

1

RESEARCH SYNTHESIS AS A SCIENTIFIC PROCESS

HARRIS COOPER

Duke University

LARRY V. HEDGES

Northwestern University

JEFFREY C. VALENTINE

University of Louisville

C O N T E N T S

1.1 Introduction	4
1.1.1 Replication and Research Synthesis	4
1.2. Research Synthesis in Context	4
1.2.1 A Definition of the Literature Review	4
1.2.1.1 Definitions of Research Synthesis	6
1.3. A Brief History of Research Synthesis as a Scientific Enterprise	7
1.3.1 Early Developments	7
1.3.2 Research Synthesis Comes of Age	8
1.3.3 Rationale for the <i>Handbook</i>	11
1.4 Stages of Research Synthesis	12
1.4.1 Formulating a Problem for a Research Synthesis	12
1.4.2 Searching the Literature	12
1.4.3 Evaluating Study Methodology and Extracting Information from Study Reports	13
1.4.4 Statistically Describing and Combining Study Outcomes	13
1.4.5 Interpreting Synthesis Outcomes	14
1.4.6 Presenting Synthesis Results	15
1.5 References	15

It is necessary, while formulating the problems of which in our further advance we are to find the solutions, to call into council the views of those of our predecessors who have declared an opinion on the subject, in order that we may profit by whatever is sound in their suggestions and avoid their errors.

Aristotle, *De Anima*, Book 1, Chapter 2

1.1 INTRODUCTION

From the moment we are introduced to science, we are told it is a cooperative, cumulative enterprise. Like the artisans who construct a building from blueprints, bricks, and mortar, scientists contribute to a common knowledge structure. Theorists provide the blueprints and researchers collect the data that are the bricks. To extend the analogy further yet, we might say that research synthesists are the bricklayers and hodcarriers of the science guild. It is their job to stack the bricks according to the blueprints and apply the mortar that allows the structure to take shape.

Anyone who has attempted a research synthesis is entitled to a wry smile as the analogy continues. They know that several sets of theory-blueprints often exist, describing structures that vary in form and function, with no a priori criteria for selecting between them. They also know that our data-bricks are not all six-sided with right angles. They come in a baffling array of sizes and shapes. Making them fit, securing them with mortar, and seeing whether the resulting construction looks anything like what the blueprint suggests is a challenge worthy of the most dedicated, inspired artisan.

1.1.1 Replication and Research Synthesis

Scientific literatures are cluttered with repeated studies of the same phenomena. Multiple studies on the same problem or hypothesis arise because investigators wish to verify and extend (that is, generalize or search for influences on) previous findings. Experience has shown that even when considerable effort is made to achieve direct replication, results across studies are rarely identical at any high level of precision (Valentine et al. 2011; Open Science Collaboration 2015), even in the physical sciences (Hedges 1987). No two bricks are exactly alike. Still, the value and need for replication in the social sciences has received increased attention recently, due in part to concerns about questionable data practices, such as selective reporting of findings. For example, the journal *Perspectives on Psychological Science* published a special issue on replication (Pashler and Wagonmakers 2012).

How should scientists proceed when study results differ? First, it is clear how they should *not* proceed. They should not decide that results are not replicated simply because some results reject the null hypothesis and the others do not, in part because the outcome of null hypothesis significance tests does not imply a difference in effect size (Gelman and Stern 2006). Differences in statistical power might explain this, as well as expected sampling variation. Even results suggesting that the relation of interest is in different directions are predictable, depending on the size of the underlying effect, its sensitivity to contextual variation, and the number of times it has been tested. Certainly, scientists should not decide that one study (perhaps the most recent one, or the one they conducted, or a study chosen via some other equally arbitrary criterion) produces the correct finding and others can be ignored. If results that are expected to be similar show variability, the scientific instinct should be to account for the variability by further systematic work. This is where research synthesis comes in.

1.2 RESEARCH SYNTHESIS IN CONTEXT

1.2.1 A Definition of the Literature Review

The American Psychological Association's *PsycINFO* reference database defines a literature review as "the process of conducting surveys of previously published material" (<http://psycnet.apa.org/psycinfo/1994-97192-000>). Common to all definitions of literature reviews is the notion that they are "not based primarily on new facts and findings, but on publications containing such primary information, whereby the latter is digested, sifted, classified, simplified, and synthesized" (Manten 1973, 75).

Table 1.1 presents a taxonomy of literature reviews that capture six distinctions that review authors use to describe their own work (Cooper 1988). The taxonomy can be applied to literature reviews appearing throughout a broad range of both the behavioral and physical sciences. The six features and their subordinate categories permit a rich level of distinction among works of synthesis.

Table 1.1 A Taxonomy of Literature Reviews

Characteristic	Categories
Focus	Research findings Research methods Theories Practices or applications
Goal	Integration Generalization Conflict resolution Linguistic bridge-building Criticism Identification of central issues
Perspective	Neutral representation Espousal of position
Coverage	Exhaustive Exhaustive with selective citation Representative Central or pivotal
Organization	Historical Conceptual Methodological
Audience	Specialized scholars General scholars Practitioners or policy makers General public

SOURCE: Cooper 1988. Reprinted with permission from Transaction Publishers.

The first distinction among literature reviews concerns the *focus* of the review, the material that is of central interest to the reviewer. Most literature reviews center on one or more of four areas: the findings of individual primary studies, not necessarily but often empirical in nature; the methods used to carry out research; theories meant to explain the same or related phenomena; and the practices, programs, or treatments being used in an applied context.

The second characteristic of a literature review is its *goals*. Goals concern what the preparers of the review hope to accomplish. The most frequent goal for a review is to *integrate* past literature that is believed to relate to a common topic. Integration includes formulating general statements that characterize multiple specific instances (or research, methods, theories, or practices); resolving conflict between contradictory research results, ideas, or statements of fact by proposing a new conception that accounts for the inconsistency; and bridging the gap

between concepts or theories by creating a new, common linguistic framework.

Another goal for literature reviews can be to *critically analyze the existing literature*. Unlike a review that seeks to integrate the existing work, one that involves a critical assessment does not necessarily summate conclusions or compare the covered works one with another. Instead, it holds each work up against a criterion and finds it more or less acceptable. Most often, the criterion will include issues related to the methodological quality of empirical studies; the logical rigor, completeness, or breadth of explanation if theories are involved; or comparison with the ideal treatment, when practices, policies, or applications are involved.

A third goal that often motivates literature reviews is to *identify issues central to a field*. These issues may include questions that have given rise to past work, questions that should stimulate future work, or methodological problems or problems in logic and conceptualization that have impeded progress within a topic area or field.

Of course, reviews more often than not have multiple goals. So, for example, it is rare to see a review that integrates or critically examines existing work without identifying central issues for future endeavors.

A third characteristic that distinguishes among literature reviews, *perspective*, relates to whether the reviewers have an initial point of view that might influence the discussion of the literature. The endpoints on the continuum of perspective might be called *neutral representation* and *espousal of a position*. In the former, reviewers attempt to present all arguments or evidence for and against various interpretations of the problem. The presentation is meant to be as similar as possible to those that would be provided by the originators of the arguments or evidence. At the opposite extreme of perspective, the viewpoints of reviewers play an active role in how material is interpreted and presented. The reviewers accumulate and synthesize the literature in the service of demonstrating the value of the particular point of view that they espouse. The reviewers muster arguments and evidence so that it presents their contentions in the most convincing manner.

Of course, reviewers attempting to achieve complete neutrality are likely doomed to failure. Further, reviewers who attempt to present all sides of an argument do not preclude themselves from ultimately taking a strong position based on the cumulative evidence. Similarly, reviewers can be thoughtful and fair while presenting conflicting evidence or opinions and still advocate for a particular interpretation.

The next characteristic, *coverage*, concerns the extent to which reviewers find and include relevant works in their paper. It is possible to distinguish at least four types of coverage. The first type, *exhaustive coverage*, suggests that the reviewers hope to be comprehensive in the presentation of the relevant work. An effort is made to include the entire literature and to base conclusions and discussions on this comprehensive information base. The second type of coverage also bases conclusions on entire literatures, but only a selection of works is actually described in the literature review. The authors choose a purposive sample of works to cite but claim that the inferences drawn are based on a more extensive literature. Third, some reviewers will present works that are broadly representative of many other works in a field. They hope to describe just a few exemplars that are descriptive of numerous other works. The reviewers discuss the characteristics that make the chosen works paradigmatic of the larger group. In the final coverage strategy, reviewers concentrate on works that were highly original when they appeared and influenced the development of future efforts in the topic area. These may include materials that initiated a line of investigation or thinking, changed how questions were framed, introduced new methods, engendered important debate, or performed a heuristic function for other scholars.

A fifth characteristic of literature reviews concerns a paper's *organization*. Reviews may be arranged historically, so that topics are introduced in the chronological order in which they appeared in the literature; conceptually, so that works relating to the same abstract ideas appear together; or methodologically, so that works employing similar methods are grouped together.

Finally, the intended *audiences* of reviews can vary. Reviews can be written for groups of specialized researchers, general researchers, policymakers, practitioners, or the general public. As reviewers move from addressing specialized researchers to addressing the general public, they use less technical jargon and detail and often pay greater attention to the implications of the work being covered.

1.2.1.1 Definitions of Research Synthesis The terms *research synthesis* or *research review* or *systematic review* are often used interchangeably in the social science literature, though they sometimes connote subtly different meanings. Regrettably, no consensus has been reached about what these meaningful differences might be. Therefore, we use the term *research synthesis* most frequently throughout this book. The reason for this choice is sim-

ple. In addition to its use in the context of research synthesis, the term *research review* is also used to describe the activities of evaluating the quality of research. For example, a journal editor will obtain research reviews when deciding whether to publish a manuscript. Because research syntheses often include this type of evaluative review of research, using the term *research synthesis* avoids confusion. The term *systematic review* is less often used in the context of research evaluation, though the confusion is still there, and the specification that it is the results of research that are being synthesized is missing. The Cochrane Collaboration uses *systematic review* but has moved toward using *Cochrane review* to signify the use of its distinct tools and methodology (<http://community.cochrane.org>). The Campbell Collaboration (<http://www.campbellcollaboration.org>) also uses the term *systematic review* to label all its reviews, whether quantitative or qualitative.

A research synthesis can be defined as the conjunction of a particular set of literature review characteristics. Most distinctive about research syntheses are their primary focus and goal: 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 of generalizations. Also, research syntheses almost always pay attention to relevant theories, critically analyze the research they cover, try to resolve conflicts in the literature, and attempt to identify central issues for future research. According to Derek Price, research syntheses are intended to “replace those papers that have been lost from sight behind the research front” (1965, 513). Research synthesis is one of a broad array of integrative activities that scientists engage in; its intellectual heritage can be traced back at least as far as Aristotle.

Using the described taxonomy, we can make further specifications concerning the type of research syntheses that are the focus of this book. With regard to perspective, readers will note that much of the material is meant to help synthesists produce neutral statements about evidence, that is, avoid being affected by many types of bias including their own subjective outlooks. For example, the material on searching the literature for evidence is meant to help synthesists uncover all the evidence, not simply positive studies that might be overrepresented in published research, or evidence that is easy for them to find and therefore might be overly sympathetic to their point of view. The material on the reliability of extracting information from research reports and how methodologi-

cal variations in research should be handled is meant to increase transparency and interjudge reliability when these activities are carried out. The methods proposed for the statistical integration of findings are meant to ensure the same rules about data analysis are applied to the next users of data as were required of the data generators. Finally, the material on explicit and exhaustive reporting of methods is meant to assist both producers and consumers of research syntheses in evaluating if or where bias may have crept into the synthesis process and to replicate findings if they choose to do so.

Finally, the term *meta-analysis* often is used as a synonym for research synthesis. However, in this volume, it is used in its more precise and original meaning—to describe the quantitative procedures that a research synthesist may use to statistically combine the results of studies. Gene Glass coined the term *meta-analysis* to refer to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (1976, 3). The authors of this book reserve meta-analysis to refer specifically to statistical analysis in research synthesis and not to the entire enterprise of research synthesis. Not all research syntheses are appropriate for meta-analysis.

1.3 A BRIEF HISTORY OF RESEARCH SYNTHESIS AS A SCIENTIFIC ENTERPRISE

1.3.1 Early Developments

In 1971, Kenneth Feldman published an article titled “Using the Work of Others” in which he demonstrated remarkable prescience: “Systematically reviewing and integrating . . . the literature of a field may be considered a type of research in its own right—one using a characteristic set of research techniques and methods” (86). He described four steps in the reviewing process: sampling topics and studies, developing a scheme for indexing and coding material, integrating the studies, and writing the report.

The same year, Richard Light and Paul Smith presented what they called a cluster approach to literature reviewing that was meant to redress some of the deficiencies in the existing strategies for integration (1971). They argued that, if treated properly, the variation in outcomes among related studies could be a valuable source of information rather than merely a source of consternation, as it appeared to be when treated with traditional reviewing methods.

Three years later, Thomas Taveggia struck a complementary theme. He wrote,

A methodological principle overlooked by [reviewers] . . . is that research results are probabilistic. What this principle suggests is that, in and of themselves, the findings of any single research are meaningless—they may have occurred simply by chance. It also follows that, if a large enough number of researches has been done on a particular topic, chance alone dictates that studies will exist that report inconsistent and contradictory findings! Thus, what appears to be contradictory may simply be the positive and negative details of a distribution of findings. (1974, 397–98)

Taveggia went on to describe six common tasks in research syntheses: selecting research; retrieving, indexing, and coding information from studies; analyzing the comparability of findings; accumulating comparable findings; analyzing distributions of results, and; reporting of results.

The development of meta-analytic techniques extends back further in time but their routine use by research synthesists is also relatively recent. Where Glass gave us the term meta-analysis in 1976, in 1990 Ingram Olkin pointed out that ways to estimate effect sizes have existed since the turn to the twentieth century. For example, Karl Pearson took the average of estimates from five separate samples of the correlation between inoculation for enteric (or typhoid) fever and mortality (1904). He used this average to better estimate the typical effect of inoculation and to compare it with that of inoculation for other diseases. Early work on the methodology for combination of estimates across studies includes papers in the physical sciences by Raymond Birge (1932) and in statistics by William Cochran (1937) and Frank Yates and Cochran (1938). Although they have fallen out of use today, methods for combining probabilities across studies also have a long history (Tippett 1931; Fisher 1932; Mosteller and Bush 1954).

Still, the use of quantitative synthesis techniques in the social sciences was rare before the 1970s. Late in that decade, several applications of meta-analytic techniques captured the imagination of behavioral scientists. Included among these were: in clinical psychology, Mary Smith and Gene Glass’s meta-analysis of psychotherapy research (1977); in industrial-organizational psychology, Frank Schmidt and John Hunter’s validity generalization of employment tests (1977); in social psychology, Robert Rosenthal and Donald Rubin’s integration of interpersonal expectancy effect research (1978); and in education,

Glass and Smith's synthesis of the literature on class size and achievement (1978).

1.3.2 Research Synthesis Comes of Age

Two papers that appeared in the *Review of Educational Research* in the early 1980s brought the meta-analytic and research synthesis-as-research perspectives together. The first, by Gregg Jackson, proposed six reviewing tasks "analogous to those performed during primary research" (1980, 441). Jackson portrayed meta-analysis as an aid to the task of analyzing primary studies but emphasized its limitations as well as its strengths. Also noteworthy about his paper was his use of a sample of thirty-six review articles from prestigious social science periodicals to examine the methods used in integrative empirical reviews. For example, Jackson reported that only one of the thirty-six reported the indexes or retrieval systems used to locate primary studies. His conclusion was that "relatively little thought has been given to the methods for doing integrative reviews. Such reviews are critical to science and social policy making and yet most are done far less rigorously than is currently possible" (459).

The first half of the 1980s also witnessed the appearance of four books primarily devoted to meta-analytic methods. The first, in 1981, by Glass, Barry McGaw, and Smith, presented meta-analysis as a new application of analysis of variance and multiple regression procedures, with effect sizes treated as the dependent variable. In 1982, Hunter, Schmidt, and Jackson introduced meta-analytic procedures that focused on comparing the observed variation in study outcomes to that expected by chance (the statistical realization of a point Taveggia made in 1974) and correcting observed effect-size estimates and their variance for known sources of bias (such as sampling error, range restrictions, unreliability of measurements). In 1984, Rosenthal presented a compendium of meta-analytic methods covering, among other topics, the combining of significance levels, effect-size estimation, and the analysis of variation in effect sizes. Rosenthal's procedures for testing moderators of variation in effect sizes were not based on traditional inferential statistics, but on a new set of techniques involving assumptions tailored specifically for the analysis of study outcomes. Finally, in 1985, with the publication of *Statistical Procedures for Meta-Analysis*, Larry Hedges and Olkin helped elevate the quantitative synthesis of research to an independent specialty within the statistical sciences. This book, summarizing and expanding nearly a decade

of programmatic developments by the authors, not only covered the widest array of meta-analytic procedures but also presented rigorous statistical proofs establishing their legitimacy.

Harris Cooper drew the analogy between research synthesis and primary research to its logical conclusion and presented a five-stage model of the integrative review as a research project (1982). For each stage, he codified the research question, its primary function in the review, and the procedural differences that might cause variation in review conclusions. In addition, he applied the notion of threats-to-inferential-validity—which Donald Campbell and Julian Stanley introduced for evaluating the utility of primary research designs (1966)—to the conduct of research synthesis (also see Shadish, Cook, and Campbell 2002). Cooper identified ten threats to validity specifically associated with reviewing procedures that might undermine the trustworthiness of the findings of a research synthesis. He also suggested that other threats might exist and that any particular synthesis' validity could be threatened by consistent deficiencies in the set of studies that formed its database. Table 1.2 presents a recent revision of this schema, which proposes a seven-stage model for conducting a research synthesis, separating the original coding stage into coding and study evaluation, the analysis stage into separate analyses, and interpretation stage into two distinct stages (Cooper 2017).

Another text that appeared in 1984 also helped elevate research synthesis to a more rigorous level. In it, Light and David Pillemer focused on the use of research reviews to help decision making in the social policy domain. Their approach placed special emphasis on the importance of meshing both numbers and narrative for the effective interpretation and communication of synthesis results.

Numerous books have appeared on research synthesis and meta-analysis since the mid-1980s—in fact, too many to mention all of them. Some focus on research synthesis in general (Card 2012; Lipsey and Wilson 2001; Petticrew and Roberts 2006; Schmidt and Hunter 2015); others treat it from the perspective of particular research designs (Bohning, Kuhnert, and Rattanasiri 2008; Eddy, Hassleblad, and Schachter 1992). Still others are tied to particular software packages (Arthur, Bennett, and Huffcutt 2001; Chen and Peace 2013; Comprehensive Meta-Analysis 2015). In 1994, the first edition of this book was published; the second edition appeared in 2009. Readers interested in a popular history of the origins of meta-analysis in the social sciences can consult

Table 1.2 Research Synthesis Conceptualized as a Research Process

Step in Research Synthesis	Research Question Asked at This Stage of the Synthesis	Primary Function Served in the Synthesis	Procedural Variation That Might Produce Differences in Conclusions
Formulating the problem	What research evidence will be relevant to the problem or hypothesis of interest in the synthesis?	Define the variables and relationships of interest so that relevant and irrelevant studies can be distinguished	Variation in the conceptual breadth and distinctions within definitions might lead to differences in the research operations deemed relevant and/or tested as moderating influences
Searching the literature	What procedures should be used to find relevant research?	Identify sources (such as reference databases, journals) and terms used to search for relevant research	Variation in searched sources might lead to systematic differences in the retrieved research
Gathering information from studies	What information about each study is relevant to the problem or hypothesis of interest?	Collect relevant information about studies in a reliable manner	Variation in information gathered might lead to differences in what is tested as an influence on cumulative results, in coder training might lead to differences in entries on coding sheets, or in rules for deciding what study results are independent tests of hypotheses might lead to differences in the amount and specificity of data used to draw cumulative conclusions
Evaluating the quality of studies	What research should be included in the synthesis based on the suitability of the methods for studying the synthesis question or problems in research implementation?	Identify and apply criteria that separate studies conducted in ways that correspond with the research question from studies that do not	Variation in criteria for decisions about study methods to include might lead to systematic differences in which studies remain in the synthesis
Analyzing and integrating the outcomes of studies	What procedures should be used to condense and combine the research results?	Identify and apply procedures for combining results across studies and testing for differences in results between studies	Variation in procedures used to summarize and compare results of included studies (such as narrative, vote count, averaged effect sizes) can lead to differences in cumulative results
Interpreting the evidence	What conclusions can be drawn about the cumulative state of the research evidence?	Summarize the cumulative research evidence with regard to its strength, generality, and limitations	Variation in criteria for labeling results as important and attention to details of studies might lead to differences in interpretation of findings
Presenting the results	What information should be included in the report of the synthesis?	Identify and apply editorial guidelines and judgment to determine aspects of methods and results readers of the report will need to know	Variation in reporting might lead readers to place more or less trust in synthesis outcomes and influence others' ability to replicate results

SOURCE: Authors' compilation.

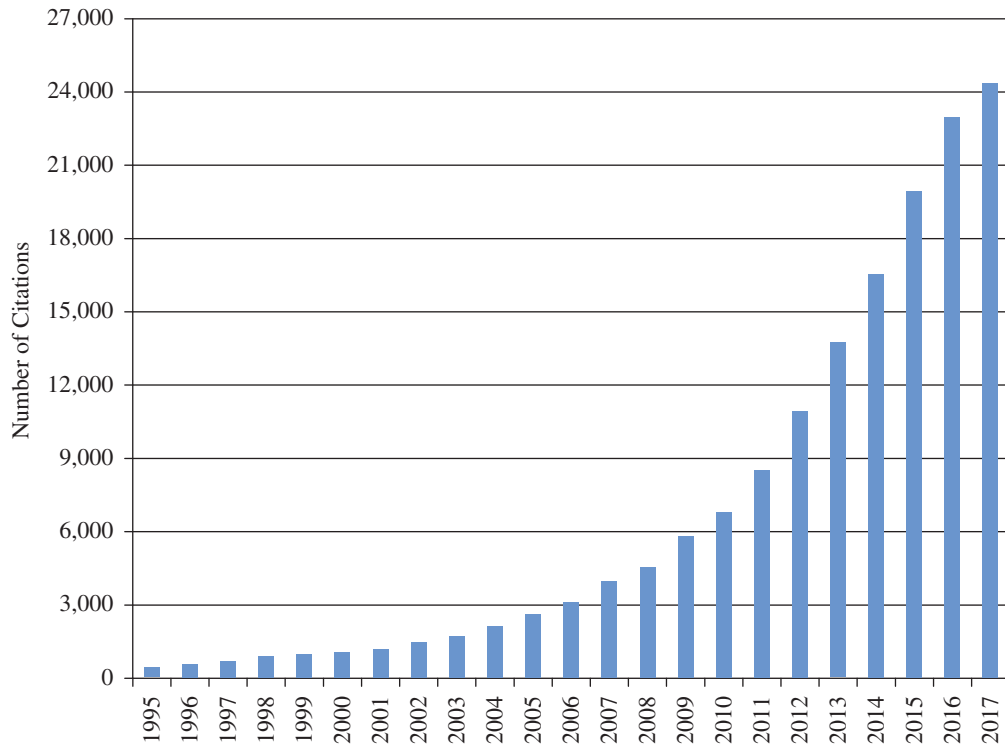


Figure 1.1 Citations to Articles Including the Terms *Research Synthesis*, *Systematic Review*, or *Meta-Analysis* in Their Titles

SOURCE: Authors' compilation based on Web of Science Core Collection (Clarivate Analytics 2018).

NOTE: Bars chart the growth in the number of titles of documents including the terms *research synthesis*, *systematic review*, or *meta-analysis* during the years following the publication of the first edition of the *Handbook of Research Synthesis*.

How Science Takes Stock: The Story of Meta-Analysis (Hunt 1997). Most recently, the journal *Research Synthesis Methods* published a special issue on the origins of modern meta-analysis (Shadish 2015).

Literally thousands of research syntheses have been published since the first edition of this book. Figure 1.1 presents some evidence of the increasing impact of research syntheses on knowledge in the sciences and social sciences. The figure is based on entries in the Web of Science Core Collection reference database (Clarivate Analytics 2018). It charts the growth in the number of document titles including the terms *research synthesis*, *systematic review*, or *meta-analysis* in their title or during the years 1995 to 2017. The figure indicates that documents in the database titles using these terms has risen

every year without exception and the growth is accelerating. Clearly, the role that research syntheses play in our knowledge claims is large and growing larger.

The use of research synthesis has spread from psychology and education through many disciplines, especially in medicine, social policy analysis, and economics. Indeed, the development of scientific methods for research synthesis has its own largely independent history in the medical sciences (see Chalmers, Hedges, and Cooper 2002). A most notable event in medicine was the establishment of the UK Cochrane Center in 1992. The center was meant to facilitate the creation of an international network to prepare and maintain systematic reviews of the effects of interventions across the spectrum of health-care practices. At the end of 1993, an international network of

individuals, the Cochrane Collaboration (<http://www.cochrane.org>), emerged from this initiative (Chalmers 1993; Bero and Rennie 1995). By 2019, the Cochrane Collaboration was an internationally renowned initiative that counted thousands of people in more than ninety countries contributing to its work. The Cochrane Collaboration is now the leading producer of research syntheses in health care and is considered by many to be the gold standard for determining the effectiveness of different health-care interventions. Its library of systematic reviews numbers in the thousands. In 2000, an initiative known as the Campbell Collaboration was launched with similar objectives for the domain of social policy analysis, focusing initially on policies concerning education, social welfare, and crime and justice (<http://www.campbellcollaboration.org>).

Because of the efforts of scholars who chose to apply their skills to how research syntheses might be improved, syntheses written since the 1980s have been held to standards far more demanding than those applied to their predecessors. The process of elevating the rigor of syntheses has continued into the twenty-first century.

1.3.3 Rationale for the *Handbook*

The Handbook of Research Synthesis and Meta-Analysis is meant to be the definitive *vade mecum* for behavioral and social scientists intent on applying the synthesis craft. It distills the products of forty years of developments in how research integrations should be conducted so as to minimize the chances of conclusions that do not truly reflect the cumulated evidence. Research synthesis in the 1960s was at best an art, at worst a form of yellow journalism. Today, the summarization and integration of studies is viewed as a research process in its own right, is held to the standards of a scientific endeavor, and entails the application of data gathering and analyses techniques developed for its unique purpose.

Numerous excellent texts on research synthesis exist. However, none is as comprehensive and detailed as this volume. Some texts focus on statistical methods. These often emphasize different aspects of statistical integration (such as combining probabilities, regression-analog models, estimating population effects from sampled effects with known biases) and often approach research accumulation from different perspectives. Although these texts are complete within their domains, no single sourcebook describes and integrates all the meta-analytic approaches in most common use.

This volume incorporates quantitative statistical techniques from all the synthesis traditions. It brings the leading authorities on the various meta-analytic perspectives together in a single volume. In doing so, it is an explicit statement that all the statistical approaches share a common assumptive base. This base is not only statistical but also philosophical. Philosophically, all the approaches rest on the presupposition that research syntheses need to be held to the same standards of rigor, systematicity, and transparency as the research on which they are based. The second and later users of data must be held as accountable for the validity of their methods as were the first.

Several problems arising in the course of conducting a quantitative synthesis have not received adequate treatment in any existing text. These include nonindependence of data sets, synthesis of multivariate data sets, and sensitivity analysis, to name just a few. Every research synthesist faces these problems and has developed strategies for dealing with them. Some of their solutions are published in widely scattered journals; others are often passed on to colleagues through informal contacts. They have never received complete treatment within the same text. This *Handbook* brings these topics together in a single volume.

Further, texts focusing on the statistical aspects of integration tend to give only passing consideration to other activities of research synthesis. These activities include the unique characteristics of problem formulation in research synthesis; methods of literature search; coding and evaluation of research reports; and the meaningful interpretation and effective communication of synthesis results. The existing texts that focus on these aspects of research synthesis tend not to be comprehensive in their coverage of statistical issues. Fully half of the chapters in this volume deal with issues that are not statistical in nature, evidencing the authors' collective belief that high-quality syntheses require considerably more than simple application of quantitative procedures.

Finally, this volume is meant for those who carry out research syntheses. Discussions of theory and proof are kept to a minimum in favor of descriptions of the practical mechanics needed to apply well the synthesis craft. The chapters include multiple approaches to problem solving and discuss the strengths and weaknesses of each approach. Readers with a comfortable background in analysis of variance and multiple regression and who have access to a research library should find the chapters accessible. The *Handbook* authors want to supply working synthesists with the needed expertise to interpret their

blueprints, to wield their mortar hoe and trowel as accurately as possible.

1.4 STAGES OF RESEARCH SYNTHESIS

The description of the stages of research synthesis presented in table 1.2 provides the conceptual organization of this book. In this section, we raise the principal issues associated with each stage. This allows us to briefly introduce the content of each of the chapters that follow.

1.4.1 Formulating a Problem for a Research Synthesis

The one major constraint on problem formulation in research synthesis is that primary research on a topic must exist before a synthesis can be conducted. How much research? The methods of meta-analysis can be applied to literatures containing as few as two hypothesis tests (Valentine, Pigott, and Rothstein 2010). Under certain circumstances—for instance, researchers synthesizing a pair of replicate studies from their own lab—the use of meta-analysis in this fashion might be sensible. Yet, most scientists would argue that the benefits of such a review would be limited (and its chances for publication even more limited).

A more general answer to the “How much research?” question is that it varies depending on a number of characteristics of the problem. All else being equal, conceptually broad topics would seem to profit from a synthesis only after the accumulation of a more varied and larger number of studies than a narrowly defined topic would (see chapter 2). Similarly, literatures that contain diverse types of operations also would seem to require a relatively large number of studies before firm conclusions could be drawn from a synthesis. Ultimately, the arbiter of whether a synthesis is needed will not be numerical standards, but the fresh insights a synthesis can bring to a field. Indeed, although a meta-analysis cannot be performed without data, many social scientists see value in “empty” syntheses that point to important gaps in our knowledge. When done properly, empty syntheses should proceed through the stages of research synthesis, including careful problem formulation.

Once enough literature on a problem has collected, then the challenge, and promise, of research synthesis becomes evident. The problems that constrain primary researchers—small and homogeneous samples, limited time and money for turning constructs of interest into

multiple operations—are less severe for synthesists. They can capitalize on the diversity in methods that has occurred naturally across primary studies. The heterogeneity of methods across studies may permit tests of theoretical hypotheses concerning the moderators and mediators of relations that have never been tested in any single primary study. Conclusions about the population and ecological validity of relations uncovered in primary research may also receive more thorough tests in syntheses.

Part II of this book focuses on issues in problem formulation. In chapter 2 (“Hypotheses and Problems in Research Synthesis”), Harris Cooper discusses in detail the issues just mentioned. In chapter 3 (“Statistical Considerations”), Larry Hedges looks at the implications of different problem definitions for how study results will be statistically modeled. The major issues involve the populations of people and measurements that are the target of a review’s inferences; how broadly the key constructs are defined, especially in terms of whether fixed- or random-effect models are envisioned; and how choices among models influence the precision of estimates and the statistical power of meta-analytic tests.

1.4.2 Searching the Literature

The literature search is the stage of research synthesis that is most different from primary research. Still, culling through the literature for relevant studies is not unlike gathering a sample of primary data. The target of a literature search that is part of a synthesis attempting exhaustive coverage would be “all the research conducted on the topic of interest.”

In contrast to the (relatively) well-defined sampling frames available to primary researchers, literature searchers confront the fact that any single source of primary reports will lead them to only a fraction of the relevant studies, and a biased fraction at that. For example, the most inclusive sources of literature are the reference databases, such as Google Scholar, Science Direct, PsycINFO, ERIC, and Medline. Still, many of these broad, nonevaluative systems exclude much of the unpublished literature. Conversely, the most exclusive literature searching technique involves accessing close colleagues and other researchers with an active interest in the topic area. Despite the obvious biases, there is no better source of unpublished and recent works. Further complicating the sampling frame problem is that the relative utility and biases associated with any single source will vary as a function of characteristics of the research problem, includ-

ing, for example, how long the topic has been the focus of study and whether it is interdisciplinary.

These problems imply that research synthesists must carefully consider multiple channels for accessing literature and how the channels they choose complement one another. The three chapters in part III are devoted to helping the synthesist consider and carry out this most unique task. In chapter 4 (“Scientific Communication and Literature Retrieval”), Howard White presents an overview of searching issues from the viewpoint of an information scientist. In chapter 5, “Searching Bibliographic Databases,” Julie Glanville provides strategies for using electronic databases (such as reference databases, citation indexes, research registries) to assist researchers with finding and accessing scholarship that is relevant to their work. In chapter 6 (“Retrieving Grey Literature, Information, and Data in the Digital Age”), Dean Giustini discusses the practical considerations of how to find research that is not indexed in the usual academic databases.

1.4.3 Evaluating Study Methodology and Extracting Information from Study Reports

Part IV offers four chapters on the evaluation of the study designs and implementation and retrieving information from studies. Once the synthesists have gathered the relevant literature, they must extract from each document those pieces of information that will help answer the questions that impel research in the field. This step includes judgments about the critical aspects of each study’s research design, measurements, and procedures, and how variations in these relate to the inferences the synthesists wish to make. The problems faced during data coding provide a strong test of the synthesists’ knowledge of the research area, thoughtfulness, and ingenuity. The decisions made during coding will have a profound influence on the contribution of the synthesis.

The aspect of coding studies that engenders the most debate involves how synthesists should represent differences in the design and implementation of primary studies that contribute to their data. What is meant by *study quality* when we are evaluating research methods? Should studies be given more or less credibility and therefore weighted differently in a meta-analysis if they differ in quality? Should studies be excluded if they contain too many flaws? How does one rate the quality of studies described in incomplete research reports? In chapter 7 (“Incorporating Judgments About Study Quality into

Research Syntheses”), Jeffrey Valentine examines the alternative approaches available to synthesists for representing primary research methodology.

But judging a studies credibility is only the beginning. Synthesists must make decisions about other classes of variables that are of potential interest to them. These can relate to variables that predict outcomes, potential moderators and mediators of effects, and the differences in how outcomes are conceptualized (and, therefore, measured). Might the type of participants and the context of the study influence its outcomes? What about characteristics of the experimental manipulation (for example, intensity, duration) and measurements (for example, reliability, timing)? If a synthesist chooses not to code a particular feature of studies, then it cannot be considered in the analysis of results.

General guidelines for what information should be extracted from primary research reports are difficult to develop beyond recommendations that are general and abstract. Instead, direction will come from the issues that have arisen in the particular literature, coupled with the synthesist’s personal insights into the topic. Still, commonalities emerge about what information is important to collect and how to think about what information to retrieve from studies. Mark Lipsey, in chapter 8 (“Identifying Potentially Interesting Variables and Analysis Opportunities”), and David Wilson, in chapter 9 (“Systematic Coding for Research Synthesis”), present complementing templates for what generally should be included on coding frames.

Once decisions on what to code have been made, synthesists need to consider how to carry out the coding (for example, who will retrieve information, how will they be trained) and how to assess the trustworthiness with which the coding frame is implemented. Numerous indexes of coder reliability are available, each with different strengths and weaknesses. In chapter 10 (“Evaluating Coding Decisions”), Jack Vevea, Nicole Zelinsky, and Robert Orwin describe strategies for reducing the amount of error that enters a synthesis during the coding of the literature’s features. Their description of reliability assessment focuses on three major approaches: sources of coding error, strategies for reducing coding error, and strategies for statistically assessing and quantifying coding error.

1.4.4 Statistically Describing and Combining Study Outcomes

As our brief history of research synthesis suggests, techniques for the analysis of accumulated research outcomes

is an area of statistics abundant in dramatic developments. Four decades ago, the mechanics of integrating research usually involved intuitive processes taking place inside the heads of the reviewers. Meta-analysis made these processes public and based them on explicit, shared, statistical assumptions (however well met). We would not accept as valid a primary researcher's conclusion if it were substantiated solely by the statement "I looked at the treatment and control scores and I found the treated group did better." We would demand statistical testing (for example, a simple *t*-test) to back up the claim. Likewise, we should no longer accept "I examined the study outcomes and find the treatment is effective" as sufficient warrant for the conclusion of a research synthesis.

Part V covers the components of synthesis dealing with combining study results. Chapter 11, by Michael Borenstein and Larry Hedges on effect sizes, covers methods for estimating the outcomes of studies using a common metric. Thirty years ago, Jacob Cohen defined an effect size as "the *degree* to which the phenomenon is present in the population, or the degree to which the null hypothesis is false" (1988, 9–10).

To most research synthesists, the search for influences on study results is the most interesting and rewarding part of the synthesis process. The next two chapters deal with techniques for analyzing whether and why there are differences in the outcomes of studies. As an analog to analysis of variance or multiple regression procedures, effect sizes can be viewed as dependent or criterion variables and the features of study designs as independent or predictor variables. However, because effect-size estimates do not all have the same sampling uncertainty, they cannot simply be inserted into traditional inferential statistics. In chapter 12 ("Statistically Analyzing Effect Sizes: Fixed- and Random-Effects Models"), Spyros Konstantopoulos and Larry Hedges discuss the difference between fixed- and random-effects models of effect-size homogeneity, the conceptual and statistical considerations involved in choosing an analytic model, and the statistical power of homogeneity tests. Chapter 13, by Larry Hedges, addresses recent advances in multivariate meta-analysis, in particular the use of meta-regression. This chapter also provides guidance to help reviewers avoid common mistakes when multivariate data are used in meta-analysis.

Part V delves into other approaches to the statistical combination of study results. In chapter 14, Rebecca Turner and Julian Higgins describe Bayesian meta-analysis, including Bayesian meta-regression and the

advantages and limitations of this approach. Effect-size estimates may be affected by factors that attenuate their magnitudes. These may include, for example, a lack of reliability in the measurement instruments or restrictions in the range of values in the subject sample. These attenuating biases may be estimated and corrected using the procedures Frank Schmidt, Huy Le, and In-Sue Oh describe in chapter 15. In chapter 16, Betsy Becker and Ariel Aloe introduce model-based meta-analysis and how to use this approach to investigate partial effects, indirect effects (including mediation), and to address questions that have not been explicitly addressed in any individual studies.

Part VI addresses two important complications that arise when working with meta-analytic data that all research synthesists must attend to. In chapter 17, Terri Pigott takes up handling missing data. She addresses different types of missing data (missing studies, effect sizes, study descriptors), provides an overview and critique of commonly used methods, discusses model-based methods for addressing missing data, and outcome reporting biases. In chapter 18, which takes up publication bias, Jack Vevea, Kathleen Coburn, and Alexander Sutton introduce methods to identify the presence, assess the impact, and adjust results for the synthesists who want to examine whether the published literature might be a biased sample all the studies that have been conducted.

1.4.5 Interpreting Synthesis Outcomes

Estimating and averaging effect sizes and searching for moderators of their variability is how the interpretation of cumulative study results begins. However, it must be followed by other procedures that help the synthesists properly interpret what they have discovered. Proper interpretation of the results of a research synthesis requires careful use of declarative statements regarding claims about the evidence, specification of what results warrant each claim, and any appropriate qualifications to claims that need to be made.

Part VII examines two important issues in data interpretation. In chapter 19 ("Interpreting Effect Sizes"), Jeffrey Valentine, Ariel Aloe, and Sandra Jo Wilson discuss methods for interpreting effect sizes in real-world terms. In chapter 20 ("Heterogeneity in Meta-Analysis"), Michael Borenstein introduces important considerations when thinking about differences between studies, including the distinction between observed and true effects,

statistics for assessing and describing heterogeneity, the null hypothesis of effect-size homogeneity, and common mistakes in thinking about heterogeneity.

1.4.6 Presenting Synthesis Results

Presenting the background, methods, results, and meaning of a research synthesis' findings are the final challenges to the synthesists' skill and intellect. These are addressed in the summary section, part VIII. In chapter 21, Evan Mayo-Wilson and Sean Grant describe the standards for reporting meta-analysis. As is true of the coding frame, no simple reporting scheme fits all syntheses. However, certain commonalities do exist. Not too surprisingly, the organization that emerges bears considerable resemblance to that of a primary research report although, also obviously, the content differs dramatically. In chapter 22 ("Threats to the Validity of Generalized Inferences from Research Syntheses"), Georg Matt and Thomas Cook provide an overall appraisal of how inferences from research syntheses may be restricted or faulty. This chapter brings together many of the concerns expressed throughout the book by the various chapter authors. Finally, chapter 23 ("Potentials and Limitations of Research Synthesis"), Harris Cooper, Larry Hedges, and Jeffrey Valentine pay special attention to possible future developments in synthesis methodology, the feasibility and expense associated with conducting a sound research synthesis, and a broad-based definition of what makes a literature review good or bad.

No secret will be revealed by stating our conclusion in advance. If procedures for the synthesis of research are held to standards of objectivity, systematicity, and rigor, then our knowledge edifice will be made of bricks and mortar. If not, it will be a house of cards.

1.5 REFERENCES

Arthur, Winfred, Jr., Winston Bennett Jr., and Allen I. Huffcutt. 2001. *Conducting Meta-Analysis Using SAS*. Mahwah, N.J.: Lawrence Erlbaum.

Bero, Lisa, and Drummond Rennie. 1995. "The Cochrane Collaboration: Preparing, Maintaining, and Disseminating Systematic Reviews of the Effects of Health Care." *Journal of the American Medical Association* 274(24): 1935–38.

Birge, Raymond T. 1932. "The Calculation of Error by the Method of Least Squares." *Physical Review* 40(2): 207–27.

Bohning, Dankmar, Ronny Kuhnert, and Sasivimol Rattanasiri. 2008. *Meta-Analysis of Binary Data Using Profile Likelihood*. Boca Raton, FL: Taylor and Francis.

Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Card, Noel A. 2012. *Applied Meta-Analysis for Social Science Research*. New York: Guilford.

Chalmers, Iain. 1993. "The Cochrane Collaboration: Preparing, Maintaining and Disseminating Systematic Reviews of the Effects of Health Care." *Annals of the New York Academy of Sciences* 703:156–63.

Chalmers, Iain, Larry V. Hedges, and Harris Cooper. 2002. "A Brief History of Research Synthesis." *Evaluation and the Health Professions* 25(1): 12–37.

Chen, Ding-Geng, and Karl E. Peace. 2013. *Applied Meta-Analysis with R*. Boca Raton, FL: Taylor and Francis.

Clarivate Analytics. 2018. "Web of Science Core Collection." Accessed May 7, 2018. <https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection>.

Cochran, William G. 1937. "Problems Arising in the Analysis of a Series of Similar Experiments." Supplement, *Journal of the Royal Statistical Society* 4: 102–118.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, N.J.: Lawrence Erlbaum.

Comprehensive Meta-Analysis. 2015. *Comprehensive Meta-Analysis*. Accessed November 20, 2018. <http://www.meta-analysis.com/index.php>.

Cooper, Harris M. 1982. "Scientific Guidelines for Conducting Integrative Research Reviews." *Review of Educational Research* 52(2): 291–302.

———. 1988. "Organizing Knowledge Synthesis: A Taxonomy of Literature Reviews." *Knowledge in Society* 1:104–26.

———. 2017. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*, 5th ed. Thousand Oaks, Calif.: Sage Publications.

Eddy, David M., Vic Hasselblad, and Ross Schachters. 1992. *Meta-Analysis by the Confidence Profile Approach*. Boston, Mass.: Academic Press.

Feldman, Kenneth A. 1971. "Using the Work of Others: Some Observations on Reviewing and Integrating." *Sociology of Education* 4:86–102.

Fisher, Ronald A. 1932. *Statistical Methods for Research Workers*, 4th ed. London: Oliver and Boyd.

Gelman, Andrew, and Hal Stern. 2006. "The Difference Between 'Significant' and 'Not Significant' Is Not Itself Statistically Significant." *American Statistician* 60(4): 328–31.

- Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis." *Educational Researcher* 5(1): 3–8.
- Glass, Gene V., Barry McGaw, and Mary L. Smith. 1981. *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications.
- Glass, Gene V., and Mary L. Smith. 1978. *Meta-Analysis of Research on the Relationship of Class Size and Achievement*. San Francisco: Far West Laboratory for Educational Research and Development.
- Hedges, Larry V. 1987. "How Hard Is Hard Science, How Soft Is Soft Science?" *American Psychologist* 42(2): 443–55.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Hunt, Morton. 1997. *How Science Takes Stock: The Story of Meta-Analysis*. New York: Russell Sage Foundation.
- Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson. 1982. *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, Calif.: Sage Publications.
- Jackson, Gregg B. 1980. "Methods for Integrative Reviews." *Review of Educational Research* 50(3): 438–60.
- Light, Richard J., and David B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Cambridge, Mass.: Harvard University Press.
- Light, Richard J., and Paul V. Smith. 1971. "Accumulating Evidence: Procedures for Resolving Contradictions Among Research Studies." *Harvard Educational Review* 41(4): 429–71.
- Lipsey, Mark W., and David B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, Calif.: Sage Publications.
- Manten, Arie A. 1973. "Scientific Literature Review." *Scholarly Publishing* 5:75–89.
- Mosteller, Frederick, and Robert R. Bush. 1954. "Selected Quantitative Techniques." In *Handbook of Social Psychology*, vol. 1. *Theory and Method*, edited by Gardner Lindzey. Cambridge, Mass.: Addison-Wesley.
- Olkin, Ingram. 1990. "History and Goals." In *The Future of Meta-Analysis*, edited by Kenneth W. Wachter and Miron L. Straf. New York: Russell Sage Foundation.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251): aac4716.
- Pashler, Harold, and Eric-Jan Wagenmakers. 2012. "Editor's Introduction to the Special Issue on Replicability in Psychological Science: A Crisis of Confidence?" *Perspectives on Psychological Science* 7(6): 528–30.
- Pearson, Karl. 1904. "Report on Certain Enteric Fever Inoculation Statistics." *British Medical Journal* 3:1243–46.
- Petticrew, Mark, and Helen Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden, Mass.: Blackwell.
- Price, Derek J. de Solla. 1965. "Networks of Scientific Papers." *Science* 149(3683): 510–15.
- Rosenthal, Robert. 1984. *Meta-Analytic Procedures for Social Research*. Beverly Hills, Calif.: Sage Publications.
- Rosenthal, Robert, and Donald B. Rubin. 1978. "Interpersonal Expectancy Effects: The First 345 Studies." *Behavioral and Brain Sciences* 3(6): 377–86.
- Schmidt, Frank L., and John E. Hunter. 1977. "Development of a General Solution to the Problem of Validity Generalization." *Journal of Applied Psychology* 62(5): 529–40.
- . 2015. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 3rd ed. Thousand Oaks, Calif.: Sage Publications.
- Shadish, William R. 2015. "Introduction to the Special Issue on the Origins of Modern Meta-Analysis." *Research Synthesis Methodology* 6(3): 219–20.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Mass.: Houghton Mifflin.
- Smith, Mary L., and Gene V. Glass. 1977. "Meta-Analysis of Psychotherapy Outcome Studies." *American Psychologist* 32(9): 752–60.
- Taveggia, Thomas C. 1974. "Resolving Research Controversy Through Empirical Cumulation: Toward Reliable Sociological Knowledge." *Sociological Methods & Research* 2(4): 395–407.
- Tippett, Leonard Henry Caleb. 1931. *The Methods of Statistics*. London: Williams and Norgate.
- Valentine, Jeffrey C., Anthony Biglan, Robert F. Boruch, Filipe G. Castro, Linda M. Collins, Brian R. Flay, Sheppard Kellam, Eve K. Moscicki, and Steven P. Schinke. 2011. "Replication in Prevention Science." *Prevention Science* 12(2): 103–17.
- Valentine, Jeffrey C., Theresa D. Pigott, and Hannah R. Rothstein. 2010. "How Many Studies Do You Need? A Primer on Statistical Power for Meta-Analysis." *Journal of Educational and Behavioral Statistics* 2(2): 215–47.
- Yates, Frank, and William G. Cochran. 1938. "The Analysis of Groups of Experiments." *Journal of Agricultural Science* 28(4): 556–80.