

# Chapter 19

## Meta-analysis in Prevention Science

Sandra Jo Wilson and Emily E. Tanner-Smith

### Introduction

This chapter describes the fundamental elements of meta-analysis, with particular emphasis on its relevance to prevention science. The goal is to provide readers with a basic understanding of what a meta-analysis is, how to identify meta-analysis topics appropriate to prevention science, how to interpret results from meta-analysis, and how to identify some of the potential biases in meta-analysis; in short, our goal is to create intelligent consumers of meta-analysis. Armed with knowledge about some of the common ways that meta-analytic techniques can be used in prevention science research, we encourage readers interested in conducting a meta-analysis to seek more comprehensive resources on the statistical methods unique to this form of research (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009; Lipsey & Wilson, 2001).

### What Is Meta-analysis?

The term “meta-analysis,” coined by Glass (1976), encompasses a range of techniques for quantitatively collecting data from a preselected set of primary research studies and applying specialized statistical analyses that synthesize findings across studies. In contrast to primary studies that may use individuals, families, classrooms, or schools as the unit of analysis, it is the primary studies themselves that are the unit of analysis in a meta-analysis. Because different primary studies often use different measures to represent the same underlying constructs, meta-analysts are tasked with standardizing findings across studies to make results comparable.

---

S.J. Wilson (✉)  
Peabody Research Institute, Vanderbilt University, Peabody #181, 230 Appleton Place,  
Nashville, TN 37203, USA  
e-mail: [sandra.j.wilson@vanderbilt.edu](mailto:sandra.j.wilson@vanderbilt.edu)

This is achieved by calculating *effect sizes* from the data reported in each primary study. An effect size is a standardized, quantitative index representing the magnitude and direction of a relationship. By representing the findings of each study included in a meta-analysis in the same form, the effect size permits a synthesis of those findings across studies. There are a wide variety of effect size indices that can be meta-analyzed, including those representing measures of central tendency (means, proportions), pretest–posttest contrasts, group contrasts (mean differences), and associations between two variables (correlations, odds ratios); the effect size index chosen for a given meta-analysis will depend on both the goals of the meta-analysis and the types of statistical information commonly reported in the primary research literature of interest.

Meta-analysis is closely related to, and often overlaps with, another form of research synthesis called a systematic review. A systematic review seeks to identify, collate, and systematically summarize all empirical evidence on a specific research topic, using explicit, systematic, transparent, replicable methods that are designed to minimize bias. This is in contrast to traditional narrative reviews, which often lack explicit eligibility criteria and transparency and tend to summarize the results of included studies in a subjective manner. Like systematic reviews, meta-analyses also seek to identify, collate, and summarize empirical evidence, but they use statistical methods to produce a quantitative summary of the results of a set of studies. While many systematic reviews use meta-analysis as the method of statistically synthesizing the included studies' findings, not all systematic reviews will include a meta-analysis (e.g., they may only provide a narrative summary of the studies). Meta-analyses are not always based on systematic reviews and may instead be based on unsystematic or non-exhaustive searches of the literature. Most meta-analyses, however, endeavor to be exhaustive in their search so that results may generalize to a broader population of studies.

## Rationale for Meta-analysis

Several features of meta-analysis make it particularly applicable to prevention science. First, by coding detailed information about the characteristics of the included studies into a database that can be analyzed statistically, meta-analysis provides an organized, systematic, and comprehensive approach to representing the nature and findings of numerous complex and diverse individual studies (Cooper, 1984). Second, the statistical aggregation across multiple respondent samples involved in meta-analysis yields more statistically reliable estimates than individual studies and, in particular, inhibits misinterpretation of sampling error as real differences among studies (Hunter, Schmidt, & Jackson, 1982; Schmidt, 1992). However, where there are real differences in findings among studies, meta-analysis permits statistically sophisticated analysis of the source of those differences. Statistical models tailored for meta-analytic data can reveal common findings across studies despite differences in study method and procedure that may obscure the real

relationships of interest (Cooper et al., 2009; Hunter & Schmidt, 1990). Finally, by dealing with the full range of samples, variables, and relationships in a body of research, meta-analysis can potentially present a synthesis of empirical findings on a topic that has more scope, depth, and generality than any one primary study can provide. Indeed, single studies sometimes have an enormous impact on public policy, without adequate consideration given to the possibility of sampling error as well as limitations of generalizability. Meta-analysis provides a set of techniques for summarizing a body of research in ways that can provide comprehensive policy-relevant information that is more reliable and defensible than results from any primary study alone.

Of course, meta-analysis requires a body of empirical research to be on hand for synthesis, which means that meta-analysis can be applied only when there is sufficient primary research to review. It also requires that the research to be synthesized be of sufficient similarity to be compared and that the research yield quantitative results that can be represented as effect sizes. Thus, while meta-analysis can provide an additional tool in the arsenal of research methods applicable to prevention scientists, it cannot stand alone and requires primary research for its existence. We now turn to a discussion of the basic phases or steps in a meta-analysis. As with any type of research, there are potential threats to the validity of the inferences that can be drawn from meta-analysis. As we outline the steps in a meta-analysis, we will weave in a discussion of some of the more common sources of bias that represent threats to validity in meta-analysis (for more detailed discussions see Cooper, 2010; Matt & Cook, 2009). Bias refers to a systematic error in findings, in contrast to the imprecision associated with sampling error inherent in all inferential statistics. It is important to note that traditional narrative reviews are subject to many of the same biases as meta-analysis; in fact, narrative reviews are generally at greater risk for bias because of the subjective way in which studies are included and interpreted. Fortunately, meta-analysts can minimize most of the biases we discuss below through careful planning at all stages of the meta-analysis. Meta-analysts can also empirically assess the possibility of some sources of bias, something not possible with narrative reviews.

## **Basic Elements of a Meta-analysis**

### ***Developing a Research Question***

While meta-analyses can differ widely in scope and purpose, all generally involve the same basic steps, the first of which is developing a research question. Three broad types of research questions are particularly appropriate for meta-analysis in the field of prevention science: (1) questions about the etiology or epidemiology of particular social problems; (2) questions about the efficacy or effectiveness of interventions for solving those social problems; and (3) questions about group

differences, either between naturally occurring groups (i.e., males and females) or between groups defined by researchers (e.g., between different diagnostic groups).

One of the pillars of prevention science research involves understanding the etiology and epidemiology of social problems, and one form of meta-analysis is well-suited to such questions of etiology—meta-analyses of correlational relationships between risk and protective factors and concurrent or later problems. Such meta-analyses can form the foundation for a risk-reduction/protection-enhancement approach to prevention (Battin-Pearson et al., 2000; Kraemer et al., 1997; Mrazek & Haggerty, 1994) by seeking to identify the significant predictors of the problem of interest. For example, one recent meta-analysis synthesized longitudinal correlation effect sizes from 41 studies to examine the predictive strength of risk factors for later delinquent and criminal behavior among children and adolescents, focusing on the differential strength of risk factors across developmental life stages and domains of risk (Tanner-Smith, Wilson, & Lipsey, 2011). One of the conclusions from the study was that family risk factors (e.g., harsh parenting, low family cohesion) occurring during childhood were strong risk factors for later adolescent crime and delinquency. Meta-analyses such as these are thus particularly useful for prevention scientists and can support two central elements of prevention programming: (1) identification and selection of individuals or groups of sufficiently high risk to be appropriate for services and/or (2) development of service programs to ameliorate targeted risk factors and/or enhance selected protective factors with the expectation that this will prevent, or at least mitigate, problematic outcomes or improve positive ones (Catalano, Hawkins, Berglund, Pollard, & Arthur, 2002; Kellam, Koretz, & Mościcki, 1999).

The second category of research questions particularly relevant to prevention scientists is questions about intervention effectiveness, and meta-analyses focusing on intervention effectiveness are quite common in the field. The Campbell Collaboration (<http://www.campbellcollaboration.org/>) publishes systematic reviews and meta-analyses of intervention effects, many of which focus on prevention and intervention programs. For example, one such meta-analysis synthesized results from 73 studies to examine the effects of universal school-based social information-processing interventions on school-age children's aggressive and disruptive behavior; the authors found an overall beneficial effect of the interventions, although the magnitude of effect was somewhat small and varied across different levels of treatment dosage (Wilson & Lipsey, 2006). Another Campbell Collaboration review synthesized results from 23 studies to examine the effects of exercise on children and youth's self-esteem, with results indicating overall beneficial effects on self-esteem outcomes (Ekeland, Heian, Hagen, Abbott, & Nordheim, 2005). Intervention meta-analyses can also be important for identifying harmful intervention approaches, as in the case of a meta-analysis of nine studies that examined whether Scared Straight juvenile awareness programs were effective for preventing juvenile delinquency; in that meta-analysis, the results indicated that participation in juvenile awareness programs was associated with higher levels of juvenile delinquency (Petrosino, Petrosino, & Buehler, 2004). Meta-analyses focusing on questions of intervention effectiveness are clearly applicable to prevention science

because they can identify effective program strategies, reveal ineffective or harmful strategies, and, as we will describe in more detail below, examine the conditions under which programs may be more or less effective.

The third major class of research questions appropriate for meta-analysis and of interest to prevention scientists has to do with group differences. Understanding the nature of differences between males and females on math achievement or studying the achievement gap for minority (or low socioeconomic status) students versus majority students is critical for understanding the nature of social problems and can inform the design of interventions that might be appropriate for different subgroups of individuals. For instance, one recent meta-analysis (Lindberg, Hyde, Linn, & Petersen, 2010) synthesized results from 242 studies that examined gender differences in mathematics achievement. The authors concluded that there was not a significant gender gap in mathematics achievement among school-age youth, although there was some evidence of a gender gap in math achievement among high school and college samples (vs. elementary and middle schools). Meta-analyses addressing research questions of group differences are particularly relevant for understanding whether certain groups of individuals may or may not need targeted prevention programs. Another application of group differences meta-analysis involves comparing groups of individuals created by the researchers, such as studies comparing outcomes or symptomatology for attention deficit hyperactive disorder patients to normal individuals (e.g., Bálint et al., 2009).

### **Potential Bias at the Research Question Stage**

When developing the research questions of interest for a meta-analysis, an important issue to consider is whether the types of primary studies to be included are similar enough to be synthesized in a single meta-analysis. Meta-analyses that include diverse types of primary studies in a single analysis have been criticized for mixing “apples and oranges,” which critics surmise to yield meaningless results. Robert Rosenthal once responded to this critique by stating that in some cases, a meta-analyst may be interested in neither apples nor oranges, but fruit salad (as cited in Borenstein et al., 2009). In the end, decisions about the breadth of studies to include in a meta-analysis depend on (and will vary according to) the stated goals of the project.

For instance, a researcher may be interested in evaluating the effectiveness of the Olweus Bullying Prevention Program and, therefore, decides to synthesize results from all randomized controlled studies that compared intervention programs using the Olweus Bullying Prevention Program manual with some control condition. To draw conclusions about this specific program’s overall effectiveness, it would not make sense for this meta-analysis to also include studies of other bullying prevention programs that did not follow the Olweus manual. In contrast, if a researcher was interested in the comparative effectiveness of different types of bullying prevention programs (i.e., interested in fruit salad), they could justifiably include studies using a wide variety of bullying prevention programs (e.g., Lions Quest,

Olweus, Ripple Effects, non-manual-based bullying programs). The authors of such a meta-analysis might then empirically examine the comparative effectiveness of those programs through moderator analysis, as discussed later in this chapter. As such, it is always important to consider the breadth of studies included in a meta-analysis and whether the study results can be synthesized in a way that allows inference to a meaningful population given the research question(s) of interest.

### ***Defining Eligibility Criteria***

The next stage of a meta-analysis is to define eligibility criteria, which follow directly from the research questions. Eligibility criteria should explicitly define the types of studies that are eligible for inclusion and give the reader a clear idea of the nature of the literature being reviewed. The specifics of the eligibility criteria will vary depending on the research question, but they will generally include four primary components and several secondary components.

#### **The Topic**

Most importantly, the eligibility criteria must identify the distinguishing features of the research to be included. If the research question is one of etiology and, for example, involves studying the predictors of later antisocial behavior, the criteria should define the predictors and outcomes of interest, how they are put into operation, and the types of relationships between those variables that are eligible. If the research question focuses on the effects of an intervention, the eligibility criteria must define the critical features of that intervention so that the intervention of interest can be distinguished from other types of interventions (which may or may not be similar). Research questions on group differences necessitate eligibility criteria that demarcate the boundaries for the groups of interest and the comparisons between those groups that are relevant.

#### **The Population**

The types of research participants that are relevant must also be specified in the eligibility criteria. Most commonly, criteria with regard to the population of interest specify the gender, ethnic, socioeconomic, and age groups that are included (or excluded). When age or socioeconomic criteria are specified, the boundaries must be specific and clearly defined. For example, if middle school children are the pertinent population, the eligibility criteria should specify what age and/or grade levels constitute middle school students and how cases where middle school students are mixed in with elementary or high school students are to be handled. In addition, whether specialized populations are eligible or not should be clearly

specified. For prevention science researchers focused on educational research, for example, the eligibility criteria should clearly specify whether special education, learning disabled, or other special populations are considered eligible for the review.

### **Pertinent Variables**

The variables of interest to the researcher must also be clearly identified. For example, if the meta-analysis is focused on intervention effects, the specific outcomes of the intervention that are eligible should be specified. Criteria here would also include limitations on the timing of measurement of the outcomes, as well as any restrictions with regard to the source and nature of the measurement. For questions about etiology, specifying the eligible pertinent variables can overlap with eligibility criteria on the topic of the meta-analysis, but further clarifying information about the measurement and timing characteristics can be provided when specifying the pertinent variables. In addition, criteria with regard to the statistical findings of the primary studies would be relevant here. Primary studies need to provide sufficient quantitative information to compute effect sizes. In addition, some meta-analysts may be interested only in certain forms of data, such as continuous or binary outcomes.

### **Research Methods**

Finally, the eligible research methods must be specified. When studying questions about intervention effects, the types of experimental or quasi-experimental designs that are relevant should be clearly specified and excluded designs enumerated. For some types of interventions, pretest–posttest designs, time series, or single-subject designs might also be relevant for any given analysis. The research methodologies corresponding to a focus on etiology may use longitudinal or cross-sectional designs or both and report on the statistical associations among variables. Research methodologies for questions of group differences would include details about how contrived groups are constituted or the types of comparisons made between groups that are considered pertinent.

### **Secondary Criteria**

Other secondary eligibility criteria may involve the cultural and linguistic range of the studies included, the timeframe of the literature, and the publication types that are to be included or excluded.

## The PICOS Framework

We have described eligibility criteria for meta-analysis somewhat generically thus far, in a way that is applicable to all forms of meta-analysis. There is, however, a useful framework for developing eligibility criteria for intervention meta-analyses that deserves mention. This framework includes five primary components and is described with the acronym PICOS (population, intervention, comparison, outcome, study design; Higgins & Green, 2011). Under the PICOS framework, the eligibility criteria should specify the types of research participants in the primary studies of interest (population); the critical features of the intervention under study, as well as its dose, format, frequency, duration, timing, and so forth (intervention); the types of comparison conditions that are eligible, whether no treatment, treatment as usual, placebo, or some other type of intervention (comparison); the outcome constructs of interest, including the timing of measurement, operationalization, and source (outcome); and, finally, the types of study designs eligible for inclusion such as randomized, nonrandomized, pretest–posttest only, and so forth (study design).

## Potential Bias at the Eligibility Criteria Stage

At the eligibility criteria stage, the meta-analyst is tasked with creating clearly defined criteria that identify the characteristics of the studies eligible for inclusion in the meta-analysis. At this stage of a meta-analysis, bias may occur due to any ambiguity in the operational definitions of the key constructs of interest specified in each of the eligibility criteria (e.g., population, intervention, outcomes, research designs). Consider a researcher proposing to conduct a meta-analysis on the effects of cognitive behavioral therapy (CBT) interventions on middle school students' tobacco use. The researcher could naively rely on study authors' descriptions of the interventions and make eligible only those studies in which programs were explicitly labeled as CBT. If the researcher is truly interested in the effectiveness of programs that use the components of CBT, reliance on a program label could result in the inclusion of studies reporting implementing "CBT" when they did not, in fact, include any elements of CBT. Moreover, such reliance on a program label could also result in the exclusion of studies that actually did use cognitive behavioral principles but were not specifically billed as CBT (like a social problem-solving intervention). In this case, the population of intervention types to which results could be generalized would be somewhat ambiguous and may not provide an adequate answer to the research question.

Another type of construct ambiguity common in intervention meta-analyses is lack of specificity about the nature of the comparison groups, especially those labeled as "practice as usual." Without a clear definition of the boundaries of "practice as usual," it is difficult to assess the type and level of services received by comparison group participants, making interpretation of the resulting effect sizes

ambiguous. Indeed, it is difficult to make statements about effective interventions when the type of service received by the comparator groups is ambiguous or undefined. Similar threats to validity can occur with all relevant study constructs (e.g., populations, pertinent variables, study designs). Regardless of the goals of meta-analysis, conceptual ambiguity at the research question and eligibility criteria stages can affect later inferences and always deserves careful consideration.

## *Literature Search*

Using the eligibility criteria as a guide, meta-analysis next involves conducting and documenting a systematic search for all studies that meet the eligibility criteria. Ideally, the search should involve transparent, diverse, and iterative procedures to locate the population of studies relevant to the meta-analysis. Excellent resources are available on conducting literature searches, but the distinguishing feature of most searches for meta-analysis (and even more so for exhaustive systematic reviews) is that they involve multiple sources. The sources commonly include electronic citation databases that house published reports (e.g., ERIC, MEDLINE, PsycINFO), but also sources for unpublished (or “gray”) literature (Rothstein & Hopewell, 2008). Internet searches, hand searches of key journals, contact with experts in the field, and reference harvesting from previous meta-analyses, systematic reviews, and narrative literature reviews are commonly used to identify gray literature that may not be indexed in standard electronic bibliographic databases.

## **Potential Bias at the Literature Search Stage**

When conducting a meta-analysis, well-executed literature searches are those that are broad in scope, diverse in sources, and transparent in methods with the ultimate goal of minimizing the potential for publication bias. Publication bias is one of many types of reporting or dissemination biases that have the potential to influence the validity of generalizations from meta-analysis. It is due primarily to the fact that studies with large effects are more likely to be published than those with small or null effects (Rothstein, Sutton, & Borenstein, 2005; Sterne, Egger, & Moher, 2008). This has often been referred to as the “filedrawer” problem—that “journals are filled with the 5 % of studies that show Type I errors, while the file drawers back at the lab are filled with the 95 % of the studies that show non-significant (e.g.,  $p < 0.05$ ) results” (Rosenthal, 1979). There is, indeed, a large body of empirical literature documenting that primary studies with statistically significant and/or “positive” effects (i.e., those in the hypothesized direction) are more likely to be submitted for publication (Cooper, DeNeve, & Charlton, 1997; Dickersin, 1997). And in intervention meta-analyses, published studies do tend to yield larger treatment effect estimates compared with unpublished studies (Lipsey & Wilson, 1993).

When a meta-analysis includes only published research, there is the potential that overall effects may be inflated as a result.

There are a range of other reporting biases, all of which increase the likelihood of overestimating true study results. For instance, time-lag bias results from studies with large effects being published faster than those with small effects (Hopewell & Clarke, 2001). Multiple publication bias results when studies with large effects are published in multiple reports and are, therefore, more easily identified in searches (Reyes, Panza, Martin, & Bloch, 2011). Location bias occurs when large effect sizes are published in more easily accessible locations (Pittler, Abbot, Harkness, & Ernst, 2000). Citation bias is the result of studies with large effects being cited more often in other publications, making them more easily identifiable through reference harvesting (Gotzsche, 1987). Language bias is the result of studies with large effects being more likely to be published in English (Egger et al., 1997). Finally, outcome reporting bias occurs when primary study authors selectively report results for only those outcomes that show large or significant effects (Tannock, 1996).

Any type of reporting bias can distort the results of a meta-analysis. There are several instances in the field of medicine where meta-analysts would have reached different conclusions about the effectiveness of treatments for ovarian cancer, heart disease, or thyroid disease depending on whether results from unpublished studies were included in the meta-analyses (Chalmers, 2001; Rennie, 1997; Simes, 1986). To minimize the possibility of publication bias, most meta-analysts, therefore, attempt to identify this difficult-to-locate “gray literature” by using diverse search strategies other than standard electronic bibliographic databases such as PsycINFO or MEDLINE. Meta-analysts should also assess the possibility of publication bias through the use of exploratory statistical procedures (see Rothstein et al., 2005). Although no statistical tests can definitively answer “Does this meta-analysis suffer from publication bias?”, it is important for meta-analysts to use the exploratory tools currently available to at least acknowledge whether findings are at risk of such bias. Furthermore, consumers of meta-analysis should pay attention to how literature search procedures and publication bias analyses are reported.

### ***Data Extraction or Study Coding***

Once the set of eligible studies is identified and obtained, a meta-analysis uses objective and systematic coding procedures to extract information from the eligible studies. Meta-analysis must attend to variation in results across studies and attempt to distinguish variation attributable to systematic differences among studies from variation attributable to sampling error and other unsystematic sources. Furthermore, it is also important to distinguish whether systematic differences across studies are due to substantive, methodological, or procedural differences in those studies. Therefore, in addition to extracting the effect sizes that index each study’s findings, other types of information are also relevant. The types of information extracted from studies in a meta-analysis vary across the different types of meta-analysis,

but generally include the following information: (1) study identification, (2) study methodology, (3) research participants, (4) effect sizes and dependent variables, and, for intervention meta-analyses, (5) the characteristics of the interventions (Wilson, 2009).

### **Study Identification**

Study identification involves recording the identifying characteristics of the studies in the meta-analysis. This information typically includes the title, date, and author(s) of the study; the source of the study (e.g., journal article, conference paper, doctoral dissertation, technical report, book chapter); retrieval source (e.g., computer search, reference list); and study setting or region.

### **Study Methodology**

While methodological characteristics of studies may not be of substantive interest to practitioners or policy makers, they are a critical part of coding because the methods used to conduct research studies can influence their outcomes. For example, studies in which subjects are randomly assigned to treatment and comparison groups tend to have different effects than studies in which subjects are not randomly assigned (Lipsey & Wilson, 2001). Coding details of the study methods allows the meta-analyst to identify influential method characteristics and control for them in any statistical analysis. To illustrate, imagine a meta-analysis of therapeutic interventions for anxiety where, in the primary studies, cognitive restructuring programs tended to use more randomized experiments, while psychoanalytic programs tended to be evaluated more often with quasi-experimental designs. If the average effect size for the cognitive restructuring programs turns out to be larger than the average effect size for the psychoanalytic treatments, we cannot be sure whether cognitive restructuring programs are more effective than psychoanalysis or whether randomized experiments might tend to result in larger program effects regardless of the type of treatment. During analysis of results, having information about research design and other methodological characteristics can be immensely useful for describing the overall methodological quality of the candidate studies, separating the results for high- and low-quality studies, and using statistical methods to control for confounding methodological differences between studies.

Many different aspects of method and procedure can be coded in a meta-analysis or systematic review. Some meta-analyses may refer to methodological quality checklists that are available (Chalmers et al., 1981; Valentine, 2009); others may frame the methodological coding around the validity of the findings or the risk of bias present in the candidate studies (Higgins & Altman, 2008). Still others may perform an objective coding of the methodological characteristics of interest and examine their influence on the effect sizes later. Some common methodological characteristics included in the various meta-analysis coding schemes include

method of assignment to research conditions (e.g., random assignment, cluster randomization, matching), nature of the comparison group (e.g., received no treatment, placebo, alternate treatment), study attrition, study blinding, pretreatment equivalence of groups, outcome measures used to assess treatment effects (e.g., norm referenced, criterion referenced, rating scale), timing of follow-up outcome measures, and reliability of outcome measures.

### **Research Participants**

Coding candidate studies for participant characteristics allows the meta-analyst to determine whether included studies have similar target populations and subsequently examine whether study findings are associated with those participant characteristics. This information is important for assessing the appropriateness of different interventions for use with different types of participants and for understanding how the etiology of social problems might differ across participant subgroups. A great variety of participant characteristics may be coded, depending on the research questions and the types of participant information commonly reported in the literature being reviewed. In general, basic demographic information including gender, race/ethnicity, age or grade, and socioeconomic status is important. Risk status, diagnostic status, severity of problem behavior, education level, previous health and mental health histories, and a range of other personal characteristics might also be relevant. The information coded about research participants should be informed by the research questions and theory or empirical research that identifies participant characteristics that may be associated with the outcomes of interest.

### **Intervention Characteristics**

For meta-analyses of intervention effects, coding characteristics of the intervention programs is of primary importance, both for understanding the quality of the treatment's implementation and for determining the overall effectiveness of the treatment program(s). Examples of treatment program characteristics used to examine variation in intervention effects include length of treatment program, number of treatment sessions in program, length of each treatment session, type of treatment program, individual or group treatment sessions, fidelity of treatment (i.e., treatment implemented as described), and type and training of treatment administration personnel.

### **Effect Sizes**

Finally, perhaps the most critical pieces of information extracted during the coding phase of a meta-analysis are those related to the effect sizes (i.e., the actual statistical results of the study). The effect sizes most commonly used in prevention

science fall into three major families: those that index differences between groups on continuous measures (e.g., the standardized mean difference or Cohen's  $d$ ), those that index relationships between two continuous measures (e.g., the correlation coefficient), and those that index differences between groups on frequency or incidence (e.g., odds ratio or risk ratio).

In any coding scheme, each effect size that is coded will be accompanied by a set of codes that provide detailed information about its distinctive source and nature. Such information might include the data used in the computation, sample size on which the effect size is based, amount of attrition in the sample, manipulations used (e.g., to derive it from other statistics), and other such items that can be used in statistical analysis to examine methodological or procedural matters that may systematically influence effect size.

### **Dependent Variables**

The coded information for each effect size also typically includes information about the variables involved in the index. For instance, this coding could identify the construct represented in the index, the nature of the behavior at issue, source of the information, and relevant features of the operationalization. It is this coding that forms the basis for grouping the effect sizes that are combined in a given analysis, that is, those that are treated as representing the "same" construct for that particular analysis.

### **Potential Bias at the Coding Stage**

As with the eligibility criteria, construct ambiguity can introduce bias at the coding stage of a meta-analysis. The validity of conclusions from a meta-analysis can also be threatened by issues of reliability that arise during the coding of eligible study reports. When multiple coders extract the same type of information from multiple studies, there is always the risk of unreliability. Therefore, it is important for meta-analysts to conduct extensive training sessions with coders and to assess the reliability of coding. Ideally, all study reports would be coded by two independent coders, and all discrepancies would be resolved through discussion and further training. In practice, however, it may not be feasible (due to budgetary reasons or otherwise) to have all studies double coded; in this case, it is extremely important for inter-coder agreement to be established early on during the project and continuously monitored with subsets of studies. Validity of conclusions can be threatened when multiple coders are not reliable with one another and when individual coders are not reliable with themselves (what has been called "coder drift"). Coder drift can be the result of coder fatigue or a change in understanding of constructs over time. Therefore, it is important to conduct, at minimum, continuous monitoring of subsets of coded studies to quickly assess and remedy any reliability problems with the coding.

## ***Data Analysis and Interpretation***

As with any form of research, the final stage of the project comes when the data are analyzed and interpreted and conclusions are drawn about the body of research under study. Analysis of meta-analytic data has several statistical quirks but proceeds much like analysis of data from primary studies. Several good texts are available, and a variety of software packages and macros have been developed to handle the specific issues associated with analyzing effect sizes (Borenstein et al., 2009; Borenstein, Hedges, Higgins, & Rothstein, 2010; Lipsey & Wilson, 2001; Sterne, 2009), so we won't spend time here reviewing the analytic techniques in detail. Rather, the remainder of this section provides a basic overview of meta-analysis methods and some applications to prevention science.

In addition to summarizing the basic study characteristics of the literature reviewed, a typical prevention science meta-analysis would include the following components: (1) the average effect size and effect size distribution for each outcome of interest and an examination of the heterogeneity in the effect size distributions, (2) subgroup or moderator analysis in which the variability present in the effect size distribution is systematically analyzed to identify study characteristics that are associated with larger or smaller effect sizes, and (3) publication bias analysis and other sensitivity analyses to assess the validity of conclusions drawn. We briefly review each of these in turn.

### **Average Effect Sizes and Heterogeneity**

Most meta-analyses will present an average effect size value synthesized from the individual effect sizes extracted from the primary studies included in the review. When calculating the average effect size, each effect size is typically weighted by the inverse of its sampling variance, so that effect sizes measured with greater precision are given greater weight because they provide better estimates of the underlying population parameter(s) of interest. Meta-analysts will typically provide estimates of average effect sizes and their distribution for each outcome of interest to make broad statements about the average effect in the population.

For instance, in their meta-analysis examining the effects of social information-processing programs on students' aggressive and disruptive behavior, Wilson and Lipsey (2006) reported an overall average standardized mean difference effect size of 0.21 based on 73 studies and concluded that the intervention had a small but statistically significant effect on reducing youths' aggressive and disruptive behavior. By calculating the average effect size across all included studies, meta-analysts can thus answer the simple question of "did the interventions work?" (or for other types of research questions, "is there an association between two variables?" or "are there differences between these groups?").

In many cases, the meta-analyst may be interested, not only in the average effect, but also in the variability of those effects across different types of studies,

participant samples, and so forth. Therefore, most meta-analyses also present statistics that summarize the amount of variability between studies, test whether any observed heterogeneity may be due to chance, and summarize the proportion of observed heterogeneity that can be considered true heterogeneity rather than statistical noise (i.e., estimates of  $\tau^2$ ,  $Q$  or  $\chi^2$ , and  $I^2$ , respectively; see Borenstein et al., 2009). When heterogeneity statistics indicate that substantial heterogeneity is present, some meta-analysts may decide that the studies are, in fact, too heterogeneous to calculate an average effect or do any meaningful statistical synthesis. Although this decision may be justified in some situations, many prevention scientists may actually be interested in this heterogeneity and, therefore, choose to empirically examine it through the use of subgroup or moderator analysis. The Wilson and Lipsey (2006) review, for instance, found substantial heterogeneity in their effect size distribution and, thus, proceeded to statistically examine a variety of factors that may have contributed to that heterogeneity.

### Moderator/Subgroup Analysis

In meta-analysis, moderator analysis refers to statistical analyses that examine whether the coded study characteristics for each study are associated with the effect sizes from those studies, that is, whether coded variables can *explain* some or all of the observed heterogeneity in the effect sizes (see Lipsey & Wilson, 2001, for more technical detail). This type of analysis is called “moderator” analysis in that it examines whether a certain coded variable or variables ( $x$ ) are associated with the direction or magnitude of the effect size ( $y$ ), when the effect size is defined as an index of the association between two variables (i.e., the association between a treatment variable and an outcome variable in an intervention meta-analysis, the association between a predictor and an outcome in an epidemiological meta-analysis, or the association between group membership and an outcome variable in a group differences meta-analysis). Thus, the covariate  $x$  is framed as a moderator of the relationship between the two variables encapsulated in the effect size  $y$ . Moderator analysis is conducted using analogs to ANOVA and linear regression that are modified for use with meta-analytic data. The choice between the ANOVA and regression frameworks depends on the measurement level of the covariate(s) of interest. Typically, “subgroup” analysis refers to a moderator analysis of categorical covariates in the ANOVA analog framework.

Moderator analysis thus allows meta-analysts to examine myriad factors that may be associated with the study findings. Questions that may be particularly relevant to prevention scientists include examining the conditions under which and for whom certain interventions may be most effective, certain risk factors may be most important, or differences between groups may be largest. For instance, the Wilson and Lipsey (2006) meta-analysis found that studies with primarily low socioeconomic status participants, those with more frequent treatment contact with participants, those in which interventions were delivered for research and demonstration purposes, and those with no obvious implementation difficulties produced

the largest intervention effects. Opportunities for moderator analysis in a meta-analysis are limited only by the size of the literature under review and the characteristics of the study variables coded for each study. That said, however, it is important for meta-analysts to identify moderators of interest a priori—not only so that the moderators can be coded during the data collection phase, but also to minimize any data dredging at the analysis phase that might capitalize on chance.

Returning to the three general types of research questions outlined above, meta-analyses that focus on questions of etiology, epidemiology, and the development of social problems present several interesting analysis opportunities. First, these correlational meta-analyses produce a quantitative summary of the strength of relationships among the variables of interest. That information can be fed into analyses that examine the differential predictive strength of different risk or protective factors or variation in the predictive strength of risk or protective factors for individuals or groups with different characteristics. This information can be used to identify both the target behaviors for intervention as well as the best individuals to target for intervention services. In a similar fashion, group differences meta-analyses lend themselves to questions about diagnostic groups that might be particularly amenable to treatment or at particular risk for later problems or comorbidities; they can also identify intervention targets that may differ across different subgroups of the population. Finally, intervention meta-analyses provide a variety of analysis opportunities for producing policy-relevant results. Questions of what works best under what conditions, in what types of settings, and for what types of individuals are important here and are most defensible when meta-analysts do a careful assessment of the risk of bias of the included studies and are also mindful of the influence that study methods can have on research findings.

### **Publication Bias and Sensitivity Analysis**

As previously mentioned, researchers must acknowledge the possibility of publication bias and how it may affect the results of a meta-analysis. There are several exploratory statistical procedures that meta-analysts can use to examine the possibility of publication bias (see Rothstein et al., 2005). The most commonly reported procedures include visual inspections of funnel plots or regression-based tests for effect sizes based on continuous outcome data (Egger, Davey Smith, Schneider, & Minder, 1997) or dichotomous outcome data (Harbord, Egger, & Sterne, 2006; Peters, Sutton, Jones, Abrams, & Rushton, 2006; Rucker, Schwarzer, & Carpenter, 2008). Other methods that have been used to assess publication bias in the past, such as the rank correlation test and variations on the fail-safe  $N$ , are no longer recommended for use given their known limitations (Becker, 2005; Sterne, Gavaghan, & Egger, 2000). Statistical development of publication bias analysis methods is constantly evolving, however, and there are currently no consistently agreed-on standards in the field. This, along with the known limitations of currently used methods (e.g., low power), means that there is no simple solution for detecting publication bias in a meta-analysis. Nonetheless, meta-analysts must be sensitive to

the possibility of publication bias and its potential effect on the conclusions that can be drawn from the results.

In addition to conducting exploratory analysis to assess the possibility of publication bias in a meta-analysis, it is also common practice for meta-analysts to conduct sensitivity analyses. As with any analysis of data in a primary study, the meta-analyst makes decisions during the data collection and analysis phases that could conceivably influence study results and conclusions. Therefore, it is important to conduct sensitivity analyses that explore whether those decisions had an appreciable impact on the meta-analysis findings. For instance, sensitivity analyses may explore the impact of (1) only including randomized studies of intervention effectiveness, (2) only including studies published in English, (3) the use of any statistical adjustments (4) how outlier cases were handled, (5) potential confounding among moderators, (6) how missing data were handled, and so on.

### Potential Biases at the Data Analysis Stage

Several threats to validity can occur at the analysis stage of a meta-analysis. A comprehensive understanding of these threats presumes a working knowledge of meta-analytic methods and standards, of course, but we will review some of the more common issues here (we again refer interested readers to comprehensive books on meta-analysis methods for a further understanding of the issues at hand, e.g., Borenstein et al., 2009; Cooper, 2010; Cooper et al., 2009; Lipsey & Wilson, 2001; Littell, Corcoran, & Pillai, 2008).

#### Effect Size Approximations

The statistical assumptions and methods employed in a meta-analysis can influence the validity of the conclusions. First, there is the issue of computing effect size estimates—the currency and primary outcome in a meta-analysis. Most meta-analysts can attest to the frustrating reality of discovering that the information needed to calculate an effect size statistic or its corresponding standard error (e.g., sample sizes, standard deviations) has been omitted from a primary study. Meta-analysts should be transparent about how they deal with such missing data and the extent to which any effect size estimates were approximated from partially reported data. For instance, one common effect size index, the standardized mean difference (or Cohen's  $d$ ), is easily computed when study authors report sample sizes, means, and standard deviations. Algebraically equivalent formulas are available for calculating  $d$  based on other statistics, such as  $t$ -tests and  $F$ -tests, but when such information is not available, meta-analysts may choose to estimate  $d$  using formulas for algebraic approximations based on other pieces of information (e.g., statistics from a two-factor repeated measures ANOVA). The accuracy of these approximations may bias effect size estimates. However, if researchers code information regarding the estimation method or level of approximation needed to

calculate the effect size, it is possible to examine these variables as moderators or statistically control for them in the final analyses.

### Study Quality

Another important issue at all stages of meta-analysis, but particularly so at the data analysis phase, is the “quality” of the studies included in the analysis and how that may bias the findings. Indeed, the results of a meta-analysis are entirely dependent on the primary studies of which it is comprised. Many meta-analyses have been legitimately criticized for synthesizing results from low-quality studies. This “garbage-in-garbage-out” criticism of meta-analysis emphasizes the need for a careful consideration of the quality of those studies included and analyzed. The difficulty lies in defining quality, however, and there is no simple definition of quality. There are many tools available that can be used to assess the “risk of bias” that may result from including lower-quality studies in a meta-analysis (Valentine, 2009). The Cochrane Collaboration systematic reviews, for instance, include risk of bias tables that assess the possibility of bias due to the quality (or lack thereof) of the included primary studies (Higgins & Altman, 2008). Most risk of bias instruments are geared toward meta-analyses of randomized controlled trials, however, and are less applicable to etiological/epidemiological or group differences meta-analyses. Therefore, it is also common for meta-analysts to assess study quality in a post hoc, empirical fashion during the analysis stage. Namely, by coding information on study quality such as measurement validity and reliability, implementation fidelity, attrition, blinding, and so forth, the meta-analyst can conduct subgroup or moderator analyses to examine whether those quality variables are associated with the effect size or explain some of the heterogeneity across studies that might be associated with the quality indicators.

### Dependent Effect Sizes

Meta-analysis methods assume the independence of effect size estimates (i.e., that any given analysis only includes one effect size estimate per study). However, many primary studies report enough information to calculate multiple effect size estimates. For example, a researcher may propose to conduct a meta-analysis summarizing the effects of a life skills prevention program on high school students’ alcohol use. It is plausible that several eligible studies may include multiple measures of “alcohol use.” A study may include information on the number of days a student drank any alcohol in the past 30 days, the number of days a student drank any alcohol in the past 90 days, the amount of alcohol consumed on a given occasion, and so forth. Even if the meta-analyst decided these were all eligible outcomes representing the same underlying alcohol use construct of interest, most meta-analysis methods assume the use of only one effect size estimate per study in any given analysis.

To avoid statistical dependencies, most meta-analysts use one of several techniques: (1) they create one average effect size per study, thereby losing the ability to account for the distinct characteristics of the different outcomes; (2) they choose one effect size per study based on some decisional criteria, thereby throwing away information contained in the other effect sizes; or (3) they conduct several separate meta-analyses split by some characteristics of the outcomes or effect sizes, thereby preventing comparison of common moderators. Historically, the only statistically defensible alternative if the meta-analyst wished to include all effect sizes in the same analysis has been to model the dependencies among effect size estimates drawn from the same study (Gleser & Olkin, 2008). This is rarely feasible because it requires information about the correlations among the outcome variables that are virtually never reported by primary study authors. There is a newly developed technique (Hedges, Tipton, & Johnson, 2010) that estimates robust standard errors that can adjust for the lack of statistical independence so that all relevant outcome variables measuring the same outcome construct can be used in the same analysis. This is a new technique, however, and is not yet widely used. Unfortunately, many meta-analysts have altogether ignored the issue of dependent effect size estimates, incorrectly including multiple effect size estimates from the same study in an analysis, with no adjustment for the fact that doing so violates the assumptions of meta-analytic techniques. Doing so can yield incorrect results. In some cases, this underestimates standard errors and increases the possibility of finding a significant effect; in other cases, standard errors may be overestimated. It is therefore important for consumers of meta-analyses to understand whether a meta-analysis has correctly analyzed dependent effect size estimates.

## Summary

In conjunction with primary studies of etiology, epidemiology, intervention effectiveness, and group differences, meta-analysis is an important analytic method for use in the field of prevention science. By allowing researchers to systematically summarize the empirical literature in a given area, meta-analysis can be a powerful tool for informing the science of prevention. Meta-analyses that study the etiology of social problems, the effectiveness of interventions, and the differences between groups can identify high-risk groups or those in need of services, and they can aid in the development and implementation of primary, secondary, and tertiary prevention programs, most especially in the contexts under which they may have the largest impact.

## References

- Bálint, S., Czobor, P., Komlósi, S., Mészáros, Á., Simon, V., & Bitter, I. (2009). Attention deficit and hyperactivity disorder (ADHD): Gender- and age-related differences in neurocognition. *Psychological Medicine*, *39*, 1337–1345.
- Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology*, *92*, 568–582.
- Becker, B. J. (2005). Failsafe N or file-drawer number. In H. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111–125). West Sussex, UK: Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*, 97–111.
- Catalano, R. F., Hawkins, J. D., Berglund, M. L., Pollard, J. A., & Arthur, M. W. (2002). Prevention science and positive youth development: Competitive or cooperative frameworks? *Journal of Adolescent Health*, *31*(Suppl 6), 230–239.
- Chalmers, T. C. (2001). Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In M. Egger, D. Smith, & D. G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (2nd ed., pp. 429–443). London, UK: BMJ Books.
- Chalmers, T. C., Smith, H., Jr., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., et al. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials*, *2*, 31–49.
- Cooper, H. (1984). *The integrative research review: A systematic review*. Newbury Park, CA: Sage.
- Cooper, H. (2010). *Research synthesis and meta-analysis* (Applied social research methods series 4th ed., Vol. 2). Thousand Oaks, CA: Sage.
- Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, *2*, 447–452.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage.
- Dickersin, K. (1997). How important is publication bias? A synthesis of available data. *AIDS Education and Prevention*, *9*, 15–21.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.
- Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomized controlled trials published in English and German. *Lancet*, *350*, 326–329.
- Ekeland, E., Heian, F., Hagen, K. B., Abbott, J., & Nordheim, L. (2005). Exercise to improve self-esteem in children and young people. *Campbell Systematic Reviews*, *4*.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.
- Gleser, L. J., & Olkin, I. (2008). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York: Russell Sage.
- Gotzsche, P. C. (1987). Reference bias in reports of drug trials. *British Medical Journal*, *195*, 654–656.
- Harbord, R. M., Egger, M., & Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, *25*, 3443–3457.

- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65.
- Higgins, J. P. T., & Altman, D. G. (2008). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions (chapter 8)* (pp. 187–241). Chichester, UK: Wiley.
- Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0, updated March 2011). The Cochrane Collaboration. Retrieved from <http://www.cochrane-handbook.org>
- Hopewell, S., & Clarke, M. (2001). Methodologists and their methods: Do methodologists write up their conference presentations or is it just 15 minutes of fame? *International Journal of Technology Assessment in Health Care, 17*(4), 601–603.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.
- Kellam, S. G., Koretz, D., & Mościcki, E. K. (1999). Core elements of developmentally based prevention research. *American Journal of Community Psychology, 24*, 463–482.
- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337–343.
- Lindberg, S. M., Hyde, J. S., Linn, M. C., & Petersen, J. L. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*, 1123–1135.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181–1209.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Littell, J. A., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York: Oxford University Press.
- Matt, G. E., & Cook, T. D. (2009). Threats to the validity of research syntheses. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (2nd ed., pp. 537–560). New York: Russell Sage Foundation.
- Mrazek, P. J., & Haggerty, R. J. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association, 295*, 676–680.
- Petrosino, A., Petrosino, C. T., & Buehler, J. (2004). “Scared Straight” and other juvenile awareness programs for preventing juvenile delinquency. *Campbell Systematic Reviews, 2*.
- Pittler, M. H., Abbot, N. C., Harkness, E. F., & Ernst, E. (2000). Location bias in controlled clinical trials: A survey of three medical journals. *Journal of Clinical Epidemiology, 2000*(53), 485–489.
- Rennie, D. (1997). Thyroid storm. *Journal of the American Medical Association, 277*, 1238–1243.
- Reyes, M. M., Panza, K. E., Martin, A., & Bloch, M. H. (2011). Time-lag bias in trials of pediatric antidepressants: A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry, 50*(1), 63–72.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin, 86*, 638–641.
- Rothstein, H. R., & Hopewell, S. (2008). Grey literature. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 103–125). New York: Russell Sage.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias analysis: Prevention, assessment, and adjustments*. West Sussex, UK: Wiley.
- Rucker, G., Schwarzer, G., & Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine, 27*, 746–763.

- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173–1181.
- Simes, R. J. (1986). Publication bias: The case for an international registry of clinical trials. *Journal of Clinical Oncology*, *4*, 1529–1541.
- Sterne, J. A. C. (Ed.). (2009). *Meta-analysis in Stata: An updated collection from the Stata Journal*. College Station, TX: Stata Press.
- Sterne, J. A. C., Egger, M., & Moher, D. (2008). Addressing reporting biases. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions (chapter 10)*. Chichester, UK: Wiley.
- Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, *53*, 1119–1129.
- Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2011). Risk factors and crime. In F. Cullen & P. Wilcox (Eds.), *The Oxford handbook of criminological theory*. New York: Oxford University Press.
- Tannock, I. F. (1996). False-positive results in clinical trials: Multiple significance tests and the problems of unreported comparisons. *Journal of the National Cancer Institute*, *88*, 206–207.
- Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 129–146). New York: Russell Sage.
- Wilson, D. B. (2009). Systematic coding. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 159–176). New York: Russell Sage.
- Wilson, S. J., & Lipsey, M. W. (2006). The effects of school-based social information processing interventions on aggressive behavior, part I: Universal programs. *Campbell Systematic Reviews*, *5*.