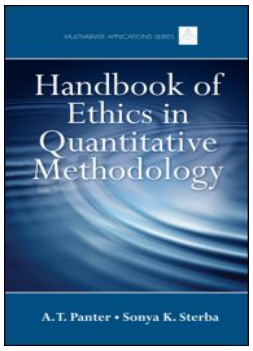

Handbook of Ethics in Quantitative Methodology

A.T. Panter, Sonya K. Sterba

Ethical Issues in the Conduct and Reporting of Meta-Analysis

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203840023.ch16
Harris Cooper, Amy Dent
Published online on: 20 Jan 2011

Section V

Ethics and Communicating Findings

16

Ethical Issues in the Conduct and Reporting of Meta-Analysis

Harris Cooper
Duke University

Amy Dent
Duke University

A research synthesis focuses on empirical studies and attempts to summarize past research by drawing overall conclusions from separate studies that address the same or related hypotheses. The research synthesist's goal is "to present the state of knowledge concerning the relation(s) of interest and to highlight important issues that research has left unresolved" (Cooper, 2010, p. 4). Meta-analysis is a type of research synthesis. It involves the statistical integration of data from separate but similar studies typically using the summary statistics presented in research reports. Meta-analysts (a) systematically collect as many published and unpublished reports addressing a topic as possible, (b) extract effect sizes from the reports, (c) statistically combine the effect sizes to obtain an estimate of the average effect size and the associated confidence interval, and (d) examine sample and study features that might influence study outcomes.

When it comes to ethical considerations, research synthesists and meta-analysts have it easy. Unlike primary researchers, they face no issues regarding the treatment of the humans or animals who participate in their work. There are no *institutional review boards* to convince that the benefits of their work outweigh the risks. Because public documents are the object of study, informed consent and confidentiality are not an issue; public documents cannot be deceived or mistreated.

Still, conducting a research synthesis or meta-analysis is not without ethical considerations. Meta-analysts face the same ethical issues faced by quantitative methodologists discussed in the other chapters in this volume but in a different context. Some of these ethical considerations relate to the process of reporting and publishing research results of any kind. For example, one treatment of ethical obligations in reporting research

can be found in the *Ethical Principles of Psychologists and Code of Conduct* (American Psychological Association [APA], 2002). Here, researchers, whether reporting a new data collection or meta-analysis, are obligated not to fabricate data, to correct errors when they are found, not to plagiarize the work of others or publish data more than once, to allocate authorship credit appropriately, and to share their data with others for purposes of verification. These ethical obligations are reproduced verbatim from the APA *Principles* in Table 16.1.

In the context of discussing the use and misuse of quantitative methods more generally, Brown and Hedges (2009) provide one of the few previous treatments of meta-analysis in an ethical context. They begin by stating one premise that informs all the chapters in this book:

> Methodological rigor is closely related to ethical vigilance: When research, statistical calculations, and data presentation can be done better and more accurately, they should be. That is, there is an ethical imperative to demand and use the highest standards of research and data presentation. (Brown & Hedges, 2009, p. 375)

With regard to meta-analysis in particular, Brown and Hedges identify three points at which a lack of methodological rigor raises ethical issues. First, they point out that meta-analysis can involve collecting, summarizing, and integrating massive amounts of data. Performing these tasks improperly—whether purposely or inadvertently—can lead to erroneous conclusions. Certainly when meta-analyses are conducted improperly on purpose the ethical violation is clear. In the inadvertent case, a lack of vigilance or the carrying out of analyses that are beyond an investigator's expertise also can suggest an ethical breach. Second, Brown and Hedges point out that the decision to include or exclude a study from a meta-analysis can raise ethical issues unless the criteria for study inclusion and exclusion have been made transparent and uniformly applied to studies. Finally, Brown and Hedges assert that it is an ethical obligation of meta-analysts to consider the possibility that publication bias may influence their results.

Thus, it appears these authors suggest three ethical dicta to be followed in conducting and reporting a meta-analysis: (a) extract and analyze your data accurately; (b) make your inclusion and exclusion criteria explicit and apply them consistently; and (c) test for publication bias. Is that it? Perhaps not. Brown and Hedges astutely point out that "what starts out as an identification of best practices can evolve into ethical expectations" (p. 378). We would add to this the suggestion that "best practice" becomes an ethical consideration when the aspect of methodology under consideration is one for which the conclusions of research are heavily dependent. How much does it influence findings if you do

TABLE 16.1

Entries in the *Ethical Principles of Psychologists and Code of Conduct* Relating to Reporting Research Results and Publication

8.10 Reporting Research Results

(a) Psychologists do not fabricate data. (See also Standard 5.01a, Avoidance of False or Deceptive Statements.)

(b) If psychologists discover significant errors in their published data, they take reasonable steps to correct such errors in a correction, retraction, erratum, or other appropriate publication means.

8.11 Plagiarism

Psychologists do not present portions of another's work or data as their own, even if the other work or data source is cited occasionally.

8.12 Publication Credit

(a) Psychologists take responsibility and credit, including authorship credit, only for work they have actually performed or to which they have substantially contributed. (See also Standard 8.12b, Publication Credit.)

(b) Principal authorship and other publication credits accurately reflect the relative scientific or professional contributions of the individuals involved, regardless of their relative status. Mere possession of an institutional position, such as department chair, does not justify authorship credit. Minor contributions to the research or to the writing for publications are acknowledged appropriately, such as in footnotes or in an introductory statement.

(c) Except under exceptional circumstances, a student is listed as principal author on any multiple-authored article that is substantially based on the student's doctoral dissertation. Faculty advisors discuss publication credit with students as early as feasible and throughout the research and publication process as appropriate. (See also Standard 8.12b, Publication Credit.)

8.13 Duplicate Publication of Data

Psychologists do not publish, as original data, data that have been previously published. This does not preclude republishing data when they are accompanied by proper acknowledgment.

8.14 Sharing Research Data for Verification

(a) After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release. This does not preclude psychologists from requiring that such individuals or groups be responsible for costs associated with the provision of such information.

(b) Psychologists who request data from other psychologists to verify the substantive claims through reanalysis may use shared data only for the declared purpose. Requesting psychologists obtain prior written agreement for all other uses of the data.

Source: From APA, *Ethical Principles of Psychologists and Code of Conduct*, APA, New York, 2002. With permission.

it right or wrong? How easy is it to "manipulate" findings (to arrive at a predetermined or biased outcome) by doing it one way or the other? Because the techniques used in meta-analysis are relatively new and still evolving, we can anticipate that standards of best practice and ethical expectations are also rapidly evolving.

Examining how the standards surrounding meta-analysis are evolving is the objective of this chapter. To begin, we will present and provide a brief background concerning a set of guidelines for the reporting of meta-analyses recently developed by the APA, called *meta-analysis reporting standards* (MARS; APA Publication and Communication Board Working Group on Journal Article Reporting Standards,[1] 2008). Then, we will describe the results of a survey conducted using the members of the Society for Research Synthesis Methodology as respondents. Participants were asked about what aspects of a meta-analysis were and were not important to report. Those that rose to a level that suggested omitting these aspects of meta-analysis might be considered a breach of ethics will be discussed in some detail. Finally, we will conclude with some other ethics-related issues that have emerged for meta-analysts, specifically, the use of *auxiliary websites*, the use of individual participant data in meta-analysis, and the identification of duplicate publications.

APA's Meta-Analysis Reporting Standards

In developing its meta-analysis reporting standards, the APA Working Group distinguished between three levels of prescription: "recommendations," "standards," and "requirements." Using Merriam Webster's Online Dictionary (2007) as its source of definitions, *to recommend* was defined as "to present as worthy of acceptance or trial … to endorse as fit, worthy, or competent …"; a *standard* was defined as "… something set up and established by authority as a rule for the measure of quantity, weight, extent, value, or quality …"; and a *requirement* was defined as something that was asked for "by right and authority … to call for as suitable or appropriate … to demand as necessary or essential …." From an ethical perspective, failing to meet requirements certainly could be considered problematic, but failing to meet recommendations or standards would be less troubling, depending on the circumstance. The APA Working Group decided that its proposals

[1] The first author of this chapter served as chair of this working group.

should "… be viewed as standards or, at least, a beginning effort at developing standards" (p. 847).

MARS was developed by integrating four efforts by other groups of researchers and editors knowledgeable about meta-analysis: the *QUOROM Statement* (*Quality of Reporting of Meta-analysis*; Moher et al., 1999) and its revision, *PRISMA* (*Preferred Reporting Items for Systematic Reviews and Meta-analyses*; Moher, Tetzlaff, Liberati, Altman, & the PRISMA Group, 2009), *MOOSE* (*Meta-analysis of Observational Studies in Epidemiology*; Stroup et al., 2000), and the Potsdam consultation on meta-analysis (Cook, Sackett, & Spitzer, 1995). The APA Working Group combined the nonredundant elements contained in these previous documents, rewrote some items for an audience of psychologists (and others who might use APA's *Publication Manual*, 2010), and added a few suggestions of its own. Then, the APA Working Group asked for additional suggestions from a subgroup of members of the Society for Research Synthesis Methodology known to have interest in psychology and from the members of APA's Publications & Communications Board. The approved final version (along with the Journal Article Reporting Standards [JARS]) appeared in the *American Psychologist* in December 2008 and was reproduced in the sixth edition of the APA *Publication Manual* (2010). The MARS is reproduced in Table 16.2.

A Survey of the Members of the Society for Research Synthesis Methodology

The MARS calls on meta-analysts to address more than 70 different aspects of methodology in their reports. Certainly, these are not of equal import, and their importance would vary as a function of the topic under consideration. Perhaps then, another level of guidance is needed: Which aspects of meta-analysis reporting are optional, context dependent, and required for the authors to have met their obligations as researchers?

To answer this question, we conducted an online survey of the 74 members of the Society for Research Synthesis Methodology. Forty-two (57%) of the Society's members responded to the survey 2 weeks after receiving the invitation (and 1 week after a reminder was sent). In addition to responding to the survey, we also asked participants several questions about their background. In one, they categorized their broad area of interest into one of five categories. Of the 42 respondents, 18 chose medicine and health; 21 chose psychology, education, or public and social policy; and 3 chose "other." Also, we asked how many years the respondents

TABLE 16.2

Meta-Analysis Reporting Standards

Paper Section and Topic	Description
Title	
	• Make it clear that the report describes a research synthesis and include "meta-analysis," if applicable • Footnote funding source(s)
Abstract	
	• The problem or relation(s) under investigation • Study eligibility criteria • Type(s) of participants included in primary studies • Meta-analysis methods (indicating whether a fixed or random model was used) • Main results (including the more important effect sizes and any important moderators of these effect sizes) • Conclusions (including limitations) • Implications for theory, policy, and/or practice
Introduction	

	• Clear statement of the question or relation(s) under investigation
	  • Historical background
	  • Theoretical, policy, and/or practical issues related to the question or relation(s) of interest
	  • Rationale for the selection and coding of potential moderators and mediators of results
	  • Types of study designs used in the primary research, their strengths and weaknesses
	  • Types of predictor and outcome measures used, their psychometric characteristics
	  • Populations to which the question or relation is relevant
	  • Hypotheses, if any

Method	
Inclusion and exclusion criteria	• Operational characteristics of independent (predictor) and dependent (outcome) variable(s) • Eligible participant populations • Eligible research design features (e.g., random assignment only, minimal sample size)

	• Time period in which studies needed to be conducted • Geographical and/or cultural restrictions
Moderator and mediator analyses	• Definition of all coding categories used to test moderators or mediators of the relation(s) of interest

Search strategies	• Reference and citation databases searched
	• Registries (including prospective registries) searched
	  • Keywords used to enter databases and registries
	  • Search software used and version (e.g., Ovid)
	• Time period in which studies needed to be conducted, if applicable
	• Other efforts to retrieve all available studies, e.g.,
	  • Listservs queried
	  • Contacts made with authors (and how authors were chosen)
	  • Reference lists of reports examined
	• Method of addressing reports in languages other than English
	• Process for determining study eligibility
	  • Aspects of reports were examined (i.e., title, abstract, and/or full text)
	  • Number and qualifications of relevance judges
	  • Indication of agreement
	  – How disagreements were resolved
	• Treatment of unpublished studies

Coding procedures	• Number and qualifications of coders (e.g., level of expertise in the area, training)
	• Intercoder reliability or agreement
	• Whether each report was coded by more than one coder and, if so, how disagreements were resolved
	• Assessment of study quality
	  • If a quality scale was employed, a description of criteria and the procedures for application
	  • If study design features were coded, what these were
	• How missing data were handled

(*Continued*)

TABLE 16.2 (Continued)

Meta-Analysis Reporting Standards

Paper Section and Topic	Description

Statistical methods	• Effect size metric(s)
	  • Effect sizes calculating formulas (e.g., means and SDs, use of univariate F to r transform, etc.)
	  • Corrections made to effect sizes (e.g., small sample bias, correction for unequal ns, etc.)
	• Effect size averaging and/or weighting method(s)
	• How effect size confidence intervals (or standard errors) were calculated
	• How effect size credibility intervals were calculated, if used
	• How studies with more than one effect size were handled
	• Whether fixed and/or random effects models were used and the model choice justification
	• How heterogeneity in effect sizes was assessed or estimated
	• Means and SDs for measurement artifacts, if construct-level relationships were the focus
	• Tests and any adjustments for data censoring (e.g., publication bias, selective reporting)
	• Tests for statistical outliers
	• Statistical power of the meta-analysis
	• Statistical programs or software packages used to conduct statistical analyses

Results	

	• Number of citations examined for relevance
	• List of citations included in the synthesis
	• Number of citations relevant on many but not all inclusion criteria excluded from the meta-analysis
	  • Number of exclusions for each exclusion criteria (e.g., effect size could not be calculated), with examples
	• Table giving descriptive information for each included study, including effect size and sample size
	• Assessment of study quality, if any
	• Tables and/or graphic summaries
	  • Overall characteristics of the database (e.g., number of studies with different research designs)
	  • Overall effect size estimates, including measures of uncertainty (e.g., confidence and/or credibility intervals)

- Results of moderator and mediator analyses (analyses of subsets of studies)
  - Number of studies and total sample sizes for each moderator analysis
  - Assessment of interrelations among variables used for moderator and mediator analyses
- Assessment of bias including possible data censoring

Discussion

- Statement of major findings
- Consideration of alternative explanations for observed results
  - Impact of data censoring
- Generalizability of conclusions, e.g.,
  - Relevant populations
  - Treatment variations
  - Dependent (outcome) variables
  - Research designs, etc.
- General limitations (including assessment of the quality of studies included)
- Implications and interpretation for theory, policy, or practice
- Guidelines for future research

Source: From Journal Article Reporting Standards Working Group, *Am. Psychol.*, 63, 839, 2008. With permission.

had been working with research synthesis methodologies. Twenty of the respondents said they had 15 or more years of experience.[2]

Each participant was told that he or she would be presented with about 70 different aspects of conducting a research synthesis or meta-analysis and would respond to the question "Generally speaking, how important is it that each be described in the report of a synthesis?"[3] The response scale was:

10 = Generally, it would be considered UNETHICAL in my field NOT TO INCLUDE this information in the report (10 on the scale was labeled "generally must include").

5 = Generally, researchers in my field MIGHT or MIGHT NOT report this information depending on characteristics of the specific literature (5 read "depends on the specific literature").

1 = Generally, it is UNNECESSARY in my field for researchers to report this information (1 read "generally unnecessary")[4]

N/A = Generally, this aspect of research synthesis is NOT APPLICABLE to my area of interest.

A comment box was provided after each request for a rating.

Table 16.3 presents the results of the survey. Note that nearly all of the elements of the MARS were included on the survey with a few exceptions we deemed to be trivial (e.g., we deemed it unnecessary to ask the Society members what the elements of a title or abstract ought to be for a synthesis report).

[2] We used these two background questions to examine whether responses to the survey were related to the participant's background. However, of 280 statistical tests we conducted (comparing the area and experience of members on the (a) frequency of scores of 10 and (b) mean response for each of the 70 items), we found two that reached statistical significance ($p < .05$). Given that this was less than the expected number of significant findings if chance were operating, we concluded that responses did not differ as a function of substantive area or experience.

[3] It might have been more interesting to poll the experts on what they considered best practice in conducting a meta-analysis rather than on reporting standards. However, issues regarding best practice in meta-analysis are more complex and in some instances still contentious. This would make the relevance of such a survey to ethical issues more difficult to discern. For example, meta-analysis experts disagree about whether it is ever appropriate to use a fixed-effect, rather than a random-effect, model for estimating error; we will see that most agree that reporting which model was used and why is a necessary part of reporting.

[4] Two respondents pointed out that the survey left it ambiguous whether answers should be based on their personal opinion or on the norms that prevailed in their field. We should have made it clearer that we were interested in norms of the field. Regardless, for the relatively broad purposes to which the survey is put, we think it is not a severe problem that responses include a mix of perceived norms and the wishes of experts.

TABLE 16.3

Results of the Survey of the Society for Research Synthesis Methodology Regarding Reporting Standards for Meta-Analysis

Aspect of Meta-Analysis Reporting	Number of 10s[a]	Number Below 5[b]	Mean
Introduction			
1. Clear statement of the research question	**34**	**0**	**9.74**
2. Narrative account of the development of the research question	8	2	6.73
3. Theoretical, policy, and/or practical issues related to the research question	9	4	7.36
4. Rationale for the selection and coding of potential moderators and mediators of results	14	2	7.62
5. Types of study designs … their strengths and weaknesses	14	3	7.80
6. Independent (predictor) and dependent (outcome) variables of primary interest	**25**	**1**	**9.07**
7. Populations to which the question is relevant	**23**	**2**	**8.45**
8. Hypotheses, if any	14	4	7.46
Methods			
1. Operational definitions of independent (predictor) and dependent (outcome) variable(s)	**23**	**2**	**8.51**
2. Eligible participant populations	**24**	**1**	**8.88**
3. Eligible research design features …	**28**	**2**	**8.90**
4. Time period in which studies needed to be conducted	20	3	8.17
5. Geographical and/or cultural restrictions	10	4	6.69
6. Whether unpublished studies were included or excluded	**24**	**2**	**8.67**
7. Reference and citation databases searched	**30**	**2**	**9.19**
8. Registries (including prospective registries) searched	17	7	7.30
9. Keywords used to enter databases and registries	17	4	8.05
10. Search software used to enter electronic databases (e.g., Ovid)	9	18	5.57
11. Conference proceedings searched	7	11	6.23

(*Continued*)

TABLE 16.3 (Continued)

Results of the Survey of the Society for Research Synthesis Methodology Regarding Reporting Standards for Meta-Analysis

Aspect of Meta-Analysis Reporting	Number of 10s[a]	Number Below 5[b]	Mean
12. Listservs queried	7	10	5.89
13. Contacts made with researchers in the field and how these researchers were chosen	5	9	6.12
14. Whether reference lists of reports were examined	12	4	7.62
15. Method of addressing reports in languages other than English	14	5	7.14
16. Aspects of reports used to determine relevance (i.e., title, abstract, and/or full text)	16	7	7.57
17. Number and qualifications of relevance judges	13	10	6.81
18. Indications of judge agreement if more than one judge examined each report	15	7	7.31
19. How judge disagreements were resolved	17	4	7.90
20. Number and qualifications of coders (e.g., level of expertise in the area, training)	10	10	6.59
21. Intercoder reliability or agreement	15	7	7.43
22. Whether each report was coded by more than one coder … how disagreements … resolved	19	3	8.02
23. How missing data were handled	17	2	8.12
24. Definitions of ALL coding categories …	13	3	7.60
25. Criteria of the quality scale and procedure for application	20	6	7.80
26. Study design features that were coded	**23**	**1**	**8.74**
27. Effect size metric(s)	**34**	**0**	**9.68**
28. Effect sizes calculating formulas …	8	11	6.24
29. Corrections made to effect sizes …	20	1	8.34
30. Effect size averaging and weighting method(s)	**31**	**0**	**9.00**
31. How effect size confidence intervals (or standard errors) were calculated	17	3	8.13
32. How effect size credibility intervals were calculated	10	4	6.89
33. How studies with more than one effect size were handled	**27**	**1**	**8.93**

34. Whether fixed and/or random effects models were used	**32**	**2**	**9.31**
35. The justification for the choice of the error model (fixed, random)	**22**	**7**	**8.00**
36. How heterogeneity in effect sizes was assessed or estimated	**23**	**1**	**8.93**
37. Means and SDs for measurement artifacts	2	7	5.24
38. Tests and any adjustments for data censoring (e.g., publication bias, selective reporting)	14	3	7.83
39. Tests for statistical outliers	9	8	6.14
40. Statistical power of the meta-analysis	3	23	4.07
41. Statistical programs or software packages used to conduct statistical analyses	16	9	7.36

[a] "Number of 10s" is the number of respondents out of 42 who said, "Generally, it would be considered UNETHICAL in my field NOT TO INCLUDE this information in the report" (10 on the scale was labeled "generally must include").

[b] "Number below 5" is the number of respondents who gave this reporting aspect a score less than 5.

[c] Questions with " … " have been shortened in the table by removing examples or clarifying information. Precise wording of MARS questions can be found in Table 16.2.

To begin interpreting Table 16.3, it is interesting to look at the responses regarding Brown and Hedges' (2009) three dicta for conducting a meta-analysis. The first of the dicta—extract and analyze your data accurately—is hard to attach to any particular question or questions because it is a broad prescription. The second dicta—make your inclusion and exclusion criteria explicit and apply them consistently—relates to several questions on the survey. These included the operational definitions of independent and dependent variables, eligible participant populations, eligible research designs, and any time period, geographic, or cultural restrictions. For these questions, between 10 and 28 respondents answered that "it would be considered UNETHICAL in my field NOT TO INCLUDE this information in the report." The third dicta—test for publication bias—relates to two questions. Twenty-four respondents believed it would be unethical not to report whether unpublished research was included or excluded from the synthesis, and 14 believed that authors must include tests and any adjustments for data censoring (e.g., publication bias, selective reporting).

Based on these results, we think it would not be unreasonable to suggest that if 21 (50%) or more of the Society's members responded that they believed not including this information in a report would be considered unethical or generally must be included that this be viewed as an indication that this reporting practice now approaches the point where best practice becomes an ethical expectation. Below we present the elements of reporting that reached this threshold and present our thinking about why this was the case; in other words, why the choices researchers make at these points can have large effects on the results of their syntheses.

Aspects of Meta-Analysis Reporting Approaching Ethical Obligation

The Problem Statement

Three aspects of reporting syntheses that reached our threshold for raising ethical concerns related to the problem statement. More than half of respondents believed that it would be unethical not to include (a) a clear statement of the research question (n = 34; 81%), (b) what were the independent (predictor) and dependent (outcome) variables of primary interest (n = 25; 60%), and (c) a description of the populations to which the question is relevant (n = 23; 55%). For example, a synthesis that claims to examine the relationship between frustration and aggression would need to provide a clear statement of how the variables of interest are

defined conceptually (e.g., frustration involves the blocking of goal attainment; aggression involves the intent to harm), what type of relationship is of interest (associational or causal), and among whom (e.g., animals or humans; children, adolescents, or adults).

A high level of concern regarding these aspects of reporting would probably be evident in a similar survey related to primary research. Without a clear statement of the problem, the variables involved, and the relevant populations, it would be impossible to evaluate the contribution the research makes to the literature, if indeed the relevant literature could be identified.

The Inclusion Criteria

Three aspects of reporting the criteria for including and excluding studies from the synthesis reached our threshold for raising ethical concerns. More than half of respondents believed it would be ethically problematic not to include in the method section (a) the operational definitions of independent (predictor) and dependent (outcome) variables (n = 23; 55%); (b) a description of the eligible participant populations (n = 24; 57%); and (c) the eligible research design features (n = 28; 67%). Nearly half (n = 20; 48%) gave the highest rating to the need to include any time period restrictions. These concerns about the inclusion and exclusion criteria parallel the conceptual concerns that arose when respondents rated the importance of aspects of the problem statement. There are good reasons for this, and these reasons are especially relevant to research syntheses.

Eligibility criteria take on unique importance because the research designs and characteristics of units sampled in a research synthesis can be considerably more varied than typically is the case for a single primary study. Research synthesists often begin their work with broad conceptual definitions. In the course of searching the literature they may come across numerous operational realizations of the concepts defined in their problem statement. For example, synthesists examining the relation between frustration and aggression might discover studies that used numerous techniques to measure or instill frustration in participants (e.g., asking them to wait in line for a long time, playing a video game that is difficult to win) and numerous ways to measure aggression (e.g., shouting, pushing, hitting). Given this variety, our respondents believed it was of great importance that readers know precisely what the synthesists defined as "in" and "out." Only with this information would readers be able to object, for example, if "shouting" is included as a measure of aggression because they believe verbal attacks are not really meant to harm.

Also, readers might have no objection to the conceptual and operational definitions of the problem but may want to judge whether the concepts and operations fit together well. They may want to determine whether the

operations used in previous research fit the concept definitions used by the synthesists, or whether a broader or narrower conceptualization would be more appropriate. For example, the synthesists might find "shouting" was used as a measure of aggression when initially they had not considered verbal aggression for inclusion. When such a circumstance arises, the synthesists must broaden their conceptual definitions to include these operations, so now aggression includes both physical and verbal assault. If this "refitting" is not done, the conclusions in the synthesis might appear to apply more generally or narrowly than warranted by the data. Readers will not be able to assess fit without a clear statement of the included and excluded operational definitions.

Similarly, synthesists need to clearly tell readers what units were and were not considered relevant to addressing the research question. Without this information, the reader cannot assess to whom the results apply. Nor can the reader object if he or she believes samples from irrelevant populations have been included in the synthesis. One respondent commented that this information was usually included but was vague and "not explicitly defined a priori—also not necessarily considered in relation to external validity." Another commented that in her or his area, people "generally use convenience samples and we agree not to talk about it."

Of equal importance, a clear description of the methodological characteristics of included and excluded studies allows the reader to gauge the fit between the included research designs, how the design was implemented, and the inferences drawn by the synthesists (Valentine & Cooper, 2008). Above, the survey respondents indicated that research synthesists must provide an explicit statement about the type of relationship under study. This aspect of rigorous research synthesis takes center stage when readers consider the correspondence between the design and implementation of individual studies and the desired inferences of the synthesis. For example, the research question "Does frustration cause aggression?" suggests that the research synthesists should focus primarily on summarizing research that used experimental and quasi-experimental designs, whereas the question "Is frustration associated with aggression?" might include cross-sectional designs as well. How readers will evaluate how well the synthesists' inferences correspond with the data will depend on what kinds of designs were admitted as evidence.

We think as well that this same line of reasoning was behind the respondents' frequent use of the highest ratings ($n = 23$; 55%) for the importance of including a thorough description of the study design features that were coded by the synthesists. Here, the description relates only to studies that were included in the synthesis, but the principle is the same. Also, 20 respondents, one short of our criteria, gave the highest rating to the importance of describing the criteria of the quality scale and procedure for its application. This element of a research report might not have reached

our threshold because not all meta-analysts think using quality scales is a good idea (Valentine & Cooper, 2008).

In sum, the inclusion–exclusion aspects of reporting that rose to the level of ethical considerations relate to the readers' ability to evaluate the fit between concepts and operations in research synthesis and the fit between the inferences drawn and what inferences the data can support. Without this information, readers will be unable to decide whether clear and legitimate linkages exist (a) between concepts and operations and (b) between research designs, study implementation, and the interpretation of results (see Cooper, 2010).

The Parameters of the Literature Search

Whether Unpublished Studies Were Included or Excluded

Not surprisingly, respondents felt strongly about the need to report whether unpublished studies were included in the research synthesis (n = 24; 57%). The concern here is that studies revealing smaller effects will be systematically omitted from the published literature, making relationships appear stronger than if all estimates were retrieved and included. Lipsey and Wilson (1993) compared the magnitudes of effects reported in published versus unpublished studies contained in 92 different meta-analyses. They reported that on average the impact of interventions in unpublished research was one-third smaller than published effects.

A reason frequently given for excluding unpublished research is that it has not undergone the peer-review process and therefore may be of lesser quality. However, researchers often do not publish their results because publication is not their objective (cf. Cooper, DeNeve, & Charlton, 2001); publication does not help them get their work before the audience they seek. For example, some research is conducted to meet degree or course requirements or as evaluations for agencies making decisions about program effectiveness. Also, research is often turned down for journal publication because it is not a novel contribution (although direct replications are of great interest to research synthesists) or because the statistical test fails to achieve standard levels of statistical significance, a problem known as "bias against the null hypothesis" (Rothstein, Sutton, & Borenstein, 2005). Conversely, some low quality research does get published.

For these reasons, it is now "best practice" in the social sciences for research synthesists to include both published and unpublished research. If the synthesists include only published research, their report must include a convincing justification. Our survey responses suggest that providing a clear description and justification for whether and why unpublished

research was or was not included in the synthesis has crossed over into an ethical obligation.

Reference and Citation Databases Searched

The sources of information that provide the most evidence that goes into a research synthesis are likely to be reference databases and citation indexes. Even though reference databases are superb sources of studies, they still have limitations. First, different reference databases restrict what is allowed to enter the system based on their topical or disciplinary coverage. Second, some reference databases contain only published research; others contain both published and unpublished research; and others contain just unpublished research (e.g., dissertation abstracts). Third, there can be a time lag between when a study is completed and when it will appear in the reference database (although technology has reduced this lag dramatically), and this may vary depending on the database. Without information on the databases used, it will be difficult for readers to assess (a) the literature coverage and (b) what studies might have been missed. Equally important, without this information it would be extremely difficult to replicate the results of the synthesis.

The Measure of Effect

The Effect Size Metric(s)

Although numerous estimates of effect size are available (Cohen, 1988), three dominate the literature: (a) the *d*-index, which is a scale-free measure of the separation between two group means calculated by dividing the difference between the two group means by either their average standard deviation or the standard deviation of the control group; (b) the *r*-index, or correlation coefficient; and (c) the odds ratio, or some variant thereof, applicable when both variables are dichotomous and findings are presented as frequencies or proportions.

The term *effect size* is sometimes used broadly to denote all measures of relationship strength, and sometimes it is used as an alternative label for the *d*-index. This is regrettable because the metrics, although translatable, are not identical. For example, a value of .40 for a *d*-index corresponds to an *r*-index value of .196. Thus, as one respondent noted, "If we don't know the metric then we don't know how to evaluate" the results. Further, it is not always the case that the choice of an *effect size metric* reflects the important design characteristics of the studies from which they are derived (specifically the dichotomous or continuous nature of the variables involved). Therefore, the survey respondents indicated that readers of research syntheses need to be explicitly informed of what indexes were used and why they were chosen ($n = 34$).

Effect Size Averaging and Weighting Method(s)

Once each effect size has been calculated, the meta-analysts next average the effects that estimate the same relationship. It is generally accepted that the individual effect sizes should be weighted by the inverse of the variance (based on the number of participants in their respective samples) before they are averaged. Sometimes, however, unweighted effect sizes are presented. Weighted and unweighted effect sizes can differ in magnitude, the difference depending on the degree of relationship between the size of the effect and the sample size. Therefore, if larger effects are associated with smaller sample sizes (a condition likely obtained if the synthesists were more likely to find studies that produced statistically significant results), an unweighted average effect size would be larger, sometime much larger, than the weighted average. For this reason, the survey respondents found it important that the procedures used to generate average effect sizes were essential to a complete synthesis report.

How Studies With More Than One Effect Size Were Handled

A problem for meta-analysts arises when a single study contains multiple effect size estimates. This is most bothersome when more than one measure of the same construct appears in a study and the measures are analyzed separately. Because the same participants provided multiple outcomes, these measures are not independent, and it would generally be inappropriate to treat them as such when combining the effect sizes across all studies. If they are, studies with more measures would get more weight in an averaged effect size, and the assumption that effect size estimates are independent would be violated in subsequent analyses.

There are several approaches meta-analysts use to handle dependent effect sizes. Some meta-analysts treat the effect size as if it were independent. Alternatively, the study might be used as the unit of analysis, taking the mean or median effect size to represent the study. Another approach is to use a shifting unit of analysis (Cooper, 2010). Here, each effect size associated with a study is first coded as if it were an independent estimate of the relationship. However, for estimating the overall effect size, these are averaged before entry into the analysis, so that the study only contributes one value. However, in analyses that examine moderators, for example, whether physical or verbal aggression is influenced more by frustration, the studies would be permitted to contribute one effect size to the estimate of each category's mean effect size. Finally, more sophisticated statistical approaches also have been suggested as a solution to the problem of dependent effect size estimates (Gleser & Olkin, 2009).

Which of the available techniques the meta-analysts use can have a large impact on the estimated average magnitude of the effect size, the

estimated variance among effect sizes, and the power of tests to uncover moderators of effects? For this reason, a majority of respondents believed meta-analysts might be ethically obligated to report which approach was used in handling nonindependent estimates of effect. One respondent also pointed out that "Specific details are preferable. For instance, simply stating that a 'shifting unit of analysis' approach was used doesn't specify how 'average' effect sizes were computed or how the conditional variances and weights were handled."[5]

Variation Among Effect Sizes

Three aspects of meta-analysis methodology related to how the variation in effect sizes was treated generated a majority of responses at the high extreme of the scale: whether fixed-effect or random-effect models were used, the justification for the use of a fixed-effect or random-effect model, and how heterogeneity in effect sizes was assessed or estimated.

One important aspect of averaging effect sizes and estimating their dispersion involves the decision about whether a fixed-effect or random-effect model underlies the generation of study outcomes. In a *fixed-effect model*, each effect size's variance is assumed to reflect sampling error of participants only, that is, error solely the result of participant differences. However, other features of studies can be viewed as additional random influences. Thus, in a random-effect analysis, study-level variance is assumed to be present as an additional source of random influence. Hedges and Vevea (1998, p. 3) state that fixed-effect models are most appropriate when the goal of the research is "to make inferences only about the effect size parameters in the set of studies that are observed (or a set of studies identical to the observed studies except for uncertainty associated with the sampling of subjects)." A further statistical consideration is that in the search for moderators fixed-effect models may seriously underestimate error variance and random-effect models may seriously overestimate error variance when their assumptions are violated (Overton, 1998). Schmidt, On, and Hayes (2009) suggest that random-effect models should always be used, whereas Cooper (2010) proposes that both models can be applied to the data and the results interpreted accordingly.

Random-effect models are typically more conservative than fixed-effect models, in the sense that they will estimate more variability around average effect sizes and therefore are less likely to reveal statistically significant effects and moderators of effects. The two models also can generate different average effect sizes, again depending on the relationship between

[5] Another respondent mused theologically that "Cardinal Newman tried to convince people of the existence of God … and he enunciated that it is more compelling to have independent evidence than repeated versions of the same evidence"

the size of effects and the size of samples. A large majority of respondents to our survey (n = 32) believed meta-analysts were ethically obligated to report whether a fixed- or random-effect model was used, why it was chosen (n = 22; 52%), and more generally how heterogeneity in effect sizes was assessed or estimated (n = 23; 55%).

Tabling Data

Not surprisingly, 26 (62%) respondents believed that meta-analysts were obligated to present a summary table of their results. Related to this table, a near majority of respondents (n = 20; 48%) used the extreme high end of the scale when rating the importance of including information on the number of studies, sample sizes, and subgroup effect sizes for each moderator analysis.

Given these results, perhaps what is surprising is that a table listing the results for the individual studies going into the analyses did not meet our threshold, although 17 (40%) respondents did give this feature of a report the "ethically-obligatory" rating. One respondent commented about tables of individual studies, "If I could have scored this '11' I would have done so!" We suspect that this table did not receive more ratings of 10 because of concerns about space limitations in many reports, especially ones being readied for publication. Two respondents supported this interpretation; one wrote, "But subjected to space limitations," and another noted, "We usually prepare the table for each meta-analysis, however, some journals do not publish it due to the limited number of tables/figures allowed." We will return to this issue below.

The Interpretation of Results

Finally, a majority of our respondents believed that four of the nine MARS elements of a discussion section were obligatory: a statement of major findings (n = 38; 90%); consideration of alternative explanations for observed results (n = 23; 55%); a discussion of the general limitations (including assessment of the quality of studies included) (n = 23; 55%)[6]; and a discussion of the implications and interpretation of findings for theory, policy, or practice (n = 24; 57%). Another element, populations to whom the results were relevant (n = 19; 45%), approached our threshold. Similar to elements of the introduction, these are likely viewed as obligatory of all good science reporting, not just research synthesis.

[6] One respondent cautioned, "Although these need careful thought to avoid biased interpretations."

Some Additional Issues Related to the Reporting of Meta-Analysis

Space Limitations and the Use of Auxiliary Websites

One issue that arises when reporting standards are discussed is the tension between the desire to exhaustively report the background, methods, and results of a study and the space limitations of journals. In fact, this issue generated the most open-ended comments from the Society members responding to our survey. One member wrote:

> Something must be said about practicality and the limitations imposed by editors. [For example] in a recent paper we had 26 pages of included references and the editors wanted us to condense a review of over 200 studies into 40 pages max with all tables. Adding all these potentially important details is impossible in most reports.

Another wrote:

> A major difficulty I've encountered with reporting research syntheses is journal space constraints. It's often infeasible to include sufficient detail about every synthesis phase, especially if there are several studies, complex results (e.g., several distinct effect sizes and multiple moderator analyses). Reporting things like details of each study is essentially impossible, especially in outlets where a certain amount of "didactic" detail about meta-analysis is required for readers unfamiliar with meta-analysis.

And a third wrote:

> We just got a meta-analysis tentatively accepted to … and they are asking us to omit nearly all of the tables and "technical" details. We do not plan to do this totally but probably will need to relegate it to appendices. We are going to resist this as much as possible and will cite JARS/MARS as part of our argument for including it.

Journals have only limited printed pages, and the detail needed to report completely one study conflicts with the desire of the journal to publish as many (worthy) studies as possible. As noted above, in research synthesis this issue arises most frequently when considering whether to publish the table of characteristics and results of individual studies. And, as one of our respondents suggested, sometimes even just the references to these studies can go on for pages.

Today the availability of the Internet eases this tension somewhat. Electronic publishing has largely removed page limitations from what

journals can publish. Still, too much length can be a problem if it leaves authors and readers feeling swamped with data that make it difficult to distinguish what is important and less important in the details of research. For this purpose, many (and an increasing number of) journals now provide auxiliary websites on which material can be placed rather than including it in the print version or the formal electronic version of the article. If the publisher does not make an auxiliary website available, some authors provide this information on their personal web pages and footnote its availability in the article. In electronic versions of articles, the supplemental information that resides on separate web pages can be linked to the article at the point in the report that they would otherwise appear.

It seems then that space limitations should no longer be a justification for the incomplete reporting of meta-analyses. However, when using auxiliary websites, another obligation arises: Authors must provide sufficient documentation accompanying the content of auxiliary websites so that readers can understand how this information is to be interpreted. For example, if meta-analysts want to share the coding sheet they used to retrieve information from studies, the coding sheets need to be completely labeled and contain the code book (which provides definitions and coding conventions) and the coding sheet itself. Similarly, a table (or spreadsheet) that contains the specific codes entered for each study needs to be accompanied by definitions for any abbreviations that are used in the table. Although this seems obvious and straightforward, many readers of web-based articles have experienced frustration when the auxiliary website contained information that was not interpretable. Such presentations do not meet the authors' obligation to report their methods and results thoroughly and clearly.

Data Sharing and Meta-Analysis With Individual Participant Data and With Aggregate Statistics

At the beginning of this chapter, we defined meta-analysis as the use of aggregate data (AD) from previous research to conduct a research synthesis. Increasingly, meta-analyses are conducted by obtaining and cumulating individual participant data (IPD). Unlike meta-analyses based on AD, IPD meta-analysis involves the collection, checking, and reanalysis of the raw data from each study to obtain combined results.

Cooper and Patall (2009) examine the relative benefits of the two types of meta-analysis. They concluded that if both IPD and AD are equally available, meta-analysis using IPD is the superior approach: IPD meta-analysis permits (a) new analyses of the data, (b) checking of the data and

original analyses for errors, (c) addition of new information to the data sets, and (d) use of different statistical methods. However, Cooper and Patall also point out that because of the cost of IPD meta-analysis and the lack of available individual participant data sets, the best strategy currently is to use both approaches in a complementary fashion; the first step in conducting an IPD meta-analysis might be to conduct a meta-analysis with AD.

Three additional ethical issues become important when we consider the differences between meta-analysis with AD and IPD. These are the issues of data sharing, authorship, and the rights to confidentiality of the participants in the primary research.

With regard to data sharing, Cooper and Patall wrote:

> The incentives for data sharing are increasing while the barriers are coming down. Advances in data storage and ease of data transfer are barriers that have largely been removed. A recent incentive is the development and heightened enforcement of policies requiring or encouraging sharing of data collected with federal funding (National Institutes of Health, 2003). (Cooper & Patall, 2009, p. 174)

Rights to authorship is an issue related to data sharing. Often in medicine, where meta-analysis with IPD is undertaken much more frequently than in the social sciences, multiple primary researchers come together and form a consortium that collects and carries out the meta-analysis. In this case, the meta-analysis may be published under the joint authorship of the consortium with the individual contributors acknowledged in an author note. If such an arrangement is not possible, it is essential that the meta-analysts come to prior agreement with the collectors of the original data regarding how authorship will be handled.

Finally, the reuse of data in an IPD meta-analysis research project also raises issues about the right to confidentiality of the research participants. Here, the ethical issues are no different from those encountered for any secondary use of data. Guidelines covering these uses are fluid. Still, whether an individual's agreement to participate in the original study also made explicit or implied consent to have data included in a secondary analysis is a question that both the original collectors of the data and the IPD meta-analysts must answer. Making data sets available to IPD meta-analysts must occur only under the same rules of confidentiality that applied when the data were first collected. Typically, if the data are not shared until they have been stripped of any and all identifying information, then the investigation is no longer research with human subjects.[7]

[7] The Office for Human Research Protections (OHRP) of the Department of Health and Human Services document *Guidance on Research Involving Coded Private Information or Biological Specimens* states, "With respect to research involving private information and

Uncovering Duplicate Publication

As a final issue, on occasion meta-analysts may find themselves acting as ethics enforcers. This occurs when in the course of gathering studies and extracting information from them, the meta-analysts identify instances in which researchers have engaged in *duplicate publication* (see point 8.13 in Table 16.1, and also Levin, Chapter 18, this volume).

Sometimes the line between what is and is not duplicate publication is clear. For example, a meta-analyst would not consider it duplicate publication if he or she came across a dissertation, convention paper, and journal article all presenting the same data (although in some fields papers presented at some meetings are considered publications). Likewise, it is certainly an ethical breach when two journal articles present the exact same data without acknowledgment of the earlier publication in the latter publication. The issue is not as clear between these extremes. Is it duplicate publication when two publications use the same data but conduct different analyses? What about publications that re-present the first wave of a longitudinal data collection, already published, in the article presenting the results of the second wave of data collection? An extended discussion of these issues is beyond our scope here, but it is important to make two points. Authors are ethically obligated to make readers (especially the reviewers who will judge the article for its original substantive contribution) aware when they re-present already-published data in subsequent articles. Meta-analysts are ethically obligated to alert the journals involved when they uncover what they consider to be duplicate publication.

Conclusion

We began this chapter by suggesting, somewhat provocatively, that meta-analysts had it easy relative to primary researcher when it came to the ethical considerations surrounding their work. If readers did not view this assertion with skepticism when we first made it, we hope they do now. Ethical issues surrounding the reporting of methods and results are as, if not more, complex for meta-analysts than for primary researchers.

specimens, the exemption that is most frequently relevant is the exemption under HHS regulations at 45 CFR 46.101(b)(4): 'Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects' " (http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm).

Respecting the rights of research participants remains an issue for those meta-analysts using IPD. And more often than primary researchers, meta-analysts may find themselves in a position in which they must report ethical lapses on the part of others in the research community.

Finally, because meta-analysis is a relatively new technique, the standards of reporting are still evolving. The ambiguities in what is and is not important for readers to know make reporting decisions more difficult for research synthesists. We hope that the results of our survey and discussion of other ethical issues will make research synthesists' decisions a bit easier. Perhaps, the broad lesson is that ethical decisions in research—be it primary, secondary, or research synthesis—are never easy and never to be taken lightly.[8]

References

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. New York: Author.

American Psychological Association. (2010). *Publication manual* (6th ed.). Washington, DC: Author.

APA Publication and Communication Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63,* 839–851.

Brown, B. L., & Hedges, D. (2009). Use and misuse of quantitative methods: Data collection, calculation, and presentation. In D. M. Mertens & P. E. Ginsberg (Eds.), *The handbook of social research ethics*. Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). New York: Academic Press.

Cook, D. J., Sackett, D. L., & Spitzer, W. O. (1995). Methodologic guidelines for systematic reviews of randomized control trails in health care from the Potsdam consultation on meta-analysis. *Journal of Clinical Epidemiology, 48,* 167–171.

Cooper, H., DeNeve, K., & Charlton, K. (2001). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods, 2,* 447–452.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis using individual participant data and aggregate data. *Psychological Methods, 14,* 165–176.

[8] Author note: The authors thank the members of the Society for Research Synthesis Methodology for their participation in the survey reported in this chapter. Correspondence can be sent to Harris Cooper, Department of Psychology & Neuroscience, Box 90086, Duke University, Durham, NC 27708-0086, or cooperh@duke.edu

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Gleser, L. J., & Olkin, I. (2009). 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 Foundation.

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

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

Merriam Webster Online. (2007). *Merriam Webster online*. Retrieved from http://www.merriam-webster.com/dictionary

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D., for the QUOROM group. (1999). Improving the quality of reporting of meta-analysis of randomized controlled trials: the QUOROM statement. *Lancet, 354,* 1896–1900.

Moher, D., Tetzlaff, J., Liberati, A., Altman, D. G., & the PRISMA group. (2009). Preferred reporting items for systematic reviews and meta-analysis: The PRISMA statement. *PLoS Medicine, 6*(7): e1000097.

National Institutes of Health. (2003). Final NIH statement on sharing research data. Retrieved from http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Oxford, UK: Wiley.

Schmidt, F. L., On, I., & Hayes, T. (2009). Fixed vs. random models in meta-analysis: Model properties and comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62,* 97–128.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., … Thancker, S. B. (2000). Meta-analysis of observational studies in epidemiology. *Journal of the American Medical Association, 283,* 2008–2012.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods, 13,* 130–149.