher's and policy maker's tool box.

3.1 A New Conceptual Framework for Meta-Analysis

Rubin (1992) has proposed a perspective on meta-analysis which is useful to consider and contrast with the goals of a traditional quantitative literature review. The idea is that the objective of combining information across studies is to build and extrapolate response surfaces in an attempt to estimate "true effects" from ideal studies rather than estimate effects from some population of "flawed" studies. In this approach, each study can be classified by two types of characteristics: S, which are variables of scientific interest (e.g., dose of treatment, subject characteristics, etc.) and D, which are design variables (e.g., sample size, whether the study was randomized or not, investigator characteristics, etc.). If T represents the observed outcome for each study, then in principle a response surface model can be built for T given (S, D) using the observed studies. However, the region where the data can fall is not necessarily of primary scientific interest because it reflects the various choices that were made by necessity about values of D. Interest would be in the extrapolated response surface, $E(T \mid S, D = D_0)$ with D fixed at specified ideal values, say D_0. The details of implementation of such an approach remain to be developed. Nevertheless, Rubin suggests a number of intriguing practical implications that follow from this perspective, such as, there is no need to be selective about which studies to include in a meta-analysis since all studies contribute to the estimate of the response surface and modeling will automatically downweight poorer studies. Also, using princi-

ples of experimental design (see *Linear Hypothesis: Experimental Design*) this approach provides guidance in the choice of new studies to add to the data base based on choosing studies, i.e., values of S and D, that increase the precision of estimation of the response surface, $E(T \mid S, D = D_0)$. The basis for this conceptualization is grounded in ideas underlying response surface methodology, hierarchical statistical models, and statistical methods for causal inference and for missing data. (See *Causal Inference and Statistical Fallacies* and *Missing Data*).

3.2 Applications of Meta-Analysis in Practice: Vignettes

The following vignettes are meant to provide a flavor for some of the more contemporary applications of meta-analysis and the central role that meta-analysis plays in explanation, program evaluation, and in informing policy decisions. Additional illustrations of this kind can be found in *Meta-Analysis for Explanation: A Casebook* (Cook et al. 1992) which presents four detailed case studies of research syntheses along with expert commentary highlighting the value of meta-analysis for scientific explanation and the challenges of implementation.

- Meta-analyses that include moderator variables provide an opportunity to investigate relationships that may never have been considered in any of the primary studies. Consider, for example, a meta-analysis based on 120 studies of whether or not homework worked. Cooper (1989) found that students given homework assignments did better than those not given assignments, and further, the more time spent on homework the better. An unexpected finding that had not been anticipated in the primary studies, however, was that although homework was very effective in high school it was almost totally ineffective in elementary school. This discovery has had important practical implications, and further, has stimulated the consideration of hypotheses as to why this might be, illustrating the impact of what Cooper has called "review-generated evidence".

- *Cross-design synthesis* is a relatively new research strategy for combining the results from diverse, complementary studies, such as randomized controlled studies

and administrative data bases, to improve knowledge about interventions and to inform policy decisions. The goal is to capture the diverse strengths of the different designs while minimizing weaknesses. The approach is anchored in meta-analytic principles and methods, but in addition, relies on methods for assessing the generalizability of data sources, and methods for statistical adjustment for potential biases in intervention comparisons (Droitcour et al. 1993).

- The goal of the Agency for Health Care Policy and Research (AHCPR) is to enhance the quality, appropriateness, and effectiveness of health services and to improve access to health care. The Federal legislation authorizing the creation of the agency (P.L. 101-239, December 1989) directs the agency to issue guidelines on clinical practice and *requires* the guidelines to be based, in part, on a systematic synthesis of research evidence. Large-scale, AHCPR multidisciplinary studies called PORTS (Patient Outcomes Research Teams) as part of this mandate carry out formal meta-analyses to determine what is known, what is not known and what needs to be learned about variations in clinical practice and outcomes for a particular disease or condition (see, for example, Lehman et al. 1995).

- When individual subject-level data is available from a number of similar studies, meta-analytic techniques can be used to combine information across the studies. Not only can study specific effects be estimated but individual subject characteristics can be used as covariates. An example of the use of such a meta-analysis is the study of nearly 600 outpatients treated for acute depression, who ranged in age from 18-80 years, from 6 studies conducted during a 10 year period. The goal of this study was to investigate the differential efficacy of psychotherapy alone, in comparison with the combination of psychotherapy and pharmacotherapy (Thase et al. 1997).

- Given the body of research available and the quality of that research with respect to a particular question, it may be determined that doing a meta-analysis is premature for that field. Nevertheless, a research synthesis can be extremely useful in identifying obstacles in the literature that need to be addressed and successes

that can be built on to guide future studies. For example, the Committee on the Assessment of Family Violence Interventions (1998) concluded that "The quality of the existing research base of evaluations of family violence interventions is therefore insufficient to provide confident inferences to guide policy and practice, except in a few areas that we identify in this report. Nevertheless, this pool of evaluation studies and additional review articles represents a foundation of research knowledge that will guide the next generation of family violence evaluation efforts and allows broad lessons to be derived from the research literature."

4. Suggestions for Further Reading

The entries on *Meta-analysis: Overview* and *Meta-analysis: Tools* provide further details on analytical aspects of doing meta-analysis. For more information on the practice of meta-analysis, see Cook et al. (1992), Cooper and Hedges (1994a), Gaver et al. (1992), Hunt (1997), Rosenthal (1984), and Stangl and Berry (2000).

Bibliography

Chalmers I, Sackett D, Silagy C 1997 The Cochrane collaboration. In: Maynard A and Chalmers I (eds.) *Non-random reflections on health services research: On the 25th anniversary of Archie Cochran's Effectiveness and Efficiency* BMJ Books, London

Committee on the Assessment of Family Violence Interventions 1998 *Violence in Families: Assessing Prevention and Treatment Programs. A Report of the National Research Council/Institute of Medicine*, Chalk R and King P A (eds.) National Academy Press, Washington, DC

Cooper, H 1989 *Homework* Longman, New York

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

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

Cooper H, Hedges L V 1994b Research synthesis as a scientific enterprise. In: Cooper H M and Hedges L (eds.) *The Handbook of Research Synthesis*, Russell Sage Foundation, New York

Draper D, Hodges J S, Mallows C L, Pregibon D L 1993 Exhcangeability and Data Analysis (with discussion). *Journal Royal Statistical Society, A* 156:9-37

Droitcour J, Silberman G, Chemlimsky E 1993 Cross-design synthesis: A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *International Journal of Technology Assessment in Health Care* 9:440-449.

DuMouchel W, Normand S-L 2000 Computer-modeling and graphical strategies for meta-analysis. In: Stangl D K and Berry D A (eds.) *Meta-analysis in Medicine and Health Policy*, Marcel Dekker, Inc., New York

Gaver D P, Draper D, Goel P K, Greenhouse J B, Hedges L V, Morris C N, Waternaux C 1992 *Combining Information: Statistical Issues and Opportunities for Research.* National Academy Press, Washington, D.C.

Greenhouse J B, Iyengar S 1994 Sensitivity analysis and diagnostics: Issues and methods. In: Cooper H M and Hedges L (eds.) *The Handbook of Research Synthesis*, Russell Sage Foundation, New York

Hunt M 1997 *How Science Takes Stock: The Story of Meta-Analysis*, Russell Sage Foundation, New York

Lehman A F, Thompson J W, Dixon L B, Scott J E (eds.) 1995 *Schizophrenia Bulletin. Special Issue Theme: Schizophrenia: Treatment Outcomes Research* 21:561-676

Mosteller F, Chalmers T C 1992 Some progress and problems in meta-analysis of clinical trials. *Statistical Science* 7:227-236

Rosenthal R 1984 *Meta-Analytic Procedures for Social Research* Sage Publications, Beverly Hills

Rubin D B 1992 Meta-Analysis: Literature synthesis or effect-size surface estimation? *Journal of Educational Statistics* 17:363-374

Stangl D K, Berry D A (eds.) 2000 *Meta-Analysis in Medicine and Health Policy* Marcel Dekker, Inc., New York

Thase M E, Greenhouse J B, Frank E, Reynolds C F, Pilkonis P A, Hurley K, Grochocinski V, Kupfer D J 1997 Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Archives of General Psychiatry* 54:1009-1015