UNIVERSITY OF
BATH

University of Bath

META-ANALYSIS: INTEGRATING ACCUMULATED KNOWLEDGE

Dhruv Grewal
Toyota Chair of Commerce and Electronic Business
Professor of Marketing, Marketing Division
213 Malloy Hall, Babson College
Babson Park, MA 02457
Phone: (781) 239-3902
e-mail: dgrewal@babson.edu

Nancy Puccinelli
Head of Division -Research
Professor of Marketing
University of Bath
and
Associate Fellow
Oxford University
N.M.Puccinelli@Bath.ac.uk

Kent B. Monroe
Professor of Marketing Emeritus
Department of Business Administration
University of Illinois
4905 Old Millrace Place
Glen Allen, VA 23059
kbmonroe@illinois.edu
kent@kentmonroe.net
804-716-2395 (TEL)
206-605-8209 (CELL)

September 29, 2017

META-ANALYSIS: INTEGRATING ACCUMULATED KNOWLEDGE

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 establishing a summary of 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 of inquiry. This article highlights the growth in the use of meta-analysis in marketing and identifies several important issues researchers must consider when conducting and reporting a meta-analysis.

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 domain. Meta-analysis represents a rigorous alternative for making sense of a rapidly expanding research literature (Glass 1976). Meta-analysis can help identify the expanding boundaries of a given research domain by highlighting current knowledge and drawing attention to 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.

The challenge is how to identify and represent this accumulated knowledge. Palmatier, Houston and Hulland (2018) highlight the importance and need for review papers. While narrative reviews can summarize collections of studies, more sophisticated meta-analytic methods for synthesizing knowledge are being used. The benefits of a meta-analysis are myriad: it tests the robustness of a finding, it helps resolve apparent conflicting findings, identifies research design issues, and proposes appropriate designs for future studies. An important benefit of meta-analysis is it provides a way to compare and combine results across a variety of studies. It provides a systematic procedure to determine consistency of results while at the same time explaining variations in observed effects. To establish the boundaries of knowledge in a research domain, it is important to determine sources of variation in results across studies. That is, what do we know and more importantly, what do we **not know**? A single study is a sample of one and can rarely, if ever, resolve a major research question (Wells 2001).

The aggregation of studies in a research domain produces accumulated knowledge by (1) helping develop and test the theoretical bases and underlying predictions, and (2) assessing the empirical evidence for a 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). Therefore, there is a need to quantitatively integrate research in various marketing domains, such as pricing, promotion, channels, new products, service quality, and numerous others. In addition, there is need to understand the magnitude of the role of drivers and mediators have on outcomes in numerous consumer behavior domains (e.g., affect, impulse buying, internal reference prices). The objective is to better understand specific questions or problems (e.g., does X Influence Y?) and to examine how specific characteristics of the various studies (e.g., type of manipulation of X, or measure of Y) in that domain influence the variation in results across the various studies. The popularity of meta-analysis to achieve knowledge extraction has grown in recent years.

The popularity of meta-analysis is growing within various sub-fields of marketing such as consumer behavior (Scheibehenne et al. 2010), branding (Keller et al. 2009), advertising (Eisend and Kuster 2011), sales (Verbeke et al. 2011), and product development (Rubera and Kirca 2012). The meta-analytic studies in these various domains also shed light on numerous methodological characteristics, such as the type of sample, experimental versus survey results, differences based on a key type of operationalization, and/or the way a central construct was measured that bear on the nature of an effect. These insights can guide future research within and across domains.

Marketing as a discipline has developed to a point that researchers have conducted 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. The advent of a second-order meta-analysis is a sign the marketing field is reaching maturity and as such should be celebrated.

Geyskens, Krishnan, Steenkamp, and Cunha (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 highlight several areas that overlap with their paper (e.g., effect sizes, homogeneity, outlier analysis, correction procedures), we discuss several other 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.

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 provide more accurate assessments (e.g., using weighting mechanisms, outlier assessments, publication bias analysis). Accordingly, we briefly outline some advantages of using meta-analysis. Next, we summarize several 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 has become the dominant research synthesis tool in most 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 (often called the effect size estimate, or effect

size for short), researchers can synthesize any set of studies that address the same fundamental relationship. To the extent 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, Schmidt and Jackson (1982), Lipsey and Wilson (2001), and 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 capture 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 then a researcher should use meta-regression procedures (Hierarchical linear meta-regression or HiLMA).

By combining a diverse 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 impact 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.

TYPES OF META-ANALYSES

Considering several influential examples, we observe some notable patterns (see Appendix 1). Appendix 1 provides a summary of 74 meta-analyses in leading marketing journals from 1981-2017. Although this summary list is not exhaustive, 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*). Marketing was relatively quick to adopt the approach; some of the first meta-analyses were published in 1985 (Churchill, Ford, Walker and Walker 1985; Peterson, Albaum and Beltramini 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.

The review of the various meta-analysis in the marketing field in the set of journals that we examine highlights that there are some inherent differences in the types of meta-analyses. In Table 1, we list the three different types of meta-analysis and the number in the set we reviewed.

Insert Table 1 about here

Type 1 or Standard Meta-Analysis. Most of the meta-analyses involved integrating insights across a domain or across a set of relationships. In our review, 52 of 74 meta-analyses fit into this category. These articles tend to use standard meta-analytic techniques recommended by several authors (e.g., Borenstein et al. (2009), Rosenthal (1984)). The sophistication of the techniques used has evolved over time. Most important is the specific key variable is an effect size. These meta-analyses tend to focus on

- Either broadly integrating numerous relationships or a focused analysis on a set of relationships.

- Examining 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.

- They 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) tend to focus on main effects and moderators.

Type 2 or Replication Analysis. Replication analysis is a term coined by Farley, Lehmann, and Ryan (1981). Such meta-analyses do not necessarily follow traditional meta-analytic procedures. However, they use some key measure (or measures) from the studies being integrated. In our review, 20 of 74 meta-analysis fit into this category. For example, Assmus, Farley and Lehmann (1984) examined studies pertaining to advertising effects on sales and analyzed estimated parameters from 128 models. The dependent measures in their ANOVAs were short term elasticity, carryover coefficient and goodness of fit. Goodness of fit measures are not effect sizes.

Similarly, Sultan, Farley, and Lehmann (1990) examine diffusion models and consider factors that influence coefficient of innovation and coefficient of imitation across 213 applications. Numerous others use similar sorts of methodologies on a variety of topics. We identify articles that fit into this type. Many meta-analytic procedures do not apply to meta-analysis of this type.

Type 3 or Second-Order Meta-Analysis. These articles take the effect sizes from published meta-analysis and look at the effects of other variables that might be influencing them, such as Peterson (2001) qualitatively and Eisend (2015) quantitatively. In our review, these two meta-analyses fit into this category. Again, many meta-analytic procedures discussed in the remainder of this article might be less pertinent for type 3 meta-analyses.

KEY CONSIDERATIONS FOR CONDUCTING A META-ANALYSIS

Many key steps or considerations need to be considered 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 1 provides a potential flow chart on how a meta-analytic researcher may proceed when conducting a Type 1 meta-analysis. Table 2 highlights some of the key considerations for this type of meta-analysis. It also highlights how these choices have changed over the decades, 1980's, 1990's, 2000's and 2010's.

Insert Figure 1 and Table 2 about here

Determine 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, Monroe and Avila 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 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, Daft, and Hulin 1982).

The research question may be substantive, such as does price influence consumers' perceptions of quality (Rao and Monroe 1989), are comparative ads more effective than non-comparative ads (Grewal et al. 1997), or is objective performance influenced more by relationship quality than commitment (Palmatier et al. 2006). Or the research domain may be more conceptual in nature, such as the theory of reasoned action (Sheppard, Hartwick and Warshaw 1988) or regulatory fit (Motyka et al. 2014). Or, it may focus on methodological issues such as research design choices (Peter and Churchill 1986). It is important to note that the research goals will drive a lot of the meta-analysis decisions.

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

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, the researchers 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 to locate studies that otherwise would be difficult to find.

The search strategy adopted affects 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 the 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** 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 seeks to identify 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 indicate their inclusion/exclusion criteria so readers can assess the validity of the meta-analysis and design meaningful future research. 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

As discussed below 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 thus report questionable findings, no single study is perfect, and 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 range 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 have been outlined. A sample checklist is provided in Table 3.

Insert Table 3 about here

Coding the Studies

In a meta-analytic review, the primary research studies provide the data for the analyses. To be able 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. One of the most important tasks of the researcher is to detail the characteristics of the studies examined (Pillemer and Light 1980).

In this regard, the researcher must take the perspective of a detective and examine each study microscopically. To facilitate the process, a coding form should be developed that reflects 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. 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.

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.

An abbreviated version of the coding guide used by Motyka et al. (2014) in their meta-analysis pf the regulatory fit literature is displayed in Table 4. This sample coding sheet highlights 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.

Insert Table 4 about here

Effect Size

Behavioral research usually relies on statistical significance tests to draw inferences (Borenstein et al. 2009; Rosenthal and DiMatteo 2001). 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. To address this issue, many studies in behavioral science report effect sizes, defined as an estimate of the difference across groups, independent of sample size (Borenstein 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 measures needed to calculate the 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 measures and how to convert from one measure 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 many 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.

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. 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 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) highlight 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 1 meta-analyses indicates 44 out of 52 meta-analyses used a correlation as an effect size, and this pattern is similar across time.

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) it is usually appropriate for researchers to 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 33 of 52 meta-analyses use effect sizes adjusted for reliability. Adjusting for

measurement error occurs 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 estimates should be reported with and without these corrections.

Multiple Outcomes from Single Studies. Another issue is when it may be 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, Schmidt, and Jackson 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.

Homogeneity of the Effect Sizes

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:

(1) $\sum_{J=1}^{n} (Z_j - Mean\ Z)^2$ distributed as χ^2 with N – 1 df.

A preferable test for the statistical significance of the homogeneity of results requires

transforming the correlations to their associated Fisher z's and conducting a chi-square test

(Rosenthal 1982). The Fisher *r* to *z* transformation becomes larger as the size of the

correlations increase, although there is little bias until *r* > .3.

(2) $\sum(Nj - 3)(z_j - Mean\ z)^2$ distributed as χ^2 with K – 1 df.

Where z_j is the transformed r and $Mean\ z = \sum(Nj - 3)z_j / \sum(Nj - 3)$

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, Hedges,

Higgins and Rothstein, 2009; Cochran 1950). Another statistic examined is the I² 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). Geyeskens et. al (2009) discuss several other homogeneity tests and suggest that 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, researchers should assess the effects of possible moderators (theoretical and methodological).

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 one 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 that excludes 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, Corey and Dunlap (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 that 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 mechanism 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).

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, the 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 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, but, 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. Therefore, 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, clearly as 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, Monroe, and Avila (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 be even 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 studies through perusal of dissertations and calls for studies through various listservs and other such means. In Table 2, 26 out 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 Duwal and Tweedie's Trim and Fill procedures (Duwal and Tweedie 2000).

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

Duwal and Tweedie's Fill and Trim procedure. The Duwal 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 trimming procedure ends up reducing the variance. Therefore. 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).

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. This is often referred to as a "subgroup" analysis within a meta-analysis. 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 relative 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 results of 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 associated with a given relationship (Borenstein et al. 2009; Lipsey and Wilson 2001). This technique is typically called meta-regression.

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 can always 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.

Meta Analytic Structural Equation Models

As we survey the meta-analytic papers in Appendix 1, we note 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., Chueng 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 Kurster 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 and colleagues (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., Ruber 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

In 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).

Meta-analysis provides an orderly procedure for integrating past research, resolving apparent inconsistencies, highlighting roles of important moderators, explicating underlying processes, and promoting innovative research on outstanding issues in a given 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 summarized many of these key decisions here but also point readers to other pertinent 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.

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 (February), 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. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity, *Research Synthesis Methods*, 8 (March), 5-18.

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. (2009). *Introduction to Meta-Analysis,* Wiley: Chichester.

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* (February), 63-77.

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 (February), 114-126.

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

Campbell, J. P., Daft, R. L. & Hulin, C. L. (1982), *What to Study: Generating and Developing Research Questions,* Beverly Hills, CA: 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 (February), 864-76.

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

Cheung, M.W.-L. (2015). *Meta-Analysis: A Structural Equation Modeling Approach. Wiley.*

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 (November), 360-375.

Churchill, G.A., Jr., Ford, N.M., Walker, S.W., & Walker, O.C., Jr. (1985). The determinants of salesperson performance: A meta-analysis. *Journal of Marketing Research*, 22 (May), 103-18.

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), 257-73.

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), 578-96.

Duval, S. J., & Tweedie, R. L. (2000b). 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, 418-40.

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), 906-21.

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), 341-55.

Farley, J.U., Lehman, D. R., & Ryan, M. J. (1981). Generalizing from 'imperfect' replication. *Journal of Business*, 54 (October); 597-610.

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 (May), 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., 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.

Geyskens, I., Steenkamp, Jan-Benedict E. M., & Kumar, N. (1999). A meta-analysis of satisfaction in marketing channel relationships. *Journal of Marketing Research*, 36 (May), 223-38.

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, CA: *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 (October), 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, 166-73.

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 (December), 268-84.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, New York, New York, USA.

Henard, D.H., & Szymanski, D.M. (2001). Why some new products are more successful than others. *Journal of Marketing Research*, 38, 362-75.

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.

Huffcutt, Allen I., and Winfred Arthur (1995). "Development of a new outlier statistic for meta-analytic data." *Journal of Applied Psychology*, 89 (2) 327-334.

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.

Hunter, J. E., & Schmidt, F. L. 1990. *Methods of Meta-Analysis. Correcting Error and Bias in Research Findings.* Thousand Oaks, CA: 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, CA: Sage Publications, Inc.

Hunter, J. E., Schmidt, F.L., & Jackson, G.B. (1982). *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, CA: Sage Publications, Inc.

Jak, S. (2015). *Meta-analytic Structural Equation Modelling*. 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, 138-49.

Keller, P.A., & Lehmann, D.R. (2008). Designing effective health communications: A meta-analysis. *Journal of Public Policy & Marketing*, 27 (2), 117-30.

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 (April), 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 (July), 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, 101-18.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications*,* 49.

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), 247-72.

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), 169-86.

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 (October), 136-53.

Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review Articles: Purpose, Process, and Structure. *Journal of the Academy of Marketing Science*. Forthcoming.

Peter, J. P., & Churchill, G.A., Jr. (1986). Relationships among research design choices and psychometric properties of rating scales: A meta-analysis. *Journal of Marketing Research*, 23 (February), 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., & Beltramini, R.F. (1985). A meta-analysis of effects sizes in consumer behavior experiments. *Journal of Consumer Research*, 12 (June), 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. & Richard J. (1980). Synthesizing outcomes: How to use research evidence from many studies. *Harvard Educational Review,* 50 (May), 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 (August), 351-57.

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 (June), 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), San Francisco: Jossey-Bass, 33-46.

Rosenthal, R. (1982). Valid interpretation of quantitative research results, in D. Brinberg and L. Kidder (eds.), *New Directions for Methodology of Social and Behavioral Science: Forms of Validity in Research*, (12), San Francisco: Jossey-Bass, 59-75.

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Research*, Beverly Hills, CA: 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 (May), 130-47.

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), 409-25.

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 (December), 325-43.

Spangenberg, E.R., & Greenwald, A.G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology*, 8 (1), 61-89.

Steenkamp, Jan-Benedict 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, 252-70.

Sultan, F., Farley, J.U., & Lehmann, D.R. (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research*, 27 (February), 70-77.

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., & Busch, P.S. (1987). Identifying the generics-prone consumer: A meta-analysis. *Journal of Marketing Research*, 24 (November), 425-31.

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 (November), 331-341.

Tellis, G.J., & Wernerfelt, B. (1987). Competitive price and quality under asymmetric information. *Marketing Science*, 6 (3), 240-53.

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 (November), 132-46.

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), 530-54.

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), 407-28.

Viswesvaran, C., & Ones, D.S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48 (4), 865-885.

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 (December), 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), 101-12.

Wright, M., & MacRae, M. (2007). Bias and variability in purchase intention scales. *Journal of the Academy of Marketing Science*, 35, 617-24.

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 (February), 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 (May), 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* (May), 140-154.

Table 1

Type 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, Ford, Walker and Walker (1985) Fern, Monroe and Avila (1986) Brown and Peterson (1993) Franke and Parke (2006) Scheer, Miao, & Palmatier (2015)
Type 2	Replication Analysis	Analysis of key statistics from past studies in a domain to explore the role of certain factors	20	Farley, Lehmann, & Ryan (1981) Sultan, Farley, & Lehmann (1990) Keller & 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.

Table 2

Summary of Meta-Analytic Choices – For Type 1 Meta-Analyses

Meta-Analytic Choices	Levels	1980-1989	1990-1999	2000-2009	2010-present	Total Frequency
		N=8	N=9	N=15	N=20	N=52
Effect Size	r/eta/phi	7	7	14	16	44
	d		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

Note: 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.

Table 3
Potential Criteria that should be reported

Meta-Analytic Trade-off	Explicitly Report the Criteria Used
Sample Domain	• Make explicit what is the domain and what studies are not being included and why
Search Strategy	• Provide details about how papers were obtained and what efforts were made to obtain unpublished papers
Effect Size	• 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	• 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?	• 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?	• BESD can be reported to increase substantive implications of results.
Are the effect sizes homogenous?	• Is one of more tests for homogeneity reported?
Was outlier analysis conducted?	• Report outlier analysis and what method was used. • Also, was analysis done with and without outliers?
Was a fixed effect model or a random effect model used?	• Report the type of model used and why.
Was publication bias tested?	• What statistic was reported?
Was the role of moderators examined?	• Was sub-group analysis conducted? • Was the effect of the various moderators assessed using meta-regression procedures?
Was a structural meta-model reported?	• How were the effect sizes for each correlation calculated? • How was the N used for the model obtained?

Table 4
SAMPLE CODING GUIDE

Promotion and Prevention Fit Coding

Construct	Definition	Example
Promotion Fit	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.	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.
Prevention Fit	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.	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.

Dependent Variable Coding

Construct	Definition	Example
Evaluation	Refers to dependent measures that assess participants' attitudes or reactions to a product, a message or other stimuli.	Asking part pants to indicate liking of a brand on a 7-point Likert scale.
Behavioral Intention	Refers to dependent measures that assess participants' willingness to engage in a behavior.	Having participants report one's willingness to pay for a product.
Behavior	Refers to dependent measures that capture participants' actual behavior.	Observing whether participants choose more of a given product.

Moderator *Variable Coding*

Source of Regulatory Focus		
Construct	Definition	Example
Chronic	Participants' predisposition toward a focus on positive or negative outcomes.	Regulatory focus is captured via standardized measures.
Self-Generated	Momentary regulatory focus is primed by having participants reflect on something that is of or related to the self.	Participants are asked to reflect on their own lives or imagine themselves in a situation as part of the prime.
Situation-Generated	Momentary regulatory focus is primed by having participants reflect on something that is external to the self.	Participants are exposed to stimuli or a context that primes a given regulatory focus.
Type of Subject		
Construct	Definition	Example
Student	Refers to studies in which participants are listed as undergraduates or students aged 22 or younger.	Having students participate for course credit.
Non-Student	Refers to studies in which any other participant type was used.	Using a community sample that might recruited through a local newspaper.

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

Figure 1: A Potential Meta-Decision Flow Chart for Type 1 Meta-Analysis
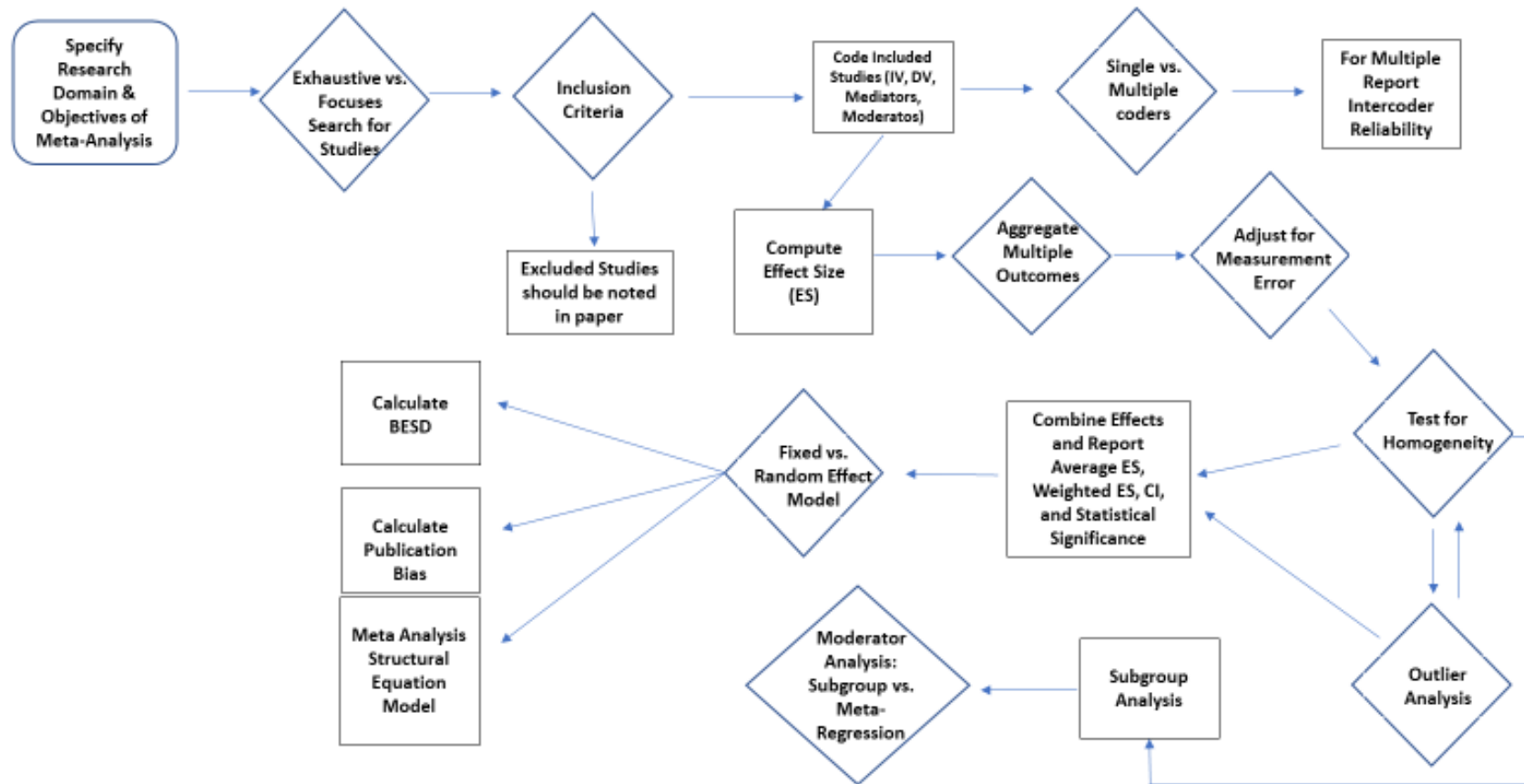
Appendix 1: Review of Meta-Analyses

Citation	Type	Substantive issue	Domain	Time Period	# Studies/ Samples	coders	# Effect Sizes	Effect Size Type	Used Weighting	ES Reliability Adjusted	Test Homogeneity	Test Publication Bias	Assess Outliers
Farley, Lehmann, & Ryan (1981)	2	Behavioral Intention	Consumer Behavior	1966-1978	37	NR	NR	reg. coeff. and R	no	no	no	no	no
Yu & Cooper (1983)	2	Response rates	Survey	1965-1981	93	NR	497	respone rate %	yes	no	no	no	no
Assmus, Farley, & Lehmann (1984)	2	Effect of Advertising on Sales	Communications	1964-1980	22	NR	128	elasticity	no	no	no	no	no
Churchill & Peter (1984)	2	Research Design	Mmethod	1964-1982	101	2	154	reliability	no	no	no	no	no
Churchill, Ford, Walker, & Walker (1985)	1	Salesperson Performance	Sales	1918-1982	116	1 (all), 2 (5%)	1653	r	yes	yes	no	no	no
Peterson, Albaum, & Beltramini (1985)	1	Effects Sizes in Consumer Behavior	Consumer Behavior	1970-1982	118	NR	1036	omega squared	no	no	no	no	no
Fern, Monroe, & Avila (1986)	1	Request strategy	Sales	NR-1986	59	NR	77	Phi	yes	no	yes	yes	no
Peter & Churchill, Jr. (1986)	1	Rating Scales	Consumer Behavior	NR-1985	NR	2	162	r	no	no	no	no	no
Szymanski & Busch (1987)	1	Generics-Prone Consumer	Product Development	1976-1986	24	1 full, 1 (25%), agree = 92%	180	r	no	no	no	no	no
Tellis & Wernerfelt (1987)	1	Price Perception	Pricing	1939-1980	9	NR	1271	r	yes	no	yes	no	no
Sheppard, Hartwick, & Warshaw (1988)	1	Theory of Reasoned Action	Consumer Behavior	1968-1985	87 each	2, agree > .90	174	r	yes	no	no	no	no
Tellis (1988)	2	Price Elasticity	Pricing	1961-1985	42	NR	367	elasticity	no	no	no	no	no
Rao & Monroe (1989)	1	Perceptions of Product Quality	Product Development	NR-1988	36	NR	85	Eta	yes	no	yes	no	yes
Sultan, Farley,& Lehmann (1990)	2	Diffusion Models	Product Development	1950s-1980s	15	NR	213	parameters (P & Q)	yes	no	no	no	no
Brown & Stayman (1992)	1	Ad Attitude	Communications	1982-1991	47	2, agree = 95%	114	r	no	yes	yes	no	yes
Brown & Peterson (1993)	1	Salesperson Satisfaction	Sales	NR-1992	59	2, agree=100%	254	r	yes	yes	yes	no	yes
Wilson & Sherrell (1993)	1	Source Effects on Persuasion	Consumer Behavior	1950-1990	114	3, agree=94%	745	omega squared	no	no	no	yes	no
Peterson (1994)	2	Alpha reliability	Method	1960-1992	1030	2, agree = 92%	4286	alpha	no	no	no	no	no
Heath & Chatterjee (1995)	2	Decoy Effects on Brand Share	Advertising	1982-1994	15	NR	92	range increase	yes	no	no	no	no

Citation	Type	Substantive issue	Domain	Time Period	# Studies/ Samples	coders	# Effect Sizes	Effect Size Type	Used Weighting	Reliability Adjusted	Test Hetero-geneity	Publication Bias	Assess Outliers
Cox III, Wogalter, Stokes, & Murff (1997)	2	Product Warnings	Product, public policy	1976-1995	15	2	79	compliance rate	yes	no	no	Yes	no
Grewal, Kavanoor, Fern, Costley, & Barnes (1997)	1	Noncomparative Advertising	Advertising	1975-1996	77	NR	262	d	yes	yes	yes	yes	no
Brown, Homer, & Inmann (1998)	1	Ad-evoked Feelings and Advertising Response	Communications	1986-1995	55	1, verified by mailing authors	134	r	yes	yes	yes	no	no
Compeau & Grewal (1998)	1	Comparative Price Advertising	Pricing	1974-1992	38	NR	86	Eta	no	no	yes	yes	Yes
Notani (1998)	1	Theory of Planned Behavior	Consumer Behavior	1989-1995	36	2, agree > 85%.	63	r	yes	no	yes	no	yes
Geyskens, Steenkamp, & Kumar (1999)	1	Marketing Channel Relationships	Sales	1970-1996	107	NR	NR	r	yes	no	no	no	no
Spangenberg & Greenwald (1999)	1	Social Influence	Consumer Behavior	1980-1997	11	NR	11	r	no	no	yes	no	no
Estelami & Lehmann (2001)	2	Consumer Price Recall	Pricing	1961-1999	22	NR	279	average deviation	no	no	no	no	no
Estelami, Lehmann, & Holden (2001)	2	Consumer Price Knowledge	Pricing	1961-1999	27	NR	297	Price Recall Error	no	no	no	no	no
Henard & Szymanski (2001)	1	New product performance	Product Development	NR-1999	41	2, agree=98.2%	798	r	yes	yes	yes	yes	no
Peterson (2001)	3	Student Samples	Method	1981-2000	28	NR	63	r, d, g, β, ω^2	no	no	no	no	no
Szymanski & Henard (2001)	1	Customer Satisfaction	Consumer Behavior	NR-1986	50	2, agree = 96%	517	r	yes	yes	yes	yes	no
Krishna, Briesch, Lehmann, & Yuan (2002)	2	Price Presentation	Pricing	1980-1998	30	2	345	% saving	no	no	no	no	no
Janiszewski, Noel, & Sawyer (2003)	1	Spacing Effect in Verbal Learning	Consumer Behavior	1887-2000	61	NR	269	r	no	no	no	yes	no
Van den Bulte & Stremersch (2004)	2	New Product Diffusion	Product Development	1969-2000	46	NR	293 observations	q/p ratio	no	no	no	no	no
Kirca, Jayachandran, & Bearden (2005)	1	Market Orientation	Strategy	-2004	studies/ 130	(35%), Reliability: .91-	418	r	yes	yes	no	Yes	no
Conchar, Crask, & Zinkhan (2005)	2	Advertising and promotion spend on market value of the firm	Strategy	1978-2004	88	NR	88	A&P Effect	yes	no	yes	yes	no

Citation	Type	Substantive issue	Domain	Time Period	# Studies/ Samples	coders	# Effect Sizes	Effect Size Type	Used Weighting	ES Reliability Adjusted	Test Hetero-geneity	Test Publication Bias	Assess Outliers
Franke & Park (2006)	1	Salesperson Behavior	Sales	1979-2006	155	2	NR	r	no	yes	yes	no	no
Palmatier, Dant, Grewal, & Evans (2006)	1	Marketing Improves Performance	Sales	1987–2004	111	2, agree > 95%	637	r	yes	yes	yes	yes	no
Szymanski, Kroff, & Troy (2007)	1	Innovativeness and New Product Success	Product Development	-2006(start not reported)	32	2, agree = 95%	95	r	yes	yes	yes	no	yes
Wright & MacRae (2007)	1	Purchase Intention Scales	Consumer Behavior	-2006 (not reported)	9	NR	52 proportion estimates	Logs Odd Ratio	yes	no	yes	no	yes
Carlson, Vincent, Hardesty, & Bearden (2008)	1	Objective and Subjective Knowledge	Consumer Behavior	1980-2007	51	2, agree = 84.4%	104	r	yes	yes	yes	yes	yes
Crosno & Dahlstrom (2008)	1	Opportunism	Channels	1982-2005	59	1	183	r	yes	no	yes	no	no
De Matos & Rossi (2008)	1	Word-of-mouth communications	Communications	1955-2007	162	NR	348	r	yes	yes	yes	yes	yes
Keller & Lehmann (2008)	2	Effective Health Communications	Communications	1961-2006	60	2	NR	(0.0 to 1.0 scale)	no	no	no	no	no
Krasnikov & Jayachandran (2008)	1	Firm Performance	Strategy	-2007 (not reported)	114	2	786	r	yes	yes	No	yes	no
Troy, Hirunyawipada, & Paswan (2008)	1	New Product Success	Product Development	- 2007 (start not reported)	25	2 coded 50%, cross-checked	146	r	yes	yes	yes	no	no
Grinstein (2008)	1	Market Orientation on Innovation	Product Development	1994-2006	56	NR	69	r	yes	yes	yes	no	yes
Eisend (2009)	1	Humor in Advertising	Communications	1960-2006	43	2, k > .9	369	r	yes	yes	yes	yes	no
Keller, Lehmann, & Milligan (2009)	2	Corporate Well-Being Program	Communications	1982-2006	60	2	76 observations	% cost reduction	no	no	no	no	no
Eisend (2010)	1	Gender Roles in Advertising	Communications	1975-2007	64	2	NR	odds-ratio	yes	yes	yes	no	yes
Orsingher, Valentini, & Angelis (2010)	1	Satisfaction With Complaint Handling	Consumer Behavior	-1997	60	NR	509	r	yes	yes	yes	yes	yes
Scheibehenne, Greifeneder, & Todd (2010)	1	Choice Overload	Consumer Behavior	2000-2009	50	NR	63 datapoints	d	Yes	no	yes	no	yes
Albers, Mantrala, & Sridhar (2010)	2	Personal Selling	Sales	-2010	88	2, agree > 90%	506	elasticity	no	no	no	no	no
Eisend & Küster (2011)	1	Publicity vs. Advertising	Communications	1971-2009	30	NR	NR	r	yes	yes	Yes	no	no
Sethuraman, Tellis, & Briesch (2011)	2	Brand Advertising Elasticity	Advertising	1960-2008	56	NR	751 short term; 402 long term	elasticity	no	no	no	no	no
Verbeke, Dietz, & Verwaal (2011)	1	Sales Performance	Sales	1982–2008	292	(20%), agree > 95%	2105	r	yes	yes	yes	yes	yes

Citation	Type	Substantive issue	Domain	Time Period	# Studies/ Samples	coders	# Effect Sizes	Effect Size Type	Used Weighting	ES Reliability Adjusted	Test Hetero-geneity	Test Publication Bias	Assess Outliers
Rubera & Kirca (2012)	1	Performance Outcomes	Product Development	-2010 (start not reported)	159	studies), agree = 90%	NR	r	yes	yes	yes	yes	no
Homburg, Klarmann, Reimann, & Schilke (2012)	1	Key informant accuracy	Method	1998-2010	127	4, reliability .71-.86	11874	r, agreements	yes	yes	no	Yes	no
Steenkamp & Geyskens (2012)	1	Transaction Cost Economics	Sales	-2011 E77	studies/13 independe	NR	211	r	no	yes	no	no	yes
Zablah, Franke, Brown, & Bartholomew (2012)	1	Customer Orientation	Sales	1979-2011	323	1, multiple for moderators	NR	r	no	yes	no	no	no
Motyka, Grewal, Puccinelli, Roggeveen, Avnet, Daryanto, Ruyter, & Wetzels, (2014)	1	Regulatory Fit Leads to More Favorable Consumer Response	Consumer Behavior	1998–2013	215	2, agree > 90%	306	r	yes	yes	yes	yes	no
Van Laer, Ruyter, Visconti, & Wetzels (2014)	1	Transportation Imagery	Consumer Behavior	NR+E2-2014	76	2, k=.75	132	r	yes	yes	no	yes	no
Pick & Eisend (2014)	1	switching costs	loyalty management	NR-2012	170	2, agree= 96%	800	r	yes	yes	yes	no	no
Zlateveska, Dubelaar, & Holden (2014)	1	Portion size	Packaging	1994-2013	104	NR	104	d	no	no	yes	yes	no
Eisend (2015)	1	Method Factors	Method	1976-2012	176	Domain: 2, agree-95% Other 2nd coded 25%,agree=100%	1841	r	no	no	no	yes	no
Scheer, Miao, & Palmatier (2015)	1	Dependence	Channels	1970-2014	218	2, agree= 96%	976	r	yes	yes	yes	yes	no
Crosno & Brown (2015)	1	Organizational Control and Channels	Channels	1975-2010	66	2, agree > 98%	165	r	yes	yes	yes	yes	yes
Watson, Beck, Henderson, & Palmatier (2015)	1	Loyalty	Loyalty	1980-2013	126/151	2, agree > 96%	713	r	yes	yes	yes	no	no
You, Vadakkepatt, & Joshi (2015)	2	Electronic Word of Mouth	Loyalty	NR-2015	51	2, k=.85	339 + 271	elasticity	no	no	no	no	yes
Rosario, Sotgiu, Valck, & Bijmolt (2016)	1	Electronic Word of Mouth	Loyalty	1999-2013	96	1 (all), 1 (subset), k=.97	1532	r	yes	no	yes	yes	yes
Chang & Taylor (2016)	1	Customer Participation in NPD	New Product Development	1998-2014	39	2, k > 82%	123	r	yes	yes	yes	yes	yes
Edeling & Fischer (2016)	2	Firm value	Strategy	1977-2013	83	1 (all), 1 (30), agree > 90%	488	elasticity	yes	no	yes	no	yes
Hogreve, Iseke, Derfuss, & Eller (2017)	1	Service Profit	Services	1994-2015	518	3, agree = 93%	687	r	yes	yes	no	yes	yes
Knoll & Matthes (2017)	1	celebrity endorsement	advertising	NR-2016	46	2, k=.74	279	d	yes	no	no	yes	yes