

Meta-analysis: integrating accumulated knowledge

Dhruv Grewal¹ · Nancy Puccinelli^{2,3} · Kent B. Monroe⁴

Received: 26 July 2017 / Accepted: 16 October 2017 / Published online: 7 November 2017
© Academy of Marketing Science 2017

Abstract Building a foundation of marketing theory requires developing effective ways to aggregate research results. Meta-analyses that accumulate knowledge within a research domain is an important means for summarizing research findings and increasingly is being conducted in various substantive marketing domains. Moderator analysis and structural models using meta-analytic inputs have emerged as a powerful means to advance current knowledge in a research domain, and, importantly, identify fruitful areas for future inquiry. This article reviews the growth of meta-analysis in marketing and identifies several important issues researchers must consider when conducting and reporting a meta-analysis.

Keywords Meta-analysis · Integrative review

Mark Houston served as Area Editor for this article.

The authors appreciate the feedback and insights shared by Michael Borenstein. They also appreciate the helpful comments of Gopalkrishnan Iyer. The authors also appreciate the feedback provided by the editor, AE and the reviewers.

✉ Dhruv Grewal
dgrewal@babson.edu

Nancy Puccinelli
N.M.Puccinelli@Bath.ac.uk

Kent B. Monroe
kbmonroe@illinois.edu; kent@kentmonroe.net

- ¹ Babson College, 213 Malloy Hall, Babson Park, MA 02457, USA
- ² University of Bath, Bath, UK
- ³ Oxford University, Oxford, UK
- ⁴ Department of Business Administration, University of Illinois, 4905 Old Millrace Place, Glen Allen, VA 23059, USA

Building a store of accumulated knowledge is critical for knowledge development in any field. The need for knowledge accumulation is important for the advancement of scientific understanding through the integration of key findings in a specific research domain. Meta-analysis is a rigorous alternative for making sense of a rapidly expanding research literature (Glass 1976). Meta-analysis can identify the expanding boundaries of a research domain by summarizing current knowledge and important unresolved conceptual, methodological, and substantive issues. Such reviews highlight empirical generalizations and draw attention to the implications of these insights both for academia (within and beyond marketing) and for practice.

Palmatier et al. (2017) discuss the importance and need for review papers. While narrative reviews can summarize collections of studies, more sophisticated meta-analytic methods for synthesizing knowledge can be used. There are benefits of a meta-analysis: it tests the robustness of a finding, helps resolve apparent conflicting findings, identifies research design issues, and suggests appropriate designs for future studies. Another important benefit of meta-analysis is that it is a way to compare and combine results across studies. It helps to determine consistency of results while at the same time explain variations in observed effects. To establish the boundaries of knowledge in a research domain, it is important to determine what we know and, more importantly, what we do *not* know.

One of the limitations of the marketing discipline is a tendency to rely on a single or few studies to indicate the current state of knowledge in a research domain. A single study is a sample of one and can rarely, if ever, provide sufficient evidence to resolve a research question (Wells 2001). Indeed, we need a very large sample size of results to obtain a correlation between two variables to be correct to two digits. Meta-analysis is the best form of literature

review to provide such a database of results for a research domain (Hunter 2001).

The aggregation of studies in a research domain produces accumulated knowledge by (1) helping to develop and test the theoretical bases and underlying predictions and (2) assessing the empirical evidence for a specific relationship across multiple studies. These tests and assessments establish the “truths,” or empirical generalizations, within a field. These “truths” in turn refine the theory and fuel further empirical efforts.

Extracting knowledge from a vast literature is a complex and important methodological problem (Glass 1976). There is a need to quantitatively integrate research in various marketing domains. There is also a need to understand the magnitude of the effect that drivers and mediators have on outcomes in these domains. The objective is to better understand specific questions or problems (e.g., does X Influence Y?) and to examine how specific characteristics of the studies (e.g., type of manipulation of X, or measure of Y) in that domain influence the variation in results across the studies.

The popularity of meta-analysis is growing within various sub-fields of marketing such as consumer behavior (Scheibehenne et al. 2010), communications (Eisend and Küster 2011), sales (Verbeke et al. 2011), and product management (Rubera and Kirca 2012). The meta-analysis studies in these domains also provide insights on the effect of numerous methodological characteristics, such as type of sample, experimental versus survey methods, variable operationalizations, and construct measurement that influence the observed effects.

Marketing as a discipline has developed to a point that researchers have even conducted a meta-analysis of meta-analyses. Eisend (2015) conducted a meta-analysis of 176 meta-analyses in marketing, finding the average effect size, $r = .24$, was moderate but higher than estimated in several other disciplines. This advent of a second-order meta-analysis is a sign that research in marketing is reaching maturity and as such should be celebrated.

Geyskens et al. (2009) reviewed 69 management meta-analysis studies that were published during 1980–2007. They offer several very important suggestions for future meta-analytic papers regarding important decisions and the trade-offs that need to be considered. Similarly, we examine meta-analytic papers in the marketing domain. Although we include several topics overlapping their paper, we discuss additional important meta-analytic issues, such as meta-analysis structural equation models and using the Binomial Effect Size Display (BESD) to consider the practical implications of the results of a meta-analysis.

As mentioned, the quantity of meta-analysis articles in marketing has increased over the past several decades. Marketing researchers have become increasingly sophisticated in conducting meta-analyses, such that the number of studies

synthesized has grown (22 in Assmus et al. 1984; 2105 in Verbeke et al. 2011). Researchers are also using more tools to assess the state of knowledge in various research domains. Nevertheless, as summarized in this article, there is a plethora of techniques and procedures that have been used. To establish validity of future meta-analyses, there is a need to establish consistency in the conduct of these reviews. That is, there is a need for an explicit review methodology. The consequences of a failure to have an explicit review methodology:

- There is an implication that little thought has been given to this issue and that sometimes less powerful review methods might be used.
- There are no standards for judging the quality of the meta-analyses.
- It is difficult to train graduate students to do quality reviews.
- It hinders the synthesis of knowledge from previous research.

The goal of our paper is to make three contributions. First, we would like to spur the accumulation and aggregation of knowledge in the marketing field, irrespective of the substantive, methodological, or conceptual domain. Through a carefully designed and executed meta-analysis, the “state of art” of an area can be readily ascertained by researchers, and they can understand the unresolved issues. Even more important, these meta-analyses can be an important basis for building theory in a variety of domains and/or testing alternative competing theories.

Second, we review 74 meta-analyses in leading marketing journals. This review is not meant to be exhaustive, as numerous marketing meta-analyses have been conducted and reported in other journals. We summarize these meta-analyses on several important dimensions including the types of methodological and reporting choices researchers have made in these reviews.

Finally, we articulate the various trade-offs that meta-analytic researchers face as they proceed through the various steps of conducting a meta-analysis. Furthermore, we offer “considerations” that researchers should follow, beginning with the important step of defining the research domain that the review and synthesis will cover and going through the many steps of data gathering, analytical procedures, and reporting the review in the final paper or article. Most of these considerations can be found in various books and meta-analytic review articles in other fields, but there is a need to provide them in one published source in the marketing discipline if for no other reason than to facilitate the training of future meta-analytic researchers and reviewers of such papers.

We briefly outline the advantages of using meta-analysis when doing literature reviews. Next, we summarize important meta-analytic issues, such as sample and search strategy, the

role of effect sizes, homogeneity of these effects, selection of an appropriate model for synthesis, testing important moderators of the effect, and understanding the mediating mechanisms using meta-analytic structural modeling procedures.

The meta-analytic advantage

Meta-analysis began to replace the narrative review in the late twentieth century, such that it is the dominant research synthesis tool in many fields. To aggregate evidence from a growing body of research, meta-analysis offers numerous advantages to researchers. Meta-analyses leverage the advantages of effect size estimates for summarizing results (Fern and Monroe 1996). With this summary effect size estimate (effect size for short), researchers can synthesize a set of studies addressing the same fundamental relationship. When these effects are consistent, meta-analysis can attest to the overall robustness of an effect. If the objective of the meta-analysis is to quantitatively combine findings from multiple studies, then standard meta-analytic methods as outlined in most meta-analytic books can be used (e.g., Borenstein et al. 2009; Glass et al. 1981; Hedges and Olkin 1985; Hunter et al. 1982; Lipsey and Wilson 2001; Rosenthal 1984).

Because effect sizes differ across studies, there will be a distribution of effect sizes as opposed to a single value that is reproduced with each additional study. Meta-analysis can leverage the study-level differences to explain variations in effect sizes and in turn assess the research domain. These study-level differences, or moderators, play a role akin to moderators in other research settings. For example, meta-analysis can indicate the strength of the effect of transaction cost economics on national governance decisions is moderated by the country's cultural values (Steenkamp and Geyskens 2012). Typically, if the objective is to examine numerous moderators simultaneously as outlined below, a researcher should use meta-regression procedures (Hierarchical linear meta-regression or HiLMA).

By combining a consistent set of studies, meta-analysis also can inform how one measure of a construct might differ from another. For example, if the objective is to determine the effect of selling-related knowledge on sales performance, we could review studies that measure self- and managerially reported sales performance. Then we could determine if selling-related knowledge has a stronger effect on self-reported or managerial assessments of salespeople's performance (Verbeke et al. 2011). If instead the goal is to estimate selling-related knowledge's effect on managerial assessments, we would examine just those studies using this measure of performance in our meta-analysis.

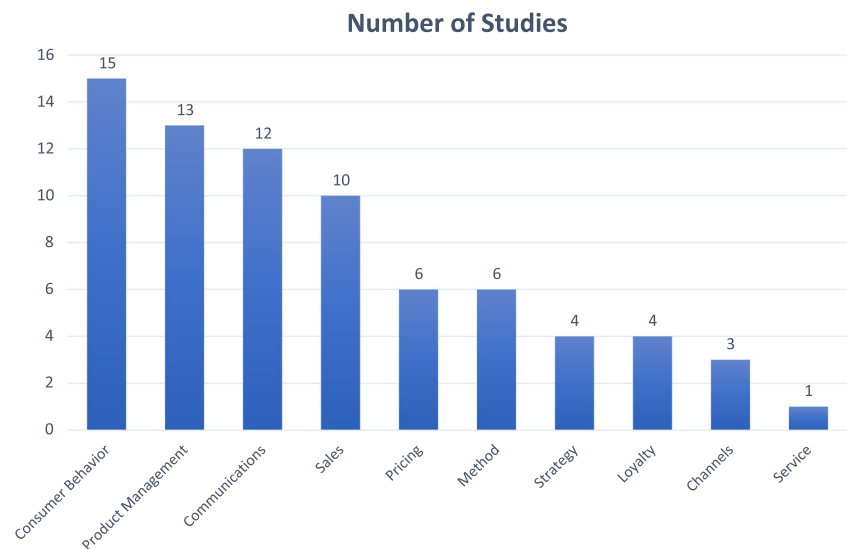
Domain of the meta-analyses reviewed

Appendix 1 (Table 5) provides a summary of 74 meta-analyses in leading marketing journals from 1981 to 2017. This summary is not exhaustive, but it illustrates meta-analyses that have been published in leading marketing journals (e.g., *Journal of Marketing*, *Journal of the Academy of Marketing Science*, *Journal of Consumer Research*, and *Journal of Marketing Research*). Some of the first meta-analyses were published in 1985 (Churchill et al. 1985; Peterson et al. 1985), and Monroe and Krishnan (1983) illustrated a procedure for integrating research outcomes across studies. While there has been consistent use of meta-analysis within marketing since that time, there has been a marked increase in popularity since 2000.

These 74 meta-analyses were classified into nine substantive areas (Fig. 1). It is important to note the robust number of marketing meta-analyses and generalizations that have appeared in consumer behavior, product management, communications, and sales. It is likely that our review underrepresents some of the other marketing areas. Meta-analyses on sales likely will also appear in *Journal of Personal Selling & Sales Management* and meta-analyses on communications in *Journal of Advertising* and *Journal of Advertising Research*. Also, retailing meta-analyses are published in *Journal of Retailing* and service meta-analyses in *Journal of Service Research*.

The domains of strategy, channels, and retailing have garnered considerable research over the last several decades, and meta-analyses on these topics have been published in other journals. It is important for meta-analyses in these domains to also be targeted for publication in our leading marketing journals. Over time, many of the meta-analyses have been influential in spurring additional research in these domains as evidenced by the high citations using web of science. Some of the most cited marketing meta-analyses include: Sheppard et al. (1988) on the theory of reasoned action (over 1100 citations), Palmatier et al. (2006) on relationship marketing (over 500 citations), Szymanski and Henard (2001) on customer satisfaction (over 450 citations), and Henard and Szymanski (2001) on new product success (over 450 citations).

Clearly, meta-analyses tend to provide a broad overview of the state of research in a chosen domain, answer many questions, lay to rest several conflicts, and bring to light new conflicts and areas of needed inquiry. Reviewing these meta-analyses highlights the need for meta-analyses in other domains that have generated considerable research attention. Some research domains that come to mind include the growing literature on mobile promotions and their consumer behavior implications, the domains of customer experience management, service recovery strategies, affect management, and so forth.

Fig. 1 Domain of meta-analyses

Types of meta-analyses

The review of the various meta-analyses in marketing (Appendix 1 (Table 5)) highlights some inherent differences in the types of meta-analyses. Table 1 lists three different types of meta-analysis and how many of each were in the set we reviewed.

Type 1 or standard meta-analysis

Fifty-two of the meta-analyses integrated effects across a domain or across a set of relationships using standard recommended meta-analytic techniques (e.g., Borenstein et al. 2009; Rosenthal 1984). The specific key variable is an effect size, and these meta-analyses either broadly integrate multiple relationships or do a focused analysis on a set of relationships. They examine the role of multiple moderators either individually and/or simultaneously. These moderators include method factors and important conceptual factors that could resolve apparent inconsistencies in the literature. Type 1 meta-

analyses have also used structural equation modeling and some have even tested alternative models.

Meta-analyses that tend to integrate research domains primarily using survey methodologies (e.g., relationship marketing, channels, and service quality) tend to conduct meta-analysis structural equation modeling (MASEM) as they have better access to correlations between all the constructs. On the other hand, meta-analyses in domains predominantly using experiments (e.g., comparative advertising, regulatory fit) focus on main effects and moderators.

Type 2 or replication analyses

Replication analyses (Farley et al. 1981) do not necessarily follow traditional meta-analytic procedures. However, they use some key measure (or measures) from the studies being integrated. Twenty meta-analyses are in this category. For example, Assmus et al. (1984) examined studies pertaining to advertising effects on sales and analyzed estimated parameters from 128 models. The dependent measures in their ANOVAs

Table 1 Types of meta-analysis

Meta-Type	Label	Description	Frequency*	Illustrations
Type 1	Traditional Meta-Analysis	Meta-analysis of effect sizes using standard meta-analytic procedures	52	Churchill et al. (1985) Fern et al. (1986) Brown and Peterson (1993) Franke and Park (2006) Scheer et al. (2015)
Type 2	Replication Analysis	Analysis of key statistics from past studies in a domain to explore the role of certain factors	20	Farley et al. (1981) Sultan et al. (1990) Keller and Lehmann (2008)
Type 3	Second Order Meta-Analysis	Meta-analysis of results of past meta-analyses to explore the role of certain factors	2	Peterson (2001) Eisend (2015)

* Based on the meta-analyses summarized in Appendix 1

were short-term elasticity, carryover coefficient and goodness of fit. (Measures of goodness of fit are not effect sizes.)

Similarly, Sultan et al. (1990) examined diffusion models considering factors that influence coefficient of innovation and coefficient of imitation across 213 applications. Several others use similar methodologies on a variety of topics. Farley et al. (1995) discuss the results of a number of such analyses, such as diffusion model (Sultan et al. 1990), buyer behavior (Farley et al. 1981) and price elasticities (Tellis 1988). They highlight the need for such meta-analyses insights “should replace the now discredited zero null hypotheses of such parameters in future work” (p. G36).

Type 3 or second-order meta-analysis

Two meta-analyses fit into this category. These analyses take effect sizes from published meta-analyses and examine the effects of other variables that might be influencing them, such as Peterson (2001) qualitatively and Eisend (2015) quantitatively. (Some of the meta-analytic procedures discussed in this article might not be pertinent for Type 3 meta-analyses.)

Key considerations for conducting a meta-analysis

Many key steps or considerations need to be followed as a researcher designs and conducts a meta-analysis. These steps include determining the research domain, identifying the central research question, specifying the sample, extracting the effect size from each study, choosing the type of model to apply, testing for heterogeneity of the effects, and identifying key moderators. Figure 2 provides a sample flow chart on how a meta-analytic researcher may proceed when conducting a Type 1 meta-analysis. Table 2 highlights some key considerations for this type of meta-analysis. It also highlights how

these choices have changed over the decades, in the 1980s, 1990s, 2000s, and 2010s.

Determining the research domain

As in any research endeavor, the first step in synthesizing research is to determine the research questions that will guide the conduct of the meta-analysis. The question may be relatively broad, such as: “Does the foot-in-the-door technique work?” (Fern et al. 1986). Or the question may be relatively narrow, such as: “Does the foot-in-the-door technique work if the multiple requests are not contiguous?”

An obvious start is to gain a thorough knowledge of the underlying theory on a topic. Another important source is previous qualitative reviews in the research area. Once a preliminary examination of the literature has been completed, the researcher may find that further refinement of the research question is appropriate. Finally, it is quite proper to search our own minds for ideas; that is, insight, intuition, and ingenuity might lead to a novel approach to a research domain (Campbell et al. 1982).

The research question may be substantive, such as: whether price influences consumers’ perceptions of quality (Rao and Monroe 1989); whether comparative ads are more effective than non-comparative ads (Grewal et al. 1997); whether objective performance is influenced more by relationship quality than commitment (Palmatier et al. 2006). Or the research domain may be conceptual in nature, such as the theory of reasoned action (Sheppard et al. 1988) or regulatory fit (Motyka et al. 2014). It could focus on methodological issues such as research design choices (Peter and Churchill 1986). The research questions will guide many of the meta-analysis decisions discussed in this article.

Establishing the underlying research question is very important, as it is probably the key source of variance in

Fig. 2 A potential meta-decision flow chart for Type 1 meta-analysis

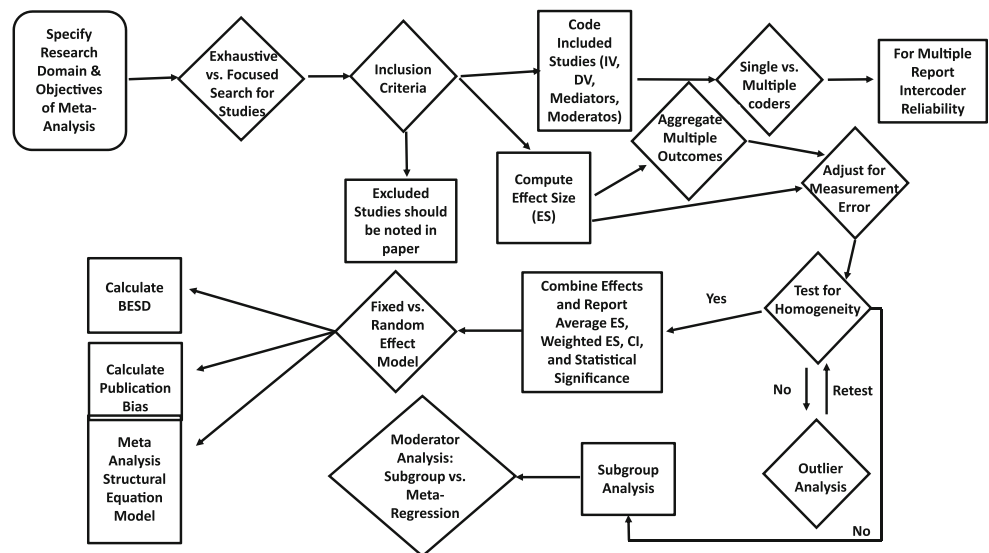


Table 2 Summary of meta-analytic choices for Type 1 meta-analyses

Meta-Analytic Choices	Levels	1980–1989 <i>N</i> = 8	1990–1999 <i>N</i> = 9	2000–2009 <i>N</i> = 15	2010–present <i>N</i> = 20	Total Frequency <i>N</i> = 52
Effect Size	<i>r</i> / η / ϕ	7	7	14	16	44
	<i>d</i>		1	3		4
	Odd-ratio			1	1	2
	Omega squared	1	1			2
Used Weighting	Yes	5	5	13	17	40
	No	3	4	2	3	12
ES Reliability Adjusted	Yes	1	4	12	16	33
	No	7	5	3	4	19
Test of Homogeneity	Yes	3	7	12	14	36
	No	5	2	3	6	16
Test of Publication Bias	Yes	1	3	9	13	26
	No	7	6	6	7	26
Assess Outliers	Yes	1	4	5	10	20
	No	7	5	10	10	32

Some of these choices may not be pertinent to some of the studies. For example, a meta-analysis from experiments looking at the effect of a manipulated independent variable on an objective measure would not likely need to adjust the effect size for reliability

conclusions across different reviews ostensibly examining the same question. It is very important to carefully articulate the operational definitions underlying the review as well as the operational detail when conducting the data gathering and analytical procedures. The validity of the conclusions of the review depends on both the conceptual definitions and the operational detail employed.

Specifying the search and sampling strategy

As in any empirical investigation, the relevant population of studies needs to be defined prior to searching for the articles/studies to include in the analysis. To locate an appropriate sample, reviewers likely will use one or more citation databases (e.g., ABI/INFORM, Proquest, Google Scholar, SSRN, EBSCO), review the bibliographies of prior reviews, and identify seminal articles. Journals oriented toward publishing reviews will prove especially useful. It is also helpful to send e-mails to leading scholars in the research domain. These experts likely are knowledgeable about research currently underway as well as unpublished work (e.g., dissertations or research presented at conferences). Finally, requests for articles posted on listservs, such as ELMAR and ACR, can help reviewers locate studies that otherwise would be difficult to find.

The search strategy adopted will affect the meta-analytic conclusions, simply because each study does not have the same probability of being selected. A well-established scholar who is very familiar with the research domain may have access to a more diverse body of research than a novice scholar relying on database searches. The researcher should engage in

as comprehensive a search as possible and include as many studies as possible (Cooper 1982). The more comprehensive the search strategy employed by the reviewers, the more generalizable the study results. It would also permit coding for a variety of publication variables (e.g., when it was published and the journal's SSCI impact factor it was published in) and assess whether these factors influence the size of an effect.

After determining the population of studies, the researcher must decide the sampling process to use, particularly if the research domain includes many studies. If a meta-analysis includes the entire population of studies, sampling is not an issue. However, most meta-analyses set some *inclusion criteria* that studies must meet to qualify for inclusion. For example, a meta-analysis of regulatory fit included only studies that manipulated or measured fit using a precise set of previously used tools (Motyka et al. 2014).

For example, if a meta-analysis examines the effect of too much choice on consumer responses, a pertinent inclusion criterion might be a study should manipulate the amount of choice, rather than measuring naturally occurring differences in the amount of choice (Scheibehenne et al. 2010). More narrow inclusion criteria are appropriate if the goal is to estimate a more specific effect; broader criteria are useful if the aim is to understand how other factors might influence an effect. For example, with a goal of accurately assessing the effect of the Events Reaction Questionnaire (a measure of regulatory focus) in terms of invoking regulatory fit, researchers would only include studies that rely on this measure. However, if their goal is to determine how other factors shape the effect of regulatory fit, the inclusion criteria should be broadened to include all studies that measure or manipulate

regulatory focus with any existing measures or stimuli (Motyka et al. 2014).

Meta-analytic researchers must explicitly report their inclusion/exclusion criteria as well as the time period the review covers so readers can assess the validity of the meta-analysis and design meaningful future research. For example, a meta-analysis of price perceived quality research prior to 1989 (Rao and Monroe 1989) was followed by a meta-analysis of price-perceived quality studies between 1989 and 2006 (Völckner and Hoffmann 2007). Validity issues pertain to whether the studies included are representative of the studies in the research domain. A second validity issue concerns whether the included studies provide a representative sample of “subjects,” research settings, research designs, and other methodological variables that may influence study results and eventual review conclusions.

Results reported in journals are more disposed toward the favored hypothesis than are findings reported in dissertations and theses. Moreover, statistically significant results are more likely to be published than non-significant results. Although some of these studies might be methodologically flawed, and may report questionable findings, no single study is perfect. It is difficult to determine reliably if a methodological flaw has compromised the findings. Before eliminating a study due to a suspicion the findings are methodologically flawed, the reviewer should determine whether variation in results across studies may be due to sampling error, measurement artifact, or theoretically plausible intervening variables. If these three sources cannot explain variance in results, a methodological flaw may be influencing results and would need to be addressed.

Finally, it would be useful to include studies with different designs. For example, to assess the effectiveness of relationship marketing (Palmatier et al. 2006), researchers might include a set of studies that examine different relationship elements (e.g., commitment, trust, relationship satisfaction, relationship quality). If there is some reason to believe these elements may differ in their effectiveness, the researcher should code for them, then test the relative effectiveness of each.

To establish validity, all integrative literature reviews, including meta-analyses, must report the search process adopted, sampling procedure, and criteria applied to exclude any studies. Given a thorough description of these procedures, the completeness and validity of the review can be judged. Future researchers will then be able to extend the review without having to duplicate it. In this article, many meta-analytic decisions are outlined. A sample checklist is provided in Table 3.

Coding the studies

In a meta-analytic review, the primary research studies provide the data for the analyses. To draw meaningful

conclusions, it is necessary to consider the many different characteristics of the individual studies that may be a source of variation in findings across studies. The objective is to relate the characteristics of studies to outcomes to isolate potential sources of variation in results across studies. A second objective is to quantify as much as possible the description of studies whether on a metric or non-metric basis.

Essentially, coding studies is a measurement issue. The validity issues here include the clarity of definitions, adequacy of the information provided in the original reports, the amount of inference the coder must make, the degree of coding detail. The reliability issue concerns the consistency of coding among coders and over time. It is important to standardize coding procedures and check (and correct) for inconsistency over coders and/or over time. It is important to use multiple coders and report the results of the reliability of the coders and how discrepancies were resolved. Careful planning, explicit instructions, and specific definitions should be provided at the outset. Moreover, a training period using a set of common studies will improve consistency and provide an early assessment of the extent of coder disagreement.

Even a careful review, if applied uncritically, may impede further research by producing an apparently clear result. By glossing over variations due to such differences as setting, type of respondents, measurement and instrumentation, operationalization of variables, range of treatment, and other study characteristics, such a review will be less likely to resolve conflicts among the different results. Meta-analytic researchers should capitalize on variations across studies to develop explanations for why a relationship may be significant in one study but not in another. Of importance when examining such variation in findings is there may be characteristics of the studies to help establish pertinent boundaries for the underlying phenomenon, such as methodological factors. It is very important for the researcher is to detail the characteristics of the studies examined (Pillemer and Light 1980).

The researcher must take the perspective of a detective and examine each study microscopically. To facilitate the process, a coding form should be developed reflecting the nature of the research to be examined and possible sources of variation in results. In many respects, the final selection of the sample of studies included in the meta-analysis might usefully be postponed until the coding process has been completed. The coding process may reveal several key study differences. Therefore, the reviewer may want to sample from different strata of studies to create a more representative sample.

There are two objectives to be accomplished in the coding process. First, as observed above, it is desirable to relate the characteristics of the studies to the study findings. A second objective is to quantify as much as possible the description of studies whether on a metric or non-metric basis. To accomplish these objectives, thought and care in the definition of the attributes of studies and their quantification is required.

Table 3 Potential criteria that should be reported

Meta-Analytic Trade-off	Explicitly Report the Criteria Used
Sample domain	<ul style="list-style-type: none"> • Make explicit what is the domain and what studies are not being included and why
Search strategy	<ul style="list-style-type: none"> • Provide details about how papers were obtained and what efforts were made to obtain unpublished papers
Effect size	<ul style="list-style-type: none"> • What is the effect size measure being used? r, d, etc. • How was it calculated when not reported? • What weighting method was used (if used)?
Correction for measurement error	<ul style="list-style-type: none"> • Was the reliability statistics available and were they used to correct for measurement error? • Report any other corrections, such as adjusting for range restrictions and dichotomization of variables.
How were multiple outcomes treated?	<ul style="list-style-type: none"> • Were they treated as multiple observations, or a composite created? • If a composite was created, how was it done? • If only one was retained, how was that choice made?
Was BESD reported?	<ul style="list-style-type: none"> • BESD can be reported to increase substantive implications of results.
Are the effect sizes homogenous?	<ul style="list-style-type: none"> • Is one of more tests for homogeneity reported?
Was outlier analysis conducted?	<ul style="list-style-type: none"> • Report outlier analysis and what method was used. • Also, was analysis done with and without outliers?
Was a fixed-effect model or a random-effects model used?	<ul style="list-style-type: none"> • Report the type of model used and why.
Was publication bias tested?	<ul style="list-style-type: none"> • What statistic was reported?
Was the role of moderators examined?	<ul style="list-style-type: none"> • Was sub-group analysis conducted? • Was the effect of the various moderators assessed using meta-regression procedures?
Was a structural meta-model reported?	<ul style="list-style-type: none"> • How were the effect sizes for each correlation calculated? • How was the N used for the model obtained?

Reviewers who critically examine the detail of each study likely will produce more valid conclusions because they will have more information about contextual variations that may have influenced the results across studies.

Of the reviews listed in the Appendix, 29 out of 74 do not report whether they used multiple coders to code the various factors. Several others noted using one coder. Best practice would be to use multiple coders and report the results of the reliability of the coders and how discrepancies were resolved. Careful planning, explicit instructions, and specific definitions should be provided at the outset. Moreover, a training period using a set of common studies will improve consistency and provide an early assessment of the extent of coder disagreement.

An abbreviated version of the coding guide used by Motyka et al. (2014) in their meta-analysis of the regulatory fit literature is displayed in Table 4. This coding guide indicates how promotion and prevention fit were coded (i.e., the independent variable) and how the three dependent variables (evaluation, behavioral intentions, and behaviors) were coded. The coding guide also provides definitions and examples for a conceptual moderator and a methodological moderator.

Extracting the effect size

Behavioral research usually relies on statistical significance tests to draw inferences. Underpowered hypotheses tests often cannot rule out a type-II error, so even if a researcher has discovered a viable relationship, tests based on small sample

sizes may not achieve statistical significance. Many studies in behavioral science report effect sizes, defined as an estimate of the difference across groups, independent of sample size (Borenstein et al. 2009; Fern and Monroe 1996). This standardized measure indicates both the direction and size of an effect associated with a relationship of interest. To the extent possible, meta-analytic researchers should calculate the effect size of an empirical result from the original report of the study. Usually, the data needed to calculate effect sizes are provided or can be inferred.

In an ideal world, researchers would have access to the raw data from all the publications about a phenomenon and could combine those raw data. That is rarely the case though, so there are various options for capturing the size of an effect (see Fern and Monroe 1996 for a detailed discussion on alternative effect size indicators and how to convert from one indicator to another). In practice, the reviewer calculates an effect size for each individual study and then compares the effect sizes before synthesizing these results. In a pinch, p -values are informative (Rosenthal 1984), but in some studies, these values refer to an effect that reaches some sort of threshold (e.g., $p < .05$), so combining them may not offer granularity or specify the actual magnitude of an effect. Furthermore, for p -values less than .001, studies generally do not report any other information about the size of the effect.

Gathering and selecting the effect size measure requires consideration of the kinds of data available in the domain of the meta-analysis. Studies reporting the differences between two groups (experimental and control conditions), tend to

Table 4 Sample coding guide

<i>Promotion and Prevention Fit Coding</i>	
<i>Construct</i>	<i>Definition</i>
<i>Promotion Fit</i>	Refers to regulatory fit conditions that match an individual with a primarily promotion focus with approach-related stimuli (compared to a nonfit condition in which individuals with a promotion focus are exposed to avoidance-related stimuli.
<i>Prevention Fit</i>	Refers to regulatory fit conditions that match an individual with a primarily promotion focus with avoidance-related stimuli (compared to a nonfit condition in which individuals with a prevention focus are exposed to approach-related stimuli.
<i>Dependent Variable Coding</i>	
<i>Construct</i>	<i>Definition</i>
<i>Evaluation</i>	Refers to dependent measures that assess participants' attitudes or reactions to a product, a message or other stimuli.
<i>Behavioral Intention</i>	Refers to dependent measures that assess participants' willingness to engage in a behavior.
<i>Behavior</i>	Refers to dependent measures that capture participants' actual behavior.
<i>Moderator Variable Coding</i>	
<i>Source of Regulatory Focus</i>	
<i>Construct</i>	<i>Definition</i>
<i>Chronic</i>	Participants' predisposition toward a focus on positive or negative outcomes.
<i>Self-Generated</i>	Momentary regulatory focus is primed by having participants reflect on something that is of or related to the self.
<i>Situation-Generated</i>	Momentary regulatory focus is primed by having participants reflect on something that is external to the self.
<i>Type of Subject</i>	
<i>Construct</i>	<i>Definition</i>
<i>Student</i>	Refers to studies in which participants are listed as undergraduates or students aged 22 or younger.
<i>Non-Student</i>	Refers to studies in which any other participant type was used.
<i>Example</i>	After determining a participant's orientation toward achieving positive outcomes through an individual difference measure, showing them a message that focuses on the positive outcomes that will be provided by product use.
<i>Example</i>	After determining a participant's orientation toward avoiding negative outcomes through an individual difference measure, showing them a message that focuses on the negative outcomes that will be avoided by product use.
<i>Example</i>	Asking participants to indicate liking of a brand on a 7-point Likert scale.
<i>Example</i>	Having participants report one's willingness to pay for a product.
<i>Example</i>	Observing whether participants choose more of a given product.
<i>Example</i>	Regulatory focus is captured via standardized measures.
<i>Example</i>	Participants are asked to reflect on their own lives or imagine themselves in a situation as part of the prime.
<i>Example</i>	Participants are exposed to stimuli or a context that primes a given regulatory focus.
<i>Example</i>	Having students participate for course credit.
<i>Example</i>	Using a community sample that might recruited through a local newspaper.

This table is an abbreviated version of the web-appendix associated with Motyka et al. (2014) meta-analysis

report t-values and F-values, from which an effect size indicator (e.g., eta) can be computed. Frequently, the meta-analytic researcher might find that all relevant information is not reported in the paper. In these cases, the researcher should contact the original authors for the information if they still have the information. If they do not, the choice is whether to use some analogy to estimate the information or not include that study in the meta-analysis. It might be better to code these effects separately and test whether the effect sizes estimated from partial information (e.g., df are not reported and need to be determined from other information) differ from those computed with complete information.

Survey-based researchers in the domain of sales, organizational behavior, and strategy tend to provide correlation matrices. A correlation itself can serve as an effect size. Numerous different effect sizes can be computed and analyzed. Fern and Monroe (1996) provide a number of different effect size measures, ranging from correlational effect sizes, to standardized mean difference effect sizes, to explained variance effect sizes. The review of Type I meta-analyses indicates 44 out of 52 meta-analyses used a correlation as an effect size, and this pattern is similar across time. It must be noted that some meta-analyses conduct their analysis using the r s whereas others use the Fisher r to z transformation.

In certain regression-based studies, if the correlation matrix is not provided it may be necessary to choose either using beta coefficients (partial coefficients) or not using the data. Peterson and Brown (2005) provide a procedure to impute the correlation from these partial coefficients. Reviewers should test whether the effects differ for the average imputed r as compared to the average correlation based r .

Correcting for measurement error If constructs can be captured by a single objective measure (e.g., sales), there is no need to correct for measurement error. But in marketing, many constructs usually require multi-item measures containing some measurement error. As measurement error leads to understated estimates of an effect size (i.e., with a less reliable multi-item measure, the effect sizes will be smaller). Whenever possible researchers should correct for this error. If measure reliability information is available, researchers can use it to adjust for measurement error (e.g., Palmatier et al. 2006). The following formula is useful (Hunter and Schmidt 2004): $r_c = r_{xy} / (\sqrt{r_{xx}} * \sqrt{r_{yy}})$, where r_{xx} and r_{yy} represent the measurement reliabilities of variables x and y , respectively (r_{xy} equal to $r_{contrast}$ and r_{xx} equal to 1).

Researchers can correct for other systematic errors, such as range restrictions in either variable and/or dichotomizing of a continuous variable (Geyskens et al. 2009). Table 2 indicates that 33 of 52 meta-analyses use effect sizes adjusted for reliability. Adjusting for measurement error has occurred more in recent meta-analyses. This technique can result in correlation effect sizes greater than one, so researchers might consider

capping it at 1.0. The effect size estimates should be reported with and without these corrections.

Handling multiple outcomes from single studies Another issue is when it is possible to obtain multiple effect size estimates from each study. Should these estimates be considered independent, or should they be aggregated at the study level, such that only one result contributes to the total synthesis? If the study can be separated into conceptually equivalent but statistically independent replications, each result should enter the analysis separately, such as when a study examines more than one outcome (e.g., Orsingher et al. 2010).

If multiple indicators are used to estimate a relationship between the independent and dependent variables, including the effect size of each relationship might violate the independence assumption of the statistical procedures. If so it would be better to use an average effect size, weighted by the simple sample size (Hunter et al. 1982). For a more extensive discussion of the interdependence of effect sizes and the options available to researchers see a review by Geyskens et al. (2009). They illustrate different procedures, such as using Hunter and Schmidt's (1990) formula, averaging if conceptually equivalent, or randomly selecting a given outcome. Each procedure has advantages and disadvantages. For example, random selection is likely prone to researcher bias.

Testing for homogeneity of the effect sizes To the extent the results are quantified using standardized metrics (effect sizes), studies can be compared and combined to test hypotheses about the underlying research domain. Before combining or synthesizing results, the studies should be tested for the homogeneity of results. If the results vary significantly, it may be because of the quality of the methodology, sampling error, and/or measurement error. A test for homogeneity helps the researcher establish these results (or effect sizes) come from the same underlying distribution of results. If not, the effect sizes should be separated into homogeneous subgroups (see moderator analysis section).

These comparisons might rely on p -values or effect sizes. If studies fail to report the effect size or fail to provide the necessary information to compute effect sizes, p -values provide a viable option. Assuming statistical non-significance for any findings without a corresponding p -value, equal to $p = 0.50$, and transforming any result with statistical significance of less than .01 to be equivalent to a p -value of .01, researchers can determine the standard normal deviate Z for each exact p -value with the same directional sign. All the p -values must be one-tailed. The equation for the statistical significance test of the heterogeneity of Z (Rosenthal 1982) is:

$$\sum_{j=1}^n (Z_j - \text{Mean } Z)^2 \text{ distributed as } \chi^2 \text{ with } N-1 \text{ df.} \quad (1)$$

A preferable test for the statistical significance of the homogeneity of results requires transforming each correlation to

an associated Fisher z_r and conducting a chi-square test (Rosenthal 1982). The population value of r typically is not zero and the distribution of sampled r s becomes skewed particularly as the population r moves away from zero. The Fisher r to z transformation is distributed more normally, but it increases as the size of the correlations increase, although there is little bias until $r > .3$.

$$\sum(N_j-3)(Z_{rj}-\text{Mean } z_r)^2 \text{ distributed as } \chi^2 \text{ with } K-1 \text{ df.} \quad (2)$$

Where z_{rj} is the transformed r and $\text{Mean } z_r = \sum(N_j-3)z_{rj} / \sum(N_j-3)$.

If the homogeneity of results hypothesis is rejected then the results should be partitioned in appropriate subgroups that are each consistent in their degree of association. A non-significant test of homogeneity means the sample distribution of results come from the same population of results and the results may be combined.

Cochran's Q may be used as a measure of heterogeneity and is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies, with the weights being those used in the pooling method (Borenstein et al. 2009; Cochran 1950). Another statistic is the I^2 statistic, indicating the percentage of variation across studies being examined that is accounted for by heterogeneity as opposed to chance (Higgins and Thompson 2002; Higgins et al. 2003). Geyskens et al. (2009) discuss several other homogeneity tests suggesting use of multiple tests may be advantageous. When the output from these statistical tests suggests heterogeneity, researchers should examine the data for outliers. If the heterogeneity persists despite the removal of outliers, the effects of possible moderators (theoretical and methodological) should be assessed.

Outliers To address some of these heterogeneity issues, a meta-analysis should test for any powerful outliers and ensure the results are indeed robust. If such a test reveals the meta-analytic conclusions would change if a study were dropped, it requires careful consideration by the researcher (e.g., Compeau and Grewal 1998). For example, researchers might report the results both with and without the outlier included in the analysis. The researcher also must realize that because a study effect is an outlier, it does not make the study automatically inaccurate or incorrect. The analysis excluding it provides additional context as to whether the results might be dependent on that outlier. In Table 2, 20 out of 52 studies (around 40%) explicitly report outlier analysis.

More sophisticated outlier analysis techniques are available. Huffcutt and Arthur's (1995) sample-adjusted meta-analytic deviancy (SAMD) procedure takes the sample size into account as it identifies potential outliers. Geyskens et al. (2009, p.400) note that SAMD is computed using "the

difference between each primary study's effect size and the mean sample-weighted effect size (with the latter value not including the former value); then, it adjusts that difference for the sample size of the study." Chang and Taylor (2016) in their meta-analysis of customer participation in new product development use it to identify potential outliers and demonstrate robustness of their results with and without outliers. It must be noted that Beal et al. (2002), using monte-carlo simulations, provide a caveat to SAMD, in that it tends to over identify small correlations as outliers.

Combining effect sizes Meta-analysis can leverage the strengths of the effect size by combining effect sizes from multiple primary studies. For example, three independent studies of the same effect might not provide statistically significant results individually, but combining them reveals that the effect is significant. This combined effect size also provides a more accurate estimate of the size of the effect in the real world. Imagine, for example, three studies each with a sample size of 60 respondents, and each failing to find a statistically significant effect of emotion on judgment. The effect size for each study hovers around $r = .15$, a respectable, though low effect size. Combining the standardized effect sizes increases statistical power by using 180 participants when determining the statistical significance of the effect.

Sometimes, a meta-analysis can call conventional wisdom into question (Vadillo et al. 2016). In work on ego depletion, many had thought the depletion of glucose was the reason an act of self-control reduced subsequent self-control. Vadillo and his colleagues used meta-analysis to aggregate the work testing this relationship and concluded the empirical evidence does not support this conclusion. Finally, meta-analysis integrates the effects of different studies that may have used different measures, so that it provides a means to assess the underlying construct with a single, standardized metric.

Combining the effect sizes across studies yields the combined or average effect. Various guidelines exist for choosing and computing effect sizes (e.g., Borenstein et al. 2009), and available software calculate these sizes automatically, when researchers enter summary data. The average effect size can be a simple average effect size or a weighted average effect size, where the weighting mechanism is based on sample-size or variance (see Geyskens et al. 2009 for the advantages and disadvantages). As shown in Table 1, 40 of the 52 standard meta-analyses report using a weighting mechanism.

Testing the main effect relationships

In most meta-analyses, the researcher will investigate the main effect of the different independent variables on the various dependent variables. Typically, a researcher will report the

number of effects testing a given relationship, the total sample size for all the effects, the overall sample weighted average effect size (or sample-weighted reliability adjusted effect), the 95% confidence interval, homogeneity statistics and publication bias statistics. However, the researcher needs to decide whether to use a fixed-effect versus random-effects meta-analysis models. These two models ask fundamentally different questions, which could yield different answers.

Fixed-effect models If all the studies in the analysis are based on the same population of participants/procedures and are largely identical in material ways, a fixed-effect model is preferable. The common belief that model choices (fixed vs. random) should reflect the amount of heterogeneity in the data is incorrect; it must be based on the modelers' understanding of the sample frame. A fixed-effect model assumes all error is due to sampling error within studies. Here, the word "effect" is singular, because all studies share the same underlying "true" effect, and "fixed" indicates the chosen population has been specifically designated rather than sampled at random (Borenstein et al. 2009).

Moreover, because all studies in a fixed-effect model estimate a common parameter, the only source of error is the sampled subjects in the studies V_i . Each study in the meta-analysis is weighted by the inverse of the variance, such that the weight assigned to each study is $1/V_i$. To confirm a fixed-effect model is appropriate, a test checking whether the results are homogeneous should be conducted. This is a check to ensure that no sample-level factors (e.g., when the study was conducted) might have led to the different effects.

A fixed-effect model can only generalize to studies that are from the same underlying distribution of results because it assumes all studies in the analysis are estimating the same effect. In practice, imagine a retailer in New York wants to find out how much consumer purchase intentions might increase in response to a specific advertisement. A computer draws 20 sets of names with 100 consumers in each set; each set is equivalent to a single study. The consumers in each set view the advertisement and indicate their purchase intentions. Using the results from these 20 studies, we can compute the mean score and use a meta-analysis to synthesize the results, which provides the estimates for the mean values for all consumers visiting that store. A fixed-effect model is best here because all the studies in the model estimate the same effect. The estimate from each study might differ (e.g., due to sampling error), but the underlying actual estimate will be the same for all the samples. However, we cannot generalize what the effect would be for another store not in the sample or if we were to change the protocol for testing the advertisement.

Having determined the research question, a researcher also must address the suitability of the data set for answering that question. If the studies have been drawn from different samples, the researcher needs to test for differences in the variable of interest. For example, if the time when the test was administered might influence performance, some preliminary testing should ascertain such influences. When the differences are notable, the researcher can perform a test of heterogeneity (as was discussed earlier). If the test of heterogeneity is not significant, a fixed-effect model remains appropriate; if the test of heterogeneity reveals there are real differences among the samples, use of random-effects models is necessary.

When researchers conduct a meta-analysis, a relevant consideration is how many studies are needed to perform it. The answer again should be driven by the objective or goal of the meta-analysis. If the objective is to estimate a given effect more precisely, the meta-analysis would use a fixed-effect model, and such an analysis requires a minimum of two studies. Because all the studies estimate the same effect, increasing the number of studies will lead to more accurate estimates. This type of approach is beneficial for multi-study articles, which can confirm the robustness of an effect by quantifying it through a meta-analysis (Puccinelli et al. 2013). Certainly, the extent of an effect also can be determined by a confidence interval around the mean effect size. Further, to understand its robustness researchers can perform a file drawer assessment.

Random-effects models In meta-analyses of a prior literature, a random-effects model almost invariably fits the data better (Borenstein et al. 2010), because it recognizes each study estimates a unique parameter. Therefore, these models account for two sources of error: the sampling of respondents from that specific study's population, denoted V_i , and the sampling of populations from the universe of all relevant populations, or between-study variance, denoted T^2 . The total sampling error for any study thus is $V_i + T^2$, and the weight assigned to each study is $1/(V_i + T^2)$.

With random-effects models, the outcomes are estimates of both the mean effect size and the dispersion of the effects about the mean. The prediction interval addresses the extent of dispersion, by revealing the expected range of estimated effects (formulas for computing this interval are available from Borenstein et al. 2017). If the mean effect size is .50 and the prediction interval is .30–.70, the actual effect size in most populations of participants likely falls within this range. Other statistics pertain to more technical concerns (Borenstein et al. 2009), such as:

- Q , or the sum of squared deviations of all observed effects from the mean on a standardized scale.

- I^2 , which is the proportion of variance in observed effects due to variance in true effects, rather than sampling error.
- T^2 , the variance of true effects, where T is the standard deviation of true effects.

Overall, a random-effects model treats any collection of studies as a sample from a larger, hypothetical study. Borrowing from our previous example, imagine that a chain of department stores maintains locations all around New York City. We still want to know how consumers will respond to an advertisement. But in this case, we select 20 stores at random, located in various areas of New York, and from each store, we randomly select 100 consumers. Again, the source data consists of 20 studies, but in this case, each study refers to a different store, for which the responses are likely to vary, considering the variability in store location and characteristics (i.e., some locations are in up-scale shopping districts; others may be located next to value-oriented stores). Therefore, the term “effects” is plural for these models, because they sample from a universe of multiple effects; “random” acknowledges that the selection of these effects relies on random sampling.

As noted though, meta-analysis more commonly seeks to understand not just the overall effect but also sources of any heterogeneity in this effect. This issue leads to the adoption of a random-effects model to identify study-level factors that influence an effect. Narrative reviews often describe dispersion in the effect size as conflicting evidence, however, the effects might be consistent, if the populations of participants can be identified. For example, the effect may hold for student samples but not community samples, thus if a researcher identifies these two distinct populations, students and community participants, the dispersion in the distribution of effect sizes can be explained. If the goal of the meta-analysis is to test various moderators, then it would be necessary to have a larger number of effect sizes with at least two effects in each level of the moderator. Thus, when a researcher uses methods such as meta-regression (Hierarchical linear meta-regression or HiLMA) to simultaneously test all the moderators, the number of effects needed is much larger.

The application of a random-effects model requires a reasonably accurate estimate of between-study variance (T^2), and that estimate demands a reasonable number of studies. However, what is “reasonable” is a subjective assessment. If the studies tend to be very similar (e.g., using similar procedures, sample characteristics), it is likely that variation in an effect size is going to be smaller and the meta-analysis is likely to achieve an acceptably accurate estimate with a smaller set of studies. However, if studies in a research area tend to vary on numerous

important dimensions, the effect size likely will vary substantially, necessitating a larger number of studies to get a reasonably accurate estimate of the between-study variation. We deliberately avoid putting numbers on these descriptors, for several reasons. First, there are no established rules. Second, if insufficient studies are available, there are no good alternatives, though computational procedures can adjust the confidence interval to account for the uncertainty due to the small number of studies.

BESD and substantive implications for a combined effect size Researchers also can gain a sense of the substantive effect of an independent variable on a dependent variable using an average effect size. The binomial effect size display (BESD) provides a means to quantify real differences in outcomes between treatment and control groups (Rosenthal and Rubin 1979, 1982). A good example demonstrating the utility of effect sizes is the aspirin trial (Rosenthal and DiMatteo 2001). A seemingly modest effect size $r = .034$ indicates 34 out of every 1000 people would be saved from a heart attack if they took low dose aspirin regularly. Aspirin is safe and low cost; heart attacks can be devastating. Thus, low dose aspirin is now routinely recommended for at-risk people.

BESD can be used to estimate how many consumers will buy a product, depending on whether price information is presented in red versus black (Puccinelli et al. 2013). The researchers obtained an average effect size of $r = .48$. The BESD tells us that this effect size means that out of a hypothetical set of 200 men where half saw prices in red and half saw prices in black, 74 of the men seeing prices in red would evaluate the retailer more favorably while only 26 of the men seeing the prices in black would evaluate the retailer more favorably. That is, men will be 1.85 times as likely to judge a retailer favorably if prices appeared in red instead of black. Further, testing the efficacy of employing a foot-in-the door (FITD) multiple request strategy Fern et al. (1986) obtained across multiple studies an average effect size of $\phi = .125$. This result means the FITD strategy could improve a survey’s response rate by an additional 125 per 1000 respondents. Effect sizes can reveal, in standardized terms, the implication of an effect in real-world applications.

Publication bias Journals exhibit bias against studies that report non-significant statistical results (Greenwald 1975). These studies are less likely to be published and may not even be submitted for publication consideration. If people take this bias to an extreme, they could argue that the journals publish the 5% of studies with Type I errors that reject the null when in fact it is false. This would then mean that the “file drawers” of researchers contain the 95% of studies that confirmed the null hypothesis. That

is, their “file drawers” are filled with the 95% of studies showing statistically non-significant results that could not be published (Rosenthal 1980). Therefore, it is important for the researcher to seek out these unpublished studies through perusal of dissertations and calls for studies through various listservs and other such means. In Table 2, 26 of 52 studies (50%) explicitly report publication bias analysis. However, the reporting of publication bias in recent studies is more pronounced. A clear majority of studies have used the file drawer method.

File drawer N procedure The File Drawer N procedure helps determine how many null effect studies would be needed to change a significant meta-analytic result to non-significance (Rosenthal 1979; Rosenthal and Rosnow 2008). Reviewers should adopt this procedure in their meta-analyses as it provides additional evidence regarding the robustness of the results and highlights that the results are less likely to be substantively influenced by publication bias. Other methods are outlined by Borenstein et al. (2009), such as Orwin’s Fail Safe N and Duval and Tweedie’s Trim and Fill procedures (Duval and Tweedie 2000).

Orwin fail safe N procedure The Orwin Fail Safe N procedure identifies how many missing effects are needed to bring the overall effect size to a specific non-zero value (Orwin 1983). A researcher can also specify a non-zero mean effect size for the missing values.

Duval and Tweedie’s fill and trim procedure The Duval and Tweedie’s Fill and Trim procedure involves iteratively removing the most extreme effects from the positive side of the distribution (trimming), resulting in an adjusted effect size that is theoretically unbiased. This procedure ends up reducing the variance. The underlying algorithm also adds studies that were removed back in (filling) to correct the variance of the adjusted effect size. Computer programs, such as Comprehensive Meta-Analysis provide an image of the funnel plot that depicts the observed effects (and they also plot the imputed effects).

Performing moderator analysis

Subgroup analysis Since most meta-analyses explicitly seek to estimate the distribution of effects across two (or more) sets of studies, it is useful to report the distribution of effects separately for each set of studies. The researcher can then compare the means of the different sets of studies, analogous to a one-way analysis of variance or a t-test if only two studies are being compared. These moderators are frequently separated into theoretical moderators and study-related moderators.

For example, an important theoretical moderator in the comparative advertising domain is the market position of the sponsor brand relative to the comparison brand. In their meta-analysis of comparative advertising (Grewal et al. 1997) synthesized 43 effects where the sponsor brand position was less than the comparison brand and 12 effects where it was equal or greater than the comparison brand. The moderator analysis indicated that the effect of comparative advertising was three times more effective when the relative market position of the sponsor was less than the comparison brand.

However, such a finding does not confirm causality. The researcher might conclude an effect size is higher in one group relative to another group but it is not certain it is only due to that moderator. Accordingly, upon discovering significant differences between the various subgroups, the researchers should determine whether the difference persists after controlling for potential other moderators using meta-regression procedures. Also, there might be a need to conduct additional experimental studies to validate the causal nature of novel findings.

Meta-regression If a sufficient number of studies is available, multivariate statistical techniques (e.g., meta-regression) can simultaneously investigate relationships among key characteristics of the reviewed studies, such as their research design, subjects, treatments, settings, and findings. Thus, the use of multiple regression can help reveal the relationship between multiple moderators and the effect size of a given relationship (Borenstein et al. 2009; Lipsey and Wilson 2001).

Imagine two predictors, X_1 and X_2 . If a regression uses X_1 as a sole predictor, the results will reveal the relationship between X_1 and the effect size, without considering how this relationship might also be influenced by X_2 . A regression with both X_1 and X_2 as predictors will produce the unique impact of X_1 (controlling for any influence of X_2), the unique impact of X_2 (controlling for any influence of X_1), and the joint impact of X_1 and X_2 .

Like subgroup analyses, the relationships identified through meta-regression are not causal in nature. But if we control for all potential confounds (or possible confounds) and the relationship persists, we have a better case for arguing about the implied causality. Even then, some additional, unknown confounds may exert an influence. Regressions for meta-analyses and primary studies rely on the same basic principles, but their computations differ. Researchers can use readily available macros (Lipsey and Wilson 2001) for traditional spss and sas software and/or publicly available software (e.g., comprehensive meta-analysis) to run meta-regression analysis.

Many times, sub-group meta-regression analysis might need to be conducted and reported due to a host of reasons (e.g., multicollinearity, by industry, by product). For example, Rosario et al. (2016) in their meta-analysis of the effect of electronic word of mouth on sales highlight the role of the various moderators on the overall sample versus based on type of platform (social media, reviews, e-commerce), type of goods (tangible vs. services) and type of product. Thus, such sub-group meta-regression can shed additional insight. However, the initial number of effects needs to be large to do such sub-group analysis.

Using meta-analytic structural equation models

As we survey the meta-analytic papers in [Appendix 1](#) (Table 5), we note that meta-analysis has started to include structural equation models. Such an analysis allows a greater understanding of the underlying process or mechanism by which the independent variable underlying the effect influences the dependent variable. Details on meta-analytic structural equation modeling are covered in several books (e.g., Cheung 2015; Jak 2015).

Meta-analytic structural equation models (MASEM) can demonstrate the superiority of one type of process or mechanism model over another. For example, that the positive effect of publicity relative to advertising on attitudes and purchase intentions are better explained by a source credibility model as compared to an information processing or information evaluation model (Eisend and Küster 2011). As another example, Brown and Peterson (1993) test antecedents and consequences of salesperson job satisfaction. However, this approach demands considerably more data, in that the effect sizes between any two constructs in the model must be available. Consequently, most causal models are limited to the most frequently studied constructs in the meta-analysis, and researchers often limit their models to study only those constructs for which at least three studies report effect sizes. Similarly, Palmatier et al. (2006) identify 14 constructs of interest, but only 6 of them meet the criteria for inclusion in their causal model.

Brown and Stayman (1992) and Brown and Peterson (1993) were among the first in marketing to present the use of a causal model approach. They recommend beginning with a matrix of meta-analytic correlations between constructs in the model, which then can be submitted to a causal model analysis. Brown and Stayman (1992) examined four alternative models of ad attitudes and found superior fit for a dual mediation hypothesis model.

The correlations used as the input matrix in the structural tests are typically adjusted for measurement error as described earlier. Since the various correlations in the matrix likely represent an accumulation of different effects

(and different overall sample sizes), many meta-analyses use a harmonic mean of the various samples as the N (e.g., Rubera and Kirca 2012) as opposed to an arithmetic mean in the structural model (see Viswesvaran and Ones 1995). Others have used the median sample size (e.g., Notani 1998; Orsingher et al. 2010; Palmatier et al. 2006). The main objective is to be conservative and not be influenced by extreme sample sizes and therefore the input matrix will likely be more representative of the domain and studies being examined.

Summary and conclusions

Over the past two decades, thousands of new empirical findings have been reported in the marketing literature. The discipline of marketing has become mature in numerous domains to provide sufficient studies to warrant meta-analytic examinations. The marketing discipline faces the necessary but difficult task of categorizing, organizing, and integrating this expanding body of knowledge, especially as calls for replication and extension of prior research increase. Importantly, meta-analysis has informed public policy in numerous domains (Franke 2001), such as product warnings (e.g., Cox et al. 1997) and health communications (Keller and Lehmann 2008).

When carefully conducted, a meta-analysis of a research domain is a systematic procedure for integrating past research, resolving apparent inconsistencies, identifying important moderators, explicating underlying processes, and promoting innovative research within the domain. Scholars have noted the ability of meta-analysis to distinguish between the size of an effect and its significance (Franke 2001). Best practices in meta-analysis require researchers to make a number of key decisions. We have identified many of these key decisions here. Readers should also consult other discussions on these issues (Geyskens et al. 2009; Watson et al. 2015). It is heartening to see the number of meta-analyses that are appearing in marketing publications and the role they are playing in encouraging additional research in their respective substantive, theoretical, and methodological domains.

Well-done meta-analyses are systematic and replicable. They offer researchers an opportunity to determine the extent specific research results are influenced by methodological quality. Moreover, combining homogeneous results across studies increases statistical power. As reviewed here, multivariate techniques and structural modeling are increasingly being used providing sophistication and maturity to the marketing knowledge identified. Meta-analysis helps isolate relationships among relevant variables while also obtaining more accurate estimates of effect sizes. Importantly, a meta-analysis is useful for developing theories about phenomena of interest.

Appendix 1

Table 5 Review of meta-analyses

Citation	Domain	Substantive Issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Homogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Farley et al. (1981)	Consumer Behavior Method	Behavioral Intention Response rates	2	1966–1978	37	NR	NR	reg. Coeff. and R response rate %	no	no	no	no	no
Yu and Cooper (1983)	Communications	Effect of Advertising on Sales	2	1965–1981	93	NR	497	elasticity	no	no	yes	no	no
Assmus et al. (1984)	Method	Determinants of Salesperson Performance	2	1964–1982	101	2	154	reliability	no	no	no	no	no
Churchill Jr and Peter (1984)	Sales	Effects Sizes in Consumer Behavior	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Churchill et al. (1985)	Consumer Behavior	Request strategy Rating Scales	1	1970–1982	118	NR	1036	omega squared	no	no	no	no	no
Peterson et al. (1985)	Sales	Generics-Prone Consumer Price Perception	1	1964–1982	101	2	154	reliability	no	no	no	no	no
Fern et al. (1986)	Consumer Behavior	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Peter and Churchill (1986)	Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Szymanski and Busch (1987)	Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Tellis and Wernerfelt (1987)	Consumer Behavior	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Sheppard et al. (1988)	Pricing Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Tellis (1988)	Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Rao and Monroe (1989)	Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Sultana et al. (1990)	Product Management	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Brown and Stayman (1992)	Communications	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Brown and Peterson (1993)	Sales	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Wilson and Sherrell (1993)	Consumer Behavior	Request strategy Rating Scales	1	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Peterson (1994)	Method	Request strategy Rating Scales	2	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Heath and Chatterjee (1995)	Communications	Request strategy Rating Scales	2	1918–1982	116	1 (all), 2 (5%)	1653	r	yes	no	yes	no	no
Citation	Domain	Substantive issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Homogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Cox et al. (1997)	Product Management	Product Warnings	2	1976–1995	15	2	79	compliance rate	no	no	yes	Yes	no
Grewal et al. (1997)	Communications	Comparative Versus Noncomparative Advertising	1	1975–1996	77	NR	262	d	yes	yes	yes	yes	no
Brown et al. (1998)	Communications	Ad-evoked Feelings and Advertising Response	1	1986–1995	55	1, verified by mailing authors	134	r	yes	yes	yes	no	no

Table 5 (continued)

Citation	Domain	Substantive Issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Homogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Compeau and Grewal (1998)	Pricing	Effectiveness of Comparative Price Advertising	1	1974–1992	38	NR	86	Eta	no	yes	no	yes	Yes
Notani (1998)	Consumer Behavior	Theory of Planned Behavior	1	1989–1995	36	2, agree >85%	63	r	no	yes	yes	no	yes
Ceyskens et al. (1999)	Sales	Satisfaction in Marketing Channel Relationships	1	1970–1996	107	NR	NR	r	no	no	yes	no	no
Spangenberg and Greenwald (1999)	Consumer Behavior	Social Influence	1	1980–1997	11	NR	11	r	no	yes	no	no	no
Estelami and Lehmann (2001)	Pricing	Consumer Price Recall	2	1961–1999	22	NR	279	% average deviation	no	no	no	no	no
Estelami et al. (2001)	Pricing	Consumer Price Knowledge	2	1961–1999	27	NR	297	Price Recall Error	no	no	no	no	no
Henard and Szymanski (2001)	Product Management	New product performance	1	NR-1999	41	2, agree = 98.2%	798	r	yes	yes	yes	yes	no
Peterson (2001)	Method	Student Samples	3	1981–2000	28	NR	63	r, d, g, β , ω^2	no	no	no	no	no
Szymanski and Henard (2001)	Consumer Behavior	Customer Satisfaction	1	NR-1986	50	2, agree = 96%	517	r	yes	yes	yes	yes	no
Krishna et al. (2002)	Pricing	Price Presentation	2	1980–1998	30	2	345	% saving	no	no	no	no	no
Januszewski et al. (2003)	Consumer Behavior	Spacing Effect in Verbal Learning	1	1887–2000	61	NR	269	r	no	no	no	yes	no
Van den Bulte and Stremersch (2004)	Product Management	New Product Diffusion	2	1969–2000	46	NR	293 observations	q/p ratio	no	no	no	no	no
Kirca et al. (2005)	Strategy	Market Orientation	1	NR-2004	114 studies/ 130 independent samples	1 (all), 1 (35%), Reliability: .91–1.0	418	r	yes	no	yes	Yes	no
Conchar et al. (2005)	Strategy	Advertising and promotion spend on market value of the firm	2	1978–2004	88	NR	88	A&P Effect	no	yes	yes	yes	no
Citation	Domain	Substantive issue	Type	Time Period	# Studies/ Samples	coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Heterogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Franke and Park (2006)	Sales	Salesperson Behavior	1	1979–2006	155	2	NR	r	yes	yes	no	no	no
Palmiter et al. (2006)	Sales	Relationship Marketing Improves Performance	1	1987–2004	111	2, agree >95%	637	r	yes	yes	yes	yes	no
Szymanski et al. (2007)	Product Management	Innovativeness and New Product Success	1	NR-2006	32	2, agree = 95%	95	r	yes	yes	yes	no	yes
Wright and MacRae (2007)	Consumer Behavior	Bias and Variability in Purchase Intention Scales	1	NR-2006	9	NR	52 proportion estimates	Logs Odd Ratio	no	yes	yes	no	yes
Carlson et al. (2008)	Consumer Behavior	Relationship between Objective and Subjective Knowledge	1	1980–2007	51	2, agree = 84.4%	104	r	yes	yes	yes	yes	yes
Crosno and Dahlstrom (2008)	Channels	Opportunism	1	1982-2005	59	1	183	r	no	yes	yes	no	no
de Matos and Vargas Rossi (2008)	Communications	Word-of-mouth communications	1	1955–2007	162	NR	348	r	yes	yes	yes	yes	yes
Keller and Lehmann (2008)	Communications	Effective Health Communications	2	1961–2006	60	2	NR	Intention (0.0 to 1.0 scale)	no	no	no	no	no

Table 5 (continued)

Citation	Domain	Substantive Issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Homogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Krasnikov and Jayachandran (2008)	Strategy	Firm Performance	1	NR–2007	114	2	786	r	yes	No	yes	yes	no
Troy et al. (2008)	Product Management	New Product Success	1	NR–2007	25	2 coded 50%, cross-checked	146	r	yes	yes	yes	no	no
Grinstein (2008)	Product Management	Market Orientation on Innovation	1	1994–2006	56	NR	69	r	yes	yes	yes	no	yes
Eisend (2009)	Communications	Humor in Advertising	1	1960–2006	43	2, k > .9	369	r	yes	yes	yes	yes	no
Keller et al. (2009)	Communications	Corporate Well-Being Program	2	1982–2006	60	2	76 observations	% cost reduction	no	no	no	no	no
Eisend (2010)	Communications	Gender Roles in Advertising	1	1975–2007	64	2	NR	odds-ratio	yes	yes	yes	no	yes
Orsingher et al. (2010)	Consumer Behavior	Satisfaction With Complaint Handling	1	NR–1997	60	NR	509	r	yes	yes	yes	yes	yes
Scheibehenne et al. (2010)	Consumer Behavior	Choice Overload	1	2000–2009	50	NR	63 datapoints	d	no	yes	Yes	no	yes
Albers et al. (2010)	Sales	Personal Selling	2	NR–2010	88	2, agree >90%	506	elasticity	no	no	no	no	no
Eisend and Küster (2011)	Communications	Publicity vs. Advertising	1	1971–2009	30	NR	NR	r	yes	Yes	yes	no	no
Sethuraman et al. (2011)	Communications	Brand Advertising Elasticity	2	1960–2008	56	NR	751 short term; 402 long term	elasticity	no	no	no	no	no
Verbeke et al. (2011)	Sales	Sales Performance	1	1982–2008	292	1 (all), 1 (20%), agree >95%	2105	r	yes	yes	yes	yes	yes
Citation	Domain	Substantive Issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Heterogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Rubera and Kirca (2012)	Product Management	Innovativeness and Its Performance Outcomes	1	NR–2010	159	1 (all), 1 (35 studies), agree = 90%	NR	r	yes	yes	yes	yes	no
Homburg et al. (2012)	Method	Key informant accuracy	1	1998–2010	127	4, reliability .71–.86	11,874	r, agreements	yes	no	yes	Yes	no
Sleekamp and Geyskens (2012)	Sales	Transaction Cost Economics	1	NR–2011	128 studies/13 independent samples	NR	211	r	yes	no	no	no	yes
Zablah et al. (2012)	Sales	Customer Orientation	1	1979–2011	323	1, multiple for moderators	NR	r	yes	no	no	no	no
Motyka et al. (2014)	Consumer Behavior	Regulatory Fit Leads to More Favorable Consumer Response	1	1998–2013	215	2, agree >90%	306	r	yes	yes	yes	yes	no
Van Laer et al. (2014)	Consumer Behavior	Transportation Imagery	1	NR–2014	76	2, k = .75	132	r	yes	no	yes	yes	no
Pick and Eisend (2014)	Loyalty	switching costs	1	NR–2012	170	2, agree = 96%	800	r	yes	yes	yes	no	no
Zlatevska et al. (2014)	Product Management	Portion size	1	1994–2013	104	NR	104	d	no	yes	no	yes	no
Eisend (2015)	Method	Method Factors	3	1976–2012	176	Domain: 2, agree-95% Other: 2, 2nd coded 25% agree = 100%	1841	r	no	no	no	yes	no
Scheer et al. (2015)	Channels	Dependence	1	1970–2014	218	2, agree = 96%	976	r	yes	yes	yes	yes	no
Crosno and Brown (2015)	Channels	Organizational Control and Channels	1	1975–2010	66	2, agree >98%	165	r	yes	yes	yes	yes	yes
Watson et al. (2015)	Loyalty	Loyalty	1	1980–2013	126/151	2, agree >96%	713	r	yes	yes	yes	no	no
You et al. (2015)	Loyalty	Electronic Word of Mouth	2	NR–2015	51	2, k = .85	339 + 271	elasticity	no	no	no	no	yes
Rosario et al. (2016)	Loyalty	Electronic Word of Mouth	1	1999–2013	96	1 (all), 1 (subset), k = .97	1532	r	no	yes	yes	yes	yes
			1	1998–2014	39	2, k > 82%	123	r	yes	yes	yes	yes	yes

Table 5 (continued)

Citation	Domain	Substantive Issue	Type	Time Period	# Studies/ Samples	Coders	# Effect Sizes	Effect Size Type	ES Reliability Adjusted	Test Homogeneity	Used Weighting	Test Publication Bias	Assess Outliers
Chang and Taylor (2016)	Product Management	Customer Participation in NPD	2	1977–2013	83	1 (all), 1 (30), agree >90%	488	elasticity	no	yes	yes	no	yes
Edelung and Fischer (2016)	Services	Firm value	1	1994–2015	518	3, agree = 93%	687	r	yes	no	yes	yes	yes
Hogreve et al. (2017)	Communications	Service Profit	1	NR-2016	46	2, k = .74	279	d	no	no	yes	yes	yes
Knoll and Matthes (2017)	Communications	celebrity endorsement	1	NR-2016	46	2, k = .74	279	d	no	no	yes	yes	yes

References

Albers, S., Mantrala, M. K., & Sridhar, S. (2010). Personal selling elasticities: A meta-analysis. *Journal of Marketing Research*, 47(5), 840–853.

Assmus, G., Farley, J. U., & Lehmann, D. R. (1984). How advertising affects sales: Meta-analysis of econometric results. *Journal of Marketing Research*, 21, 65–74.

Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, 87(3), 583–589.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1, 97–111.

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8, 5–18.

Brown, S. P., & Peterson, R. A. (1993). Antecedents and consequences of salesperson job satisfaction: Meta-analysis and assessment of causal effects. *Journal of Marketing Research*, 30, 63–77.

Brown, S. P., & Stayman, D. M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research*, 19, 34–51.

Brown, S. P., Homer, P. M., & Inman, J. J. (1998). A meta-analysis of relationships between ad-evoked feelings and advertising responses. *Journal of Marketing Research*, 35, 114–126.

Campbell, J. P., Daft, R. L., & Hulin, C. L. (1982). *What to study: Generating and developing research questions*. Beverly Hills: Sage Publications, Inc..

Carlson, J. P., Vincent, L. H., Hardesty, D. M., & Bearden, W. O. (2008). Objective and subjective knowledge relationships: A quantitative analysis of consumer research findings. *Journal of Consumer Research*, 35, 864–876.

Chang, W., & Taylor, S. A. (2016). The effectiveness of customer participation in new product development: A meta-analysis. *Journal of Marketing*, 80, 47–64.

Cheung, M.W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. Chichester: John Wiley & Sons

Churchill Jr, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 21, 360–375.

Churchill Jr., G. A., Ford, N. M., Hartley, S. W., & Walker Jr., O. C. (1985). The determinants of salesperson performance: A meta-analysis. *Journal of Marketing Research*, 22, 103–118.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37(3/4), 256–266.

Compeau, L. D., & Grewal, D. (1998). Comparative price advertising: An integrative review. *Journal of Public Policy & Marketing*, 17(2), 257–273.

Conchar, M. P., Crask, M. R., & Zinkhan, G. M. (2005). Market valuation models of the effect of advertising and promotional spending: A review and meta-analysis. *Journal of the Academy of Marketing Science*, 33(4), 445–460.

Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52(Summer), 291–302.

Cox III, E. P., Wogalter, M. S., Stokes, S. L., & Tipton Murff, E. J. (1997). Do product warnings increase safe behavior? A meta-analysis. *Journal of Public Policy & Marketing*, 16(2), 195–204.

Crosno, J. L., & Brown, J. R. (2015). A meta-analytic review of the effects of organizational control in marketing exchange relationships. *Journal of the Academy of Marketing Science*, 43(3), 297–314.

- Crosno, J. L., & Dahlstrom, R. (2008). A meta-analytic review of opportunism in exchange relationships. *Journal of the Academy of Marketing Science*, 36(2), 191–201.
- de Matos, C. A., & Vargas Rossi, C. A. (2008). Word-of-mouth communications in marketing: A meta-analytic review of the antecedents and moderators. *Journal of the Academy of Marketing Science*, 36(4), 578–596.
- Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.
- Edeling, A., & Fischer, M. (2016). Marketing's impact on firm value: Generalizations from a meta-analysis. *Journal of Marketing Research*, 53(4), 515–534.
- Eisend, M. (2009). A meta-analysis of humor in advertising. *Journal of the Academy of Marketing Science*, 37(2), 191–203.
- Eisend, M. (2010). A meta-analysis of gender roles in advertising. *Journal of the Academy of Marketing Science*, 38, 418–440.
- Eisend, M. (2015). Have we progressed marketing knowledge? A meta-meta-analysis of effect sizes in marketing research. *Journal of Marketing*, 79(3), 23–40.
- Eisend, M., & Küster, F. (2011). The effectiveness of publicity versus advertising: A meta-analytic investigation of its moderators. *Journal of the Academy of Marketing Science*, 39(6), 906–921.
- Estelami, H., & Lehmann, D. R. (2001). The impact of research design on consumer price recall accuracy: An integrative review. *Journal of the Academy of Marketing Science*, 29(1), 36–49.
- Estelami, H., Lehmann, D. R., & Holden, A. C. (2001). Macro-economic determinants of consumer price knowledge: A meta-analysis of four decades of research. *International Journal of Research in Marketing*, 18(4), 341–355.
- Farley, J. U., Lehman, D. R., & Ryan, M. J. (1981). Generalizing from 'imperfect' replication. *Journal of Business*, 54, 597–610.
- Farley, J. U., Lehmann, D. R., & Sawyer, A. (1995). Empirical generalizations using meta-analysis. *Marketing Science*, 13(3), G36–G46.
- Fern, E. F., & Monroe, K. B. (1996). Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research*, 23(2), 89–105.
- Fern, E. F., Monroe, K. B., & Avila, R. A. (1986). Effectiveness of multiple request strategies: A synthesis of research results. *Journal of Marketing Research*, 23, 144–152.
- Franke, G. R. (2001). Applications of meta-analysis for marketing and public policy: A review. *Journal of Public Policy & Marketing*, 20(2), 186–200.
- Franke, G. R., & Park, J.-E. (2006). Salesperson adaptive selling behavior and customer orientation: A meta-analysis. *Journal of Marketing Research*, 43 (November), 693–702.
- Geyskens, I., Steenkamp, J.-B. E. M., & Kumar, N. (1999). A meta-analysis of satisfaction in marketing channel relationships. *Journal of Marketing Research*, 36, 223–238.
- Geyskens, I., Krishnan, R., Steenkamp, J. B. E., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management*, 35(2), 393–419.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1–20.
- Grewal, D., Kavanoor, S., Fern, E. F., Costley, C., & Barnes, J. (1997). Comparative versus noncomparative advertising: A meta-analysis. *Journal of Marketing*, 61, 1–15.
- Grinstein, A. (2008). The effect of market orientation and its components on innovation consequences: A meta-analysis. *Journal of the Academy of Marketing Science*, 36, 166–173.
- Heath, T. B., & Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, 22, 268–284.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Henard, D. H., & Szymanski, D. M. (2001). Why some new products are more successful than others. *Journal of Marketing Research*, 38, 362–375.
- Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558.
- Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560.
- Hogreve, J., Iseke, A., Derfuss, K., & Eller, T. (2017). The service-profit chain: A meta-analytic test of a comprehensive theoretical framework. *Journal of Marketing*, 81(3), 41–61.
- Homburg, C., Klarmann, M., Reimann, M., & Schilke, O. (2012). What drives key informant accuracy? *Journal of Marketing Research*, 49(4), 594–608.
- Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 89(2), 327–334.
- Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, 28, 149–158.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Thousand Oaks: Sage Publications, Inc..
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks: Sage Publications, Inc..
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills: Sage Publications, Inc..
- Jak, S. (2015). *Meta-analytic structural equation modelling*. Cham: Springer International Publishing.
- Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research*, 30, 138–149.
- Keller, P. A., & Lehmann, D. R. (2008). Designing effective health communications: A meta-analysis. *Journal of Public Policy & Marketing*, 27(2), 117–130.
- Keller, P. A., Lehmann, D. R., & Milligan, K. J. (2009). Effectiveness of corporate well-being programs. *Journal of Macromarketing*, 29(3), 279–302.
- Kirca, A. H., Jayachandran, S., & Bearden, W. O. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of Marketing*, 69, 24–41.
- Knoll, J., & Matthes, J. (2017). The effectiveness of celebrity endorsements: A meta-analysis. *Journal of the Academy of Marketing Science*, 45(1), 55–75.
- Krasnikov, A., & Jayachandran, S. (2008). The relative impact of marketing, research-and-development, and operations capabilities on firm performance. *Journal of Marketing*, 72, 1–11.
- Krishna, A., Briesch, R., Lehmann, D. R., & Yuan, H. (2002). A meta-analysis of the impact of price presentation on perceived savings. *Journal of Retailing*, 78, 101–118.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (p. 49). Thousand Oaks: Sage Publications.
- Monroe, K. B., & Krishnan, R. (1983). A procedure for integrating outcomes across studies. *Advances in Consumer Research*, 10, R. P. Bagozzi and A. M. Tybout (eds.), 503–508.
- Motyka, S., Grewal, D., Puccinelli, N. M., Roggeveen, A. L., Avnet, T., Daryanto, A., & Wetzels, M. (2014). Regulatory fit: A meta-analytic synthesis. *Journal of Consumer Psychology*, 24(3), 394–410.

- Notani, A. S. (1998). Moderators of perceived behavioral control's predictiveness in the theory of planned behavior: A meta-analysis. *Journal of Consumer Psychology, 7*(3), 247–272.
- Orsingher, C., Valentini, S., & de Angelis, M. (2010). A meta-analysis of satisfaction with complaint handling in services. *Journal of the Academy of Marketing Science, 38*(2), 169–186.
- Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157–159.
- Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors influencing the effectiveness of relationship marketing: A meta-analysis. *Journal of Marketing, 70*, 136–153.
- Palmatier, R. W., Houston, M. B., & Hulland, J. (2017). Review articles: Purpose, process, and structure. *Journal of the Academy of Marketing Science, 46*(1). <https://doi.org/10.1007/s11747-017-0563-4>.
- Peter, J. P., & Churchill Jr., G. A. (1986). Relationships among research design choices and psychometric properties of rating scales: A meta-analysis. *Journal of Marketing Research, 23*, 1–10.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21*(2), 381–391.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research, 28*(3), 450–461.
- Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175–181.
- Peterson, R. A., Albaum, G., & Beltrami, R. F. (1985). A meta-analysis of effects sizes in consumer behavior experiments. *Journal of Consumer Research, 12*, 97–103.
- Pick, D., & Eisend, M. (2014). Buyers' perceived switching costs and switching: A meta-analytic assessment of their antecedents. *Journal of the Academy of Marketing Science, 42*(2), 186–204.
- Pillemer, D. B., & Light, R. (1980). Synthesizing outcomes: How to use research evidence from many studies. *Harvard Educational Review, 50*, 176–195.
- Puccinelli, N.M., et al. (2013). Are men seduced by red? The effect of red versus black prices on price perceptions. *Journal of Retailing, 89*(2), 115–125.
- Rao, A. R., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research, 26*, 351–357.
- Rosario, A. B., Sotgiu, F., De Valck, K., & Bijmolt, T. H. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research, 53*, 297–318.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641.
- Rosenthal, R. (1980). Summarizing significance levels. In R. Rosenthal (Ed.), *New Directions for Methodology of Social and Behavioral Science: Quantitative Assessment of Research Domains, (5)* (pp. 33–46). San Francisco: Jossey-Bass.
- Rosenthal, R. (1982). Valid interpretation of quantitative research results. In D. Brinberg & L. Kidder (Eds.), *New Directions for Methodology of Social and Behavioral Science: Forms of Validity in Research, (12)* (pp. 59–75). San Francisco: Jossey-Bass.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills: Sage Publications.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*(1), 59–82.
- Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). Boston: McGraw-Hill.
- Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Social Psychology, 9*(5), 395–396.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*(2), 166.
- Rubera, G., & Kirca, A. H. (2012). Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of Marketing, 76*, 130–147.
- Scheer, L. K., Miao, C. F., & Palmatier, R. W. (2015). Dependence and interdependence in marketing relationships: Meta-analytic insights. *Journal of the Academy of Marketing Science, 43*(6), 694–712.
- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research, 37*(3), 409–425.
- Sethuraman, R., Tellis, G. J., & Briesch, R. A. (2011). How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research, 48*(3), 457–471.
- Sheppard, B.H., Hartwick, J., & and Warshaw, P.R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*, 325–343.
- Spangenberg, E. R., & Greenwald, A. G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology, 8*(1), 61–89.
- Steenkamp, J.-B. E. M., & Geyskens, I. (2012). Transaction cost economics and the roles of national culture: A test of hypotheses based on Inglehart and Hofstede. *Journal of the Academy of Marketing Science, 40*, 252–270.
- Sultan, F., Farley, J. U., & Lehmann, D. R. (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research, 27*, 70–77.
- Szymanski, D. M., & Busch, P. S. (1987). Identifying the generics-prone consumer: A meta-analysis. *Journal of Marketing Research, 24*, 425–431.
- Szymanski, D. M., & Henard, D. H. (2001). Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science, 29*(1), 16–35.
- Szymanski, D. M., Kroff, M. W., & Troy, L. C. (2007). Innovativeness and new product success: Insights from the cumulative evidence. *Journal of the Academy of Marketing Science, 35*, 35–52.
- Tellis, G. J. (1988). The price elasticity of selective demand: A meta-analysis of econometric models of sales. *Journal of Marketing Research, 25*, 331–341.
- Tellis, G. J., & Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science, 6*(3), 240–253.
- Troy, L. C., Hirunyawipada, T., & Paswan, A. K. (2008). Cross-functional integration and new product success: An empirical investigation of the findings. *Journal of Marketing, 72*, 132–146.
- Vadillo, M. A., Gold, N., & Osman, M. (2016). The bitter truth about sugar and willpower: The limited evidential value of the glucose model of ego depletion. *Psychological Science, 27*(9), 1207–1214.
- Van den Bulte, C., & Stremersch, S. (2004). Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science, 23*(4), 530–554.
- Van Laer, T., De Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research, 40*(5), 797–817.
- Verbeke, W., Dietz, B., & Verwaal, E. (2011). Drivers of sales performance: A contemporary meta-analysis. Have salespeople become knowledge brokers? *Journal of the Academy of Marketing Science, 39*(3), 407–428.

- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48*(4), 865–885.
- Völckner, F., & Hoffmann, J. (2007). The price-perceived quality relationship: A meta-analytic review and assessment of its determinants. *Marketing Letters, 18*(3), 181–196.
- Watson, G. F., Beck, J. T., Henderson, C. M., & Palmatier, R. W. (2015). Building, measuring, and profiting from customer loyalty. *Journal of the Academy of Marketing Science, 43*(6), 790–825.
- Wells, W. D. (2001). The perils of $N = 1$. *Journal of Consumer Research, 28*, 494–498.
- Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science, 21*(2), 101–112.
- Wright, M., & MacRae, M. (2007). Bias and variability in purchase intention scales. *Journal of the Academy of Marketing Science, 35*, 617–624.
- You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing, 79*(2), 19–39.
- Yu, J., & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research, 20*, 36–44.
- Zablah, A. R., Franke, G. R., Brown, T. J., & Bartholomew, D. E. (2012). How and when does customer orientation influence frontline employee job outcomes? A meta-analytic evaluation. *Journal of Marketing, 76*, 21–40.
- Zlatevska, N., Dubelaar, C., & Holden, S. S. (2014). Sizing up the effect of portion size on consumption: A meta-analytic review. *Journal of Marketing, 78*, 140–154.