La méthodologie de la méta-analyse en marketing

Patrice Laroche, Sébastien Soulez

▶ To cite this version:

Patrice Laroche, Sébastien Soulez. La méthodologie de la méta-analyse en marketing. Recherche et Applications en Marketing (French Edition), 2012, 27 (1), pp.79-105. 10.1177/076737011202700104 . hal-00923909

HAL Id: hal-00923909
https://hal.science/hal-00923909

Submitted on 21 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Meta-analysis for Marketing Research

Patrice Laroche

Université de Lorraine
ISAM-IAE Nancy
CEREFIGE

Sébastien Soulez

Université de Haute-Alsace
IUT de Colmar - Département TC
CEREFIGE

ABSTRACT

The aim of this paper is to present and illustrate the procedure to follow when a researcher wants to use meta-analysis in marketing research. Meta-analysis is the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings. Meta-analysis offers new opportunities for integrating and combining the contradictory outcomes of studies and for analyzing variance in effect sizes across findings.
After a description of the different stages of meta-analysis, two applications of meta-analytic procedure in marketing are presented and recommendations are suggested for marketing researchers who want to perform a meta-analysis accurately.

Keywords: meta-analysis, marketing, methodology, publication bias, literature review.

The authors can be contacted at the following email addresses:
Patrice.Laroche@univ-lorraine.fr; *sebastien.soulez@uha.fr*

INTRODUCTION

Today researchers in marketing must cope with a growing body of empirical studies in many areas and this makes it increasingly difficult to provide a traditional review of the literature or to highlight "a pattern that repeats but need not be universal over all circumstances" (Bass, 1995) or "empirical regularities that not only recur over time and space, but which are also understood in terms of theory such that the conditions under which they exist can be specified" (Fornell, 1995) or, in other words, empirical generalizations. When a researcher wants to develop an empirical generalization he/she can use one of four possible approaches: a traditional review of the literature, a meta-analysis, content analysis and classification or he/she can search for invariants or irregularities by exploring multiple data sets (Bass, 1995). This pedagogical article presents one of these methods, i.e. meta-analysis, which has, with time, become the dominant method for reviewing scientific literature (Aguinis *et al.*, 2011). This methodology may seem less prevalent in management science, even though we can find meta-analyses in all of its sub-fields: finance (Coggin, Fabozzi and Rahman, 1993), accounting/control (Pomeroy and Thornton, 2008), HR (Subramony, 2009), strategic management (Geyskens, Steenkamp and Kumar, 2006) or information systems (Wu and Lederer, 2009). The significant place reserved for meta-analyses in academic journals demonstrates an increased interest in the method in the field of marketing (Franke, 2001). A bibliographical study of journals in management science indicates that 105 meta-analyses were published between 1976 and 2011, with 23 of 34 journals in marketing (67%) publishing at least one meta-analysis (cf. Appendix A1)[1]. Historically, the first meta-analyses in marketing have examined the influence of advertising on sales (Clarke, 1976; Assmus, Farley and Lehmann, 1984), the influence of research design on the reliability of scales (Churchill and Peter, 1984) or questionnaire response rates (Yu and Cooper, 1983), the determinants of sales staff performance (Churchill *et al.*, 1985) or consumer behavior patterns (Farley, Lehmann and Ryan, 1982). However, all fields of marketing are concerned by meta-analysis. Thus, since 1976, 59 meta-analyses have been published in operational marketing (including 25 in communications alone), 15 in strategic marketing, 19 in consumer behavior and 12 on methodological issues.

Meta-analysis is "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976). It is therefore a methodological approach that can be used to combine the results of individual studies on the

[1]The ranking of journals proposed in Appendix A1 is the one published in 2008 by section 37 of the CNRS (French National Center for Scientific Research).

same subject in order to produce a quantified and reproducible synthesis. Among the methods researchers can use to summarize the literature, meta-analysis has the advantage of reducing the arbitrary elements of traditional narrative reviews to a strict minimum through a systematic and reproducible methodology in such a way that another researcher with access to the same data can replicate it and arrive at the same conclusions (Fournier and Vauquois-Mathevet, 1999).

Initial applications of meta-analysis have above all concerned the fields of psychology (Glass, 1976) and medicine (Leizorovicz and Boissel, 1983), but, over the last thirty years, its application has been extended to other areas of research and particularly to marketing (Farley, Lehmann and Sawyer, 1995), while it has continued to benefit from methodological improvements and recommendations (Bijmolt and Pieters, 2001; Geyskens *et al.*, 2009; Carlson and Xiaoying Ji, 2010). However, meta-analysis is not often employed by French researchers, especially in marketing. Consequently, this pedagogical and methodological article has a dual ambition: to demonstrate and illustrate the interest of meta-analysis in the field of marketing and to offer recommendations for researchers in this field who wish to use it.

The purpose of this article is therefore to present and illustrate the appropriate methodological approach when conducting a meta-analysis by highlighting the contributions and limitations of this means for synthesizing marketing literature. First, it offers an overview of existing tools and procedures, underlining the challenges and debates that characterize the choice of techniques used to conduct a meta-analysis (1). Then it presents two applications in marketing, before proposing recommendations for researchers interested in using this method (2).

THE MAIN STAGES OF META-ANALYSIS

The possibility of summarizing a set of empirical results, particularly when they are contradictory, can explain the growing interest in meta-analysis. The general principle is based on the hypothesis that the incidence of one variable on another is a constant and that, consequently, each study measures the same constant. Therefore, different results observed in individual studies should only stem from random fluctuations (measurement or sampling errors).

[INSERT BOX 1]

Conducting a meta-analysis means following a specific procedure whose main stages are summarized below.

[INSERT TABLE 1]

Stage 1: Formulating a research question

The preliminary stage consists in acquiring a good understanding of the research problem and knowledge of the challenges it involves. This initial stage, as is the case for all scientific research, is used to identify a problem and justify the interest of using a meta-analysis. If the problem concerns, for example, the link between price and perceived quality, the meta-analyst must first decide on the meaning of these two concepts, the definition of variables and the inclusion, or not, of possible mediating and/or moderating effects (Völckner and Hofman, 2007). These different questions are important in that they allow the meta-analyst to select the studies and identify the information required to conduct the meta-analysis.

Stage 2: Gathering studies

After formulating a specific research question, the second stage of a meta-analysis consists in an exhaustive search for all the existing studies on subject. Computerized databases have become, over the last few years, essential tools in bibliographical research. There are many database listings of marketing studies and it is strongly recommended to consult them all for a more thorough analysis. *ABI Inform/Proquest, JStor, Ebsco, ScienceDirect* and *Emerald* offer direct access to a wide range of journals in management science with the possibility of downloading articles in PDF format. There are other databases that are less practical to access to the extent they only provide article references or abstracts (*Social Science Citation Index* (SSCI), *EconLit, Doge, Current Contents, Management Contents, Journal of Economic Literature, Marketing Abstracts,* etc.), but they can be very useful in identifying the existing literature. Today it is also very easy to complete this bibliographical research by using search engines such as *Google Scholar, EconPapers, SSRN, Scientific Commons,* etc. (after carefully choosing a series of keywords). Furthermore, it is essential to consider other sources of information. Consulting the references in the articles already collected, searching manually through the table of contents in journals or collective works and contacting specialists in the field are all complementary means of learning about existing literature on the subject.

Stage 3: Selecting studies

Before starting the coding process, it is important to define the criteria for inclusion, or exclusion, of a study in the meta-analysis. Identifying inclusion criteria involves, in particular, the question of the study's quality. Indeed, introducing a poor quality study may reduce the quality of the meta-analysis, according to the principle of garbage in/garbage out. Specialists therefore recommend excluding "poor quality" studies. But how can one determine the intrinsic quality of a study? Some would suggest eliminating all unpublished documents that have not been peer-reviewed. Others propose using the *ISI Web of Knowledge* and *Journal Citation Reports* (*JCR*®) databases, which offer a ranking of more than 1,600 international journals in social sciences. Then it is possible to use the *JCR*® score for each publication as weights (Doucouliagos and Laroche, 2003). Still others would examine the internal and external validity of each study in depth (Cooper, 2010). In addition to differences in quality between studies, the meta-analysis can detect a certain number of articles that do not have the necessary characteristics to be included in the final sample. Thus publications that rely on the results of the same study must be eliminated in order to avoid *over-representation bias*. It is not rare, in fact, for a researcher to publish several articles based on the same data. Furthermore, studies that do not provide enough information to calculate a common metric should also be excluded from the meta-analysis. The same goes for studies based on highly different research methodologies, even if they are on the same subject. It is not possible, for example, to combine the results of a logistic regression (*logit* or *probit*) with those of a linear one. Meticulous selection of the studies offers an effective response to a frequent criticism regarding meta-analyses, i.e. of "mixing apples and oranges" (Glass, McGaw and Smith, 1981), and ensures its validity and reliability.

Stage 4: Gathering data and coding the selected studies

After selecting a set of existing studies, the next stage consists in establishing a coding grid. A coding grid is a meta-analysis tool used to gather all the data from the selected studies. Most data is from the studies themselves, but there are other sources that can also be included in order to go beyond simple replication. When the number of studies is large, creating a coding grid can be an arduous task. Indeed, it involves establishing a list of all the characteristics of all the studies the meta-analyst wishes to include. This list can contain many variables likely to be associated with the results of the study (Reisinger, 1997). While the content of a coding grid will vary from one meta-analysis to another, it often contains the same data that can be classified in different categories (Cooper, 2010): the publication's characteristics, the experimental conditions of the study, the nature of the variables of interest and the way they are measured, the methodology used, the statistical results obtained and, in certain cases, information on the coding process. These different characteristics can be recorded in a spreadsheet program in order to organize statistical processing of the meta-analysis more easily. Then the meta-analyst should draw up a well-argued list of potential moderators that will be considered if the results of the studies turn out to be highly heterogeneous.

Stage 5: Analyzing the data

Once the studies have been coded, the data can be analyzed. At this stage, the data extracted from the studies can serve as the basis for various calculations to obtain a summary of existing results in the literature. This process involves three steps: (1) selecting a common metric (2) combining these effect sizes and (3) assessing the heterogeneity of the effect sizes in each study. The *effect size* is defined by estimating the degree of the relationship between two variables of interest. Today, researchers in social sciences focus more on the intensity or strength of a relationship between two phenomena than on the existence of the relationship alone. In light of these questions, meta-analysts use much more sophisticated methods for calculating effect sizes than those initially used to summarize the literature (such as a vote-counting procedure).

Choosing a common metric

There are several techniques designed to transform statistics from the studies collected in order to calculate a common metric - or standardized effect size - that will be used to integrate the results and calculate the mean effect size (Wolf, 1986, p. 35). These methods diverge according to the type of variable (continuous variables, ordinal or nominal categorical variables) and the type of study available (experimental or correlational research).

[INSERT BOX 2 HERE]

Combining "standardized" effect sizes

Once standardized effect sizes have been calculated for each study they can be combined. There are, in fact, two methods for aggregating "standardized" effects. The first is called the "fixed effects" method and the second is referred to as the "random effects" method (Erez, Bloom and Wells, 1996). The *fixed effects method* hypothesizes that there is a certain homogeneity between studies, i.e. that the expected theoretical effect for each one is the same. The only variance in results observed between two studies stems from random variations around this mean common effect, due to measurement or sampling errors. Then the general formula for the average effect size is equal to the mean of effect size observed in each study, sometimes weighted by the inverse of their estimated variance w_i and its 95% confidence interval (Hedges and Olkin, 1985).

The hypothesis of the fixed effect model is relatively strong and does not allow for variations in effect size according to the characteristics of the selected studies (business sector, type of econometric specification, etc.) Another approach, which might seem more realistic, proposes considering the variability of differences in effects observed across studies, by supposing that a random effect specific to each study is added to the above-mentioned common effect. This is referred to as the *random effects model* and includes a random component in the variance of effects (Raudenbush, 1994; Hedges and Vevea, 1998). In this method, the estimation of common effect size is a weighted mean of effects observed in each study, by integrating a term representing inter-study variability in addition to intra-study variability. In the end, the estimation of common effect size with either of these methods (fixed effects or random effects) is practically the same, but the confidence interval is different. The one obtained with the random effects method is greater and can lead to different conclusions regarding the significance of the common effect.

[INSERT BOX 3 HERE]

Evaluating heterogeneity of the results

Once the previous stages have been completed, the meta-analyst has a central trend indicator that can be used to identify the nature of the relationship (positive or negative) between two variables of interest and its intensity. However, estimation of the sampling and measurement error can help determine whether the studies share a common effect size for the population, i.e. if the results are homogeneous. The terms homogeneity test or heterogeneity test are used indifferently. The most commonly used methods are those recommended by Hedges and Olkin (1985) and Hunter, Schmidt and Jackson (1982).

[INSERT BOXES 4 AND 5]

Heterogeneity comprises two sources of dissimilarity: on the one hand, a dissimilarity in the results of individual studies, which is statistical in nature and can be explored using the statistical methods mentioned above, and, on the other hand, a dissimilarity in the very organization of the existing studies, which is either contextual or methodological. This implies answering the following question: *Which contextual factors and/or methodological features of empirical studies can explain the heterogeneity of these results?*

When a case of heterogeneity has been discovered, it is important to search for its sources. Identifying key factors linked to the structure of the study, to the type of population examined or the relationship tested, is an important stage (Aguinis and Pierce, 1998; Cortina, 2003). There are two possible methodologies the meta-analyst can use. The first consists in conducting meta-analyses in subgroups, or a stratified meta-analysis. The second method consists in a regressing the supposed moderating variables of effect size calculated for each study.

Sub-group analyses can be used to search for sources of heterogeneity in a univariate manner, by comparing results obtained between sub-groups of studies (Muller, 1988). Sub-groups are formed according to the supposed moderating factors. For example, gender can constitute a relevant explanatory variable for differences in purchase behaviors regarding a product or service. Analyses of sub-groups can lead to an uncontrolled inflation of a Type I error a . Multiplication of statistical tests (one per sub-group) increases the probability of obtaining a significant test exclusively by accident. To reduce the risk of accidental significant results in sub-group analyses, it is best to define, *a priori*, a small number of sub-groups. This approach would be similar to a hypothetical-deductive one. According to this approach, the potential moderating variables are coded on the basis of theoretical justifications.

Regression meta-analysis (or meta-regression) consists in regressing different supposed moderating variables on the "standardized" effect size, calculated for each study (Stanley, 2001). The aim of this modelization is to examine the simultaneous effect of several moderating variables on effect size. In this case, this means estimating a multiple regression model with the following form:

$$Y_i = \alpha + \beta_1 N_i + \gamma_1 X_{i1} + \ldots + \gamma_k X_{ik} + \delta_1 K_{i1} + \ldots + \delta_n K_{in} + u_i$$

where:

Y_i is the size effect of study *i*
α is the constant that can be interpreted here as the "real" effect size
N_i is the size of the sample in study *i*
X is a dummy variable representing certain features associated with study *i*
K is the mean value of a quantitative variable representing other characteristics
u_i is the random disturbance

This methodology often requires coding the moderating variables that seem, *a priori*, to influence effect size in the form of dichotomous or dummy variables (Hunter, Schmidt and Jackson, 1982, p. 119).

Once the methodological framework has been established, it is time to present and interpret the results of the meta-analysis.

Stage 6: Presenting and interpreting the results

The results of the meta-analysis are generally presented as a table that comprises: the number of studies included in the meta-analysis (k), the size of the total sample (N), the estimation of the common effect size obtained by integrating all the data from the studies included in the meta-analysis. This estimation is presented with its confidence interval (generally at the 5% level) and sometimes with the range of effect sizes and, finally, the results of the heterogeneity test. Results of the meta-analysis are also frequently presented as a graph. There are several types of graph used to represent the results of a meta-analysis. The most common form is a line graph that easily presents all the results of the meta-analysis. The effect sizes of each study and estimated effect within the population are represented on the same graph by a series of dots (square or vertical bar) alongside their confidence interval (horizontal line). This frequency distribution of values for the measured effect can be used to determine that the homogeneity hypothesis was not rejected because of a few isolated dots (Light and Pillemer, 1984).

[INSERT FIGURE 1]

In a second stage, it is necessary to present the results of the analysis of moderating factors. One example from the meta-analysis conducted by Peterson and Jolibert (1995), which examines the link between country of origin and quality of purchase intention, is provided in Table 2.

[INSERER TABLE 2]

The purpose of the following section is to illustrate, more specifically, the practical implementation of a meta-analysis and to formulate some recommendations through a detailed examination of two meta-analyses published in marketing.

DEVELOPMENT OF TWO APPLICATIONS IN MARKETING

The object of this section is to illustrate the main stages of a meta-analysis conducted using data from two meta-analyses published in marketing:

- one by Szymanski, Troy and Bharadwaj published in *The Journal of Marketing* in 1995
- one by Trappey published in *Psychology & Marketing* in1996.

The meta-analysis conducted by Szymanski, Troy and Bharadwaj (1995) aims to synthesize research on the effects of order of market entry on sales performance. This meta-analysis is particular in that it integrates studies based on a single model and with strictly similar methods. In this case, the authors have gathered non-standardized regression coefficients from each study in order to calculate a mean and standard deviation. Trappey's meta-analysis (1996), on the other hand, is devoted to the effects of marketing stimuli on consumer choice behaviors and combines a set of studies based on very different statistical methods. One is based on correlational type data (Szymanski, Troy and Bharadwaj, 1995), while the other (Trappey, 1996) relies on experimental data. Each one illustrates different meta-analytical procedures which is why these two studies were selected for this article. Furthermore, these two meta-analyses have the advantage of providing enough information to pick them apart, replicate them and re-process the data, if need be, in order to illustrate the contributions and limitations of meta-analytical tools. Indeed, it is not often that one has access to a list of a study's characteristics and the possibility of examining the methods and procedures adopted in great detail. It is important to emphasize here that these two studies do not demonstrate how the results of a meta-analysis can then be included in a causal analysis as is frequently the case in marketing today (*cf. infra*, p. 20).

Collecting and choosing studies

The first stage in a meta-analysis consists in identifying a field of research that has been sufficiently explored in empirical literature so there are comparable results that are worth synthesizing in order to produce generalizable conclusions. The question of feasibility must be raised by the researcher at the very beginning and is related to the number of studies required to conduct a meta-analysis. The answer to this question depends on the nature of the relationship studied. For example, five similar replications of a study can offer a rather good mean estimation of the parameters of a model while a set of more heterogeneous studies that differ in four or five important aspects will require collecting at least twenty studies with enough differences between them. Therefore there are no strict rules, but it is important to underline two key points when it comes to integrating the results of individual studies. First, there are often fewer available studies than the number envisaged at the outset of the meta-analysis. It is important not to become discouraged if only a dozen more or less comparable empirical studies can be found. Secondly, often there are several estimations provided in the same study because the researchers have tested different statistical models. This is the case, for example, in the two meta-analyses we propose to examine in detail here. Thus, Szymanski, Troy and Bharadwaj (1995) collected 16 empirical studies on order of market entry and its effects on market share (*cf.* Table 3). From these 16 studies they extracted a total of 64 estimations, which all represent measures of effect size between the two variables of interest.

[INSERT TABLE 3]

In principle, a meta-analysis should be based on all existing studies on the same research topic. However, it is often difficult for the meta-analyst to obtain all this research work, especially due to difficulties accessing unpublished research (theses, research journals, conference papers, etc.). In this particular case the authors collected 21 studies published in leading marketing journals, but also two studies published in the form of working papers (Kalyanaram, 1993; Bernt *et al.*, 1994). Regarding this specific point, Cooper (2010) emphasizes that unpublished research is generally of lesser quality than published studies and, consequently, it should not be included in analyses. Of the two working papers included in the meta-analysis, one was published in a collective work several years later (Bernt *et al.*, 1997, in a book entitled *The Economics of New Goods*), while Kalyanaram's study was never published. Cooper's recommendation (2010) seems relevant even if orienting the selection of studies in favor of those already published raises the question of publication bias (Laroche, 2007). Indeed, many studies have shown that research presenting significant results were easier to publish than those presenting results that were not (Rosenthal, 1979; Begg and Berlin, 1988; Rust, Lehmann and Farley, 1990).

Recommendation 1: **Do not "mix apples and oranges"**

It is important to remember that a meta-analysis can only be conducted if the studies collected use similar methodologies and provide the same type of quantitative results. The meta-analyst is not an alchemist who tries to turn "base metals into gold" (Feinstein, 1995, p. 71). Indeed, variability across studies can be observed in their quality, research design and methodologies. Feinstein (1995) warns against the excessive use of meta-analyses to integrate studies of very different levels of quality. At best one can end up with a "*mixed salad*" of "apples and oranges" and at worst "even less savory substances" (Feinstein, 1995). It is therefore important to assess the quality of the studies and select them on this basis. Instead of attributing a subjective score to each study, it is better to select only those published in peer-reviewed journals and to use rankings of marketing publications (CNRS, AERES, ISI *Web of Knowledge*, etc.) to weight the calculated effect sizes. This type of weighting system is not used enough by meta-analysts in marketing, who tend to integrate studies with varying levels of quality. This is the case of the meta-analysis produced by Szymanski, Troy and Bharadwaj (1995).

Furthermore, the authors did cite the criteria for the studies selected by presenting a list of those excluded and the reasons behind their elimination, thus ensuring the internal validity of the meta-analysis. They indicate sources of information that helped identify the studies and the keywords used for the bibliographical search. In this case it is important to eliminate the possibility of a selection process designed according to the expected results of the meta-analysis. In the meta-analysis conducted by Szymanski, Troy and Bharadwaj (1995, p 19), the authors state they only selected studies containing non-standardized regression coefficients and eliminated all those with standardized coefficients, analyses of simple correlations, logistic regression coefficients, etc. Be that as it may, the regression coefficients the authors finally kept are not, in fact, all non-standardized. Furthermore, the elimination of Robinson's study (1988) from the sample simply because it is an outlier from a statistical perspective is not sufficiently justified by the authors.

Recommendation 2: Eliminate the possibility of a selection process governed by expected results

It is advisable to present the selection criteria for the studies in order to avoid using completely arbitrary ones. At this level, a list of studies excluded and the reasons for their rejection can be useful in ensuring the internal validity of the meta-analysis. It is also a good idea to cite the sources of information used to identify the selected studies. One should indicate the keywords used for searching electronic databases and provide the names of these resources, as well as those of the experts consulted. The idea is to make it easier for another researcher to replicate the meta-analysis. Unfortunately very few meta-analyses in marketing provide a detailed list of selection criteria and the reasons behind the elimination of certain studies.

As for Trappey (1996), a bibliographical search was used to gather nine studies on subliminal advertising with a total of twenty-three estimations (*cf.* Table 4). Among these nine studies, seven had been published in peer-reviewed journals (*Journal of Marketing*, *Journal of Advertising*, etc.) and two in conference proceedings (*Advances in Consumer Research* and *Advertising and Consumer Psychology*). We can consider that Trappey (1996) collected comparable scientific studies and did not "mix apples and oranges".

[INSERT TABLE 4]

Calculating a common metric

All the estimations collected from published studies provide observations regarding the variable(s) of interest used in the meta-analysis. Estimations drawn from various studies are generally not measured the same way and a common metric must be used in order to compare them. Often this means converting indicators into correlation coefficients, elasticity, etc. Szymanski, Troy and Bharadwaj (1995) rely on a set of strictly similar data, in this case non-standardized regression coefficients from estimations of regression models designed to study the relationship between a variable of interest measuring order of market entry for a product and another variable explaining market share.

Recommendation 3: Prefer partial correlation coefficients when conducting a meta-analysis based on correlational data

One remark can be made concerning the methodological approach of Szymanski, Troy and Bharadwaj (1995). The choice of, as a common statistical indicator, non-standardized regression coefficients of the variable "order of entry" is highly questionable. Indeed, this coefficient varies according to the other variables introduced in the regression models. However, none of the studies collected estimates the exact same regression model. Some introduce one or two control variables, while others consider a dozen of them. It would have been better to use a partial correlation coefficient in order to determine the value of the correlation between two variables of interest, by holding constant the other variables introduced in the regression (*cf. infra*).

On the other hand, Trappey (1996) relies on a relatively wide range of studies whose results must be transformed into a common metric. In the studies collected by Trappey (1996), the researcher compares two experimental conditions and identifies whether there is, or not, a significant difference between the control and experimental groups with the help of a statistical test. This also involves calculating the intensity of this difference, i.e. the effect size for each existing study. When a meta-analysis concerns a set of studies comparing the means of experimental and control groups, the most commonly used measure of association is Cohen's *d* (1996). Hedges and Olkin propose variations of this statistic (Hedges' *g*, 1981; Hedges and Olkin's *d*, 1985) that are commonly used in the context of marketing research. In his meta-analysis Trappey opts for Cohen's *d*, which can be calculated using several available statistics in existing studies. Thus, Cohen's *d* can be computed using a Chi-square test, a Fisher test, Student's t-test or Pearson's correlation coefficient (cf. Wolf, 1986, p. 35). Replicating Trappey's calculations turn out to be very arduous, when one of the purposes of a meta-analysis is to allow any researcher access to the same data in order to reproduce the meta-analysis and arrive at the same conclusions. Indeed, the author does not provide enough information on how to calculate Cohen's *d*, in particular intra-group aggregated variance (cf. infra), thus making any attempt at replication impossible.

Combining "standardized" effect sizes

Szymanski, Troy and Bharadwaj (1995) choose to calculate two measures of central tendency based on non-standardized regression coefficients gathered from the selected studies. The first is simply the mean of these regression coefficients while the second is a mean weighted according to the sample size of each study. They obtain, respectively, a non-weighted mean of 3.59 (sd = 5.07; n = 64) and a weighted mean of 5.72 (sd = 5.62; n = 61), after eliminating one extreme value and a study that did not provide information on the sample size. They deduce an overall positive effect of order of market entry on market share, while indicating, however, a strong heterogeneity of results.

We have used their data to conduct a series of tests in order to confirm this. Unlike Szymanski, Troy and Bharadwaj (1995), we only kept the published studies and eliminated the two working papers from the sample in keeping with the recommendations of Cooper (2010). The sample therefore comprised fourteen studies instead of sixteen. To be completely thorough we should have added the studies eliminated by Szymanski, Troy and Bharadwaj (1995), in particular the ones including standardized regression coefficients, which the authors voluntarily excluded. Indeed, use of partial correlation coefficients allows the integration of results from analyses of multiple linear regressions, whatever the nature of the regression coefficient may be. For each of these fourteen studies we calculated partial correlation coefficients using Student's t-test or standard errors. Thus we had several coefficients for each study, i.e. as many models as regressions presented in each one[2].

The mean of these partial correlation coefficients is 0.249 and the weighted mean according to sample size in each study is 0.124, which leads us to believe that order of market entry does globally have a positive effect on market share. However, variance linked to sampling error is only 13.3%, which induces strong residual variance (>86%) and confirms the strong heterogeneity of the results in existing studies. These conclusions converge with those of Szymanski, Troy and Bharadwaj (1995) who use a different meta-analytical procedure.

In order to extend the analysis even further, we also tested for publication bias. The problem of publication bias has been widely discussed by specialists and statistical tools have been developed to control for this type of problem (for an overview see Laroche 2007).

Recommendation 4: Control for publication bias

To date very few meta-analyses in marketing have used tools to compensate for publication bias, even though they have proven effective in other fields. The simplest and most common method is based on careful examination of a graph called a "funnel plot". This type of graph represents the *estimations of effect sizes* extracted from each study according to sample size. The term "funnel plot" comes from the fact that the estimate of the "real" effect size increases in precision with the size *n* of the sample or with the inverse of the standard deviation of the estimation from each study *i* $(1/\sigma_i)$. Then, when the value of $1/\sigma_i$ from each study is represented on the vertical axis (Y), the results from small observation samples will vary with greater amplitude around the real effect size. From a graphic standpoint, these results will be spread broadly across the base of the graph while the results from large samples will be gathered more closely near the top. In the absence of publication bias, the different results obtained in the empirical studies will be spread evenly around the real effect size.

Figure 2 shows a funnel plot obtained with data from the meta-analysis in Szymanski, Troy and Bharadwaj (1995). We can easily see that the estimations from studies of small samples vary with more or less amplitude around the "real" effect size than those with a greater number of observations. When there is no bias the cloud of dots is symmetrical and forms a "funnel" on the graph (Laroche, 2007). Obviously, the different results obtained in the studies identified by Szymanski, Troy and Bharadwaj are not distributed evenly around the effect size (*cf.* Figure 2), indicating that there is a publication bias in favor of studies demonstrating a positive effect of order of entry on market share.

[INSERT FIGURE 2]

[2] In this regard there is a great deal of debate between meta-analysts as to whether it is preferable to keep only the best estimation per study, calculate the mean of all available estimations or propose a weighting system according to the number of estimations in each one.

The procedure adopted in Trappey's study is the one used in Hedges and Olkin (1985) since most of the studies conducted on the subject rely on comparisons of means between experimental and control groups to test the influence of subliminal advertising. The effect size is expressed with Hedges' *g*. This was calculated for each study with the following formula:

$$g = \frac{\overline{Y}^E - \overline{Y}^C}{s}$$

where $\overline{Y^E}$ is the mean of the experimental group, $\overline{Y^C}$ is the mean of the control group and *s* is the aggregated intra-group variance. In Trappey (1996) effect size is corrected for bias due to small sample size. Then, the corrected effect sizes are integrated using the following formula:

$$d = \frac{\sum_{i=1}^{k} w_i d_i}{\sum_{i=1}^{k} w_i}$$

where the d_i are the effect sizes corrected for bias and $w_i = \frac{1}{v_i}$ and weights each effect size according to the inverse of its variance. Cooper (2010) indicates that Hedges' weighting system must be used when calculating the effect sizes of small samples. The results obtained are presented in Table 5.

[INSERT TABLE 5]

The combined estimator of 23 effect sizes is 0.0045 with a 95% confidence interval of -0.061 to +0.070. The confidence interval is determined by the weighted mean of the effect size, more or less 1.96 standard deviation. Thus, the results do not show any positive or negative effect. In order to determine whether studies share a common metric for population, Cochran's Q test was applied. The results of this test indicate that the sample of effects is heterogeneous (with $Q_t = 63.5$ and $p > 0.001$). After excluding two papers that contributed to the heterogeneity of the studies, the author recalculated the mean effect and observed the homogeneity of the results. The latter highlight the weak effect of subliminal advertising on consumer choices (+0.118). In order to assess the relative importance of the effect, the author converted the effect size *d* into a correlation coefficient using Cooper's formula (1984):

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

The correlation coefficient obtained ($r = 0.058$) was compared to the values of Rosenthal and Rubin's (1982) binomial effect size display (BESD), which can be used to evaluate effect size.

In this case the correlation is quite weak since it is situated under r = 0.10 [3]. Various correlations have been tested to see if some variables influenced effect size. Finally, only the type of subliminal treatment (exposure to a film, text or image) seems to have some influence on the experimental effect. The meta-analysis does indeed show that subliminal advertising has little effect on consumer choices.

At this stage of the process, the meta-analyst can suggest testing the robustness of the results obtained. Calculating the "file-drawer number" is a procedure suggested by Rosenthal (1979) that can be used to assess the validity of meta-analysis results and, in marketing, it is certainly the most common method. As suggested earlier, researchers may not publish their studies if the results are not significant and therefore "file them in a drawer" (hence Rosenthal's expression). If this is indeed the case, published studies are not representative of all research conducted on a subject and conclusions based on a summary of their results can be false. Rosenthal estimates there are a large number of non-significant (unpublished) results and introducing them in the meta-analysis can change the conclusions. Consequently, Rosenthal asks how many additional studies (Fail-Safe N) must be included in the meta-analysis to reverse the conclusion that a significant relationship exists at a specific threshold level, which in general is 1 or 5%[4]. Since we could not find this information in Trappey's article (1996) we calculated the "file-drawer number" ourselves by following Orwin's procedure (1983) applied to Cohen's *d* (Wolf, 1986, p. 39)[5]. This calculation indicates that only four studies with a *d* close to 0.10 are required to challenge the results of Trappey's meta-analysis. It is easy to see why he did not calculate this robustness indicator himself!

Like Trappey (1996), Szymanski, Troy and Bharadwaj (1995) chose not to present "file-draw numbers". Therefore we calculated this indicator using information available in the article according to a procedure adapted to correlational type data (*cf.* Wolf, 1986, p. 38). Calculations were conducted exclusively with the published studies in the sample, i.e. 13 studies. In order to obtain Fail-Safe N, we used the following formula at the 5% level:

$$N_{fs0,5} = \left(\frac{\Sigma Z}{1{,}645}\right)^2 - N.$$

After applying this formula and calculating the sum of Z scores for each value of t in the studies, we obtain a Fail-Safe N of 8,304 studies. It would require 8,304 additional studies, each indicating an absence of effect (i.e. $Z = 0$) or sum of no effect (i.e. $\Sigma Z = 0$) to reverse the

[3] For more information on the use of Binomial Effect Size Display (BESD), consult Wolf (1986, p.32-33).
[4] For a more detailed description of the interest and limitations of this technique *cf.* Author (2007).
[5] For this calculation one can also use the Excel "macro" available on the following website: www.stat-help.com/spreadsheets/**Fail**%20**Safe**%20N.xls

conclusion according to which there is a positive link between order of market entry and sales performance of a product or service (with a 5% level of significance). In these conditions, we can suggest that the results of the meta-analysis in Szymanski, Troy and Bharadwaj (1995) are rather robust since it is highly unlikely there are so many studies "filed away in a drawer somewhere" on the subject.

Searching for sources of heterogeneity

Unlike many meta-analyses in marketing, Szymanski, Troy and Bharadwaj (1995) performed a variance analysis (ANCOVA) in order to identify sources of heterogeneity among results of existing studies. The principle of this analysis is to test the influence of certain characteristics in the studies themselves on the results obtained by each one. Szymanski, Troy and Bharadwaj (1995) identified several potential sources of variability between studies such as explanatory variables omitted in the regression models (for example product line breadth, advertising expenditures, price) but also characteristics of the sample (industrial products or consumer goods, brand vs. strategic business unit level) or the type of scales used to evaluate the influence of order of market entry on market share. Therefore the authors created a double-entry table with columns for each moderator and lines for each estimation drawn from the selected studies (*cf.* Table 6). For example Bernt *et al.* (1994) proposes three estimations for which we coded certain characteristics. Thus, the analyses of Bernt and his co-authors only incorporate price as a control variable, not assortment size and advertising expenditures. Similarly, the study focuses on brands (vs. strategic business units) of FMCG (fast moving consumer goods) (vs. industrial products). Furthermore, it uses relative market share as a measure of sales performance and a quantitative measure for order of entry (vs. a dichotomous variable).

[INSERT TABLE 6]

Once the table is completed the meta-analyst has a database he can use to perform multivariate analyses (in this case an ANCOVA) with effect size as the variable to be explained and the selected characteristics of the studies as explanatory variables.

The results of the ANCOVA are presented in Table 7. Here the variable to explain is the non-standardized regression coefficient calculated for each study. The sample size is 64 studies for the model with non-weighted values and 61 studies for the model with values weighted according to sample sizes. This approach is more relevant than conducting a variance analysis (ANOVA) by sub-groups of studies because it takes into account all the moderating variables simultaneously.

[INSERT TABLE 7]

The table shows that the characteristics of the existing studies explain 36 to 82% of the variance of results across studies. The most robust model shows, for example, that studies introducing the explanatory variable "marketing expenditures" in the regression models also point more often to a positive relationship between order of entry and market share. The characteristics of the models explain, in part, these differences in results. However, all other things being equal, the relationship between order of entry and market share remains positive and statistically significant, as observed in the coefficient of the ANCOVA constant.

In the field of marketing, analysis of sub-groups is widely preferred to meta-regressions even though many studies are based on correlational data. However, meta-regression, a term often used by economists (ANCOVA being a specific form of regression analysis), can be used not only to identify factors likely to vary research results, but also to identify and correct publication bias (Stanley and Jarrell, 1989; Laroche, 2007). To our knowledge, this means of searching for moderating variables is rarely used in marketing meta-analyses.

Recommendation 5: Testing for publication bias while analyzing sources of heterogeneity

Marketing studies often cite the complexity of relationships that may exist between variables of interest. It is therefore wise to incorporate tests for publication bias in multiple regression models that take into account other factors explaining variance of results across the studies collected. It is advisable to estimate a multiple regression model with the following form:

$$t = \beta_0 + \beta_1 (1/\sigma) + \delta_1 K_1 + \gamma_1 X_1 + ... + \gamma_k X_k + u \qquad (1)$$

where β_0 is the constant, $X_1 ... X_k$ are variables indicating the forms of certain characteristics of the study i; K is a quantitative variable and u is the random disturbance.

Egger *et al.* (1997) demonstrate that in the statistical test associated with the constant of the equation (1), β_0, is a test of publication bias and its estimation, b_0, indicates the direction of this bias. Testing β_0 is the same as testing the asymmetry of the funnel plot (Sutton *et al.*, 2000).

In practice, statistical software is often used to test more or less advanced (meta)regression models, ranging from simple multiple linear regressions - incorporating many supposed moderating variables - to structural equation models designed to assess causality models. On this point, structural equation models are increasingly used by meta-analysts in marketing. This involves a two-stage procedure. The first stage consists in calculating the correlation coefficients between different variables of interest, based on several independent meta-analyses. The second stage consists in testing the possible structural equations model, using the correlations matrix thus obtained, with software programs such as AMOS or LISREL (Viswesvaran and Ones, 1995; Cheung and Chan, 2005). This procedure can also be used to envisage direct and/or indirect relationships between variables of interest that have not necessarily been examined in the original studies.

In the end, three types of software programs can be used to conduct a meta-analysis. The first option consists in using an Excel-type spreadsheet. This is an excellent means to acquire the fundamentals of meta-analysis to the extent the researcher must develop his own "macros" (and therefore review his calculation formulas) in order to obtain the basic statistics (weighted means, confidence intervals, Cochran's test, etc.) required for aggregating the results and the heterogeneity tests. On a personal note, we found nothing better than an Excel spreadsheet for gathering the characteristics of each study before processing them with specialized statistical software. The spreadsheet has its limitations when it comes to more precise analyses. However, for the last few years there have been software programs like MIX 2.0 that allow the use of more elaborate meta-analytical tools under Excel. The second option consists, however, in using statistical software such as SPSS, SAS, R or Stata, which allow the most sophisticated meta-analyses (covariance analyses, meta-regressions, structural equations models...). Finally, it is always possible to use specialized meta-analysis software programs. There are more and more of them available, particularly in the field of medical research. This software can be of interest for a meta-analysis concerning the type of experimental data that is common in epidemiology, psychology and marketing. However, it is not suitable for meta-analysis of correlational data where it is preferable to use a good statistical program (SAS, Stata, LISREL...). Appendix A2 offers a comparison of the main software programs currently available to researchers.

CONCLUSION

The construction of a meta-analysis is based on a strict methodology that is justified in complex situations where the literature leads to contradictory conclusions. The undeniable benefit of a meta-analytical approach is to establish rules for integrating data that can be shared by all researchers in the often subjective exercise of reviewing the literature.

However, meta-analytical procedures are not a panacea for solving problems inherent in this process (Sackett *et al.*, 1985; Schmidt, 1992; Bobko and Stone-Romero, 1998; Hermann and Joseph, 1999; Aguinis *et al.*, 2011). Meta-analysis is a relatively recent methodological tool that is the object of recurring critiques and these will continue as long as users fail to comply with the demands its application requires. Indeed, each step in a meta-analysis is a stage in a structure whose bases must be clearly stated. The apparent precision associated with calculating a common metric can generate false certainties if the means used to produce it are inaccessible. Meta-analysis presents several limitations that have been highlighted by its detractors. First, the choice of studies included in the meta-analysis - and the risk of "mixing apples and oranges" - is often the subject of harsh criticism to the extent this selection process is sometimes based on subjective criteria, generally stemming from a narrative review of the literature! It is important to bear in mind that meta-analyses do no replace or exclude narrative reviews, but remain a complementary means of rendering results in the literature easier to understand. Another oft-cited problem is the quality of the studies selected. We have observed that it is possible, however, to work around this problem by proposing a proportional weighting of publication quality. Finally, one of the problems in conducting meta-analyses is the search for moderating variables. Identifying moderating variables depends on information available in the existing studies and the choice of these moderators is often based on preconceived notions. This approach is therefore far from satisfactory from a scientific standpoint.

Despite these limitations, it is still preferable that meta-analytical procedures, which have become mainstream in science, enrich research methods in management science and contribute, in particular, to synthesize empirical investigations in marketing. While the number of French researchers who have published meta-analyses in marketing remains limited for the time being, it is important to encourage the development of meta-analyses outside the United States (Laurent, 1999).

APPENDIX A1: Summary of meta-analyses published in marketing by type of journal and by theme

Themes \ Rank A journals*	*JCR*[a]	*JM*	*JMR*	*MKS*	Total
Methodology	3		3		6
Consumer	4			1	5
Marketing strategy		5	1		6
Price/Promotion			3	1	4
Product/Service		1	4	1	6
Communication	2	2	5		9
Distribution and sales force			4		4
TOTAL	9	8	20	3	40

[a] *JCR*: *Journal of Consumer Research*; *JM*: *Journal of Marketing*; *JMR: Journal of Marketing Research*; *MKS*: *Marketing Science*

Themes \ Rank A journals	*IJRM*[b]	*JCP*	*JIBS*	*JR*	*JAMS*	*ML*	*RAM*[1]	Total
Methodology			1			1		2
Consumer	2	2	1		2	1		8
Marketing strategy	3				2			5
Price/Promotion	2			2	1	2		7
Product/Service					1			1
Communication	1				5	1		7
Distribution and sales force				3	1			4
TOTAL	8	2	2	5	12	5	0	34

[b] *IJRM : International Journal of Research in Marketing ; JCP : Journal of Consumer Psychology ; JIBS: Journal of International Business Studies ; JR : Journal of Retailing ; JAMS : Journal of the Academy of Marketing Science ; ML : Marketing Letters ; RAM : Recherche et Applications en Marketing.*

[1] *RAM* published 5 meta-analyses, but in its international section (these were articles previously published in A* or A ranking journals).

Themes \ Rank B journals	*IJMR*[c]	*JA*	*JAR*	*JBR*	*JPSSM*	*P&M*	Total
Methodology	1		1				2
Consumer				2		2	4
Marketing strategy				2			2
Price/Promotion				1			1
Product/Service							0
Communication		2	4			1	7
Distribution and sales force				2	2	1	5
TOTAL	1	2	5	7	2	4	21

[c] *IJMR : International Journal of Market Research ; JA : Journal of Advertising ; JAR : Journal of Advertising Research ; JBR : Journal of Business Research ; JPSSM : Journal of Personal Selling and Sales Management ; P&M : Psychology & Marketing.*

Themes \ Rank C journals	*EJM*[d]	*IMR*	*JMM*	*JMTP*	*JPPM*	*JSR*	Total
Methodology		1	1				2
Consumer						2	2
Marketing strategy	1	1					2
Price/Promotion					1		1
Product/Service				1			1
Communication					2		2
Distribution and sales force							0
TOTAL	1	2	1	1	3	2	10

[d] *EJM : European Journal of Marketing ; IMR : International Marketing Review ; JMM : Journal of Marketing Management ; JMTP : Journal of Marketing Theory and Practice ; JPPM : Journal of Public Policy and Marketing ; JSR : Journal of Service Research.*

Appendix A2: Presentation of some software programs available for meta-analysts

Software name	Publisher	Website	Price	Stat. descript.	ANOVA	Regression
Spreadsheet						
Excel	Microsoft	www.microsoft.com	Office Suite	++	-	-
General statistics software						
SPSS	IBM	www.spss.com David Wilson proposes a macro for SPSS that can be used to conduct meta-analyses at the following address: http://mason.gmu.edu/~dwilsonb/ma.html	From €1,900 to €4,200	++	++	++
Stata	Stata Corps.	www.stata.com Many statisticians propose software programs for conducting meta-analysis with Stata (including D. Wilson) A book dedicated specifically to meta-analysis under Stata is available on the website http://www.stata-press.com/books/mais.html (Sterne, 2009)	Around €700 (Stata/SE, education version)	++	++	++
SAS	SAS Institute	David Wilson also proposes macros for SAS at the following address: http://mason.gmu.edu/~dwilsonb/ma.html Furthermore, other meta-analysts propose specific programs for conducting meta-analyses under SAS (*cf.* Wang et Bushman, 1999)	From €2,100 to €4,000 (education version)	++	++	++
R	R Foundation	Thomas Lumley (meta) and Guido Schwarzer (meta) propose macros for R software. "Metafor Package" can also be used to conduct meta-analyses under R (www.metafor-project.org)	Free	++	+	++
Specialized meta-analysis software						
MIX 2.0	BiostatXL	MIX 2.0 is available free of charge for conducting meta-analyses under Excel (www.meta-analysis-made-easy.com/) but is more suitable for medical research.	Free	++	+	-
Meta-Analysis 5.3	Ralph Schwartzer	http://userpage.fu-berlin.de/~health/meta_e.htm	Free but under DOS	++	+	-
Comprehensive Meta-Analysis (CMA)	Borenstein *et al.*	www.meta-analysis.com	From €300 to €900	++	+	+
Metastat	Rudner *et al*	http://echo.edres.org:8080/meta/metastat.htm Software more suitable for medical research	Free	++	+	-
RevMan	Cochrane collaborative project	http://ims.cochrane.org/revman Software more suitable for medical research	Free	++	+	-
MetaWin 2.0	Rosenberg, Adams and Gurevitch	www.metawinsoft.com	From €60 to €100	++	+	-
META	David A Kenny	http://davidkenny.net/meta.htm	Free	+	-	-

Table 1: The different stages of meta-analysis

Stage	Research question	Main objective
1. Formulate a research question	What is the purpose of this study?	Define the variables of interest and specify the relationship studied in order to identify the studies concerned.
2. Collect existing empirical studies	Which procedures should be used to find the relevant studies?	Identify the sources (databases, types of journals...) and keywords to search for the studies concerned.
3. Evaluate the consistency of the methods used in each of the empirical studies identified	Which studies should be included or excluded from the analysis according to their characteristics?	Apply these criteria in order to select the studies.
4. Code the data collected	Which information should be gathered?	Select the relevant data and establish a list of characteristics of the studies that need to be collected.
5. Analyze and integrate the results of individual empirical studies	Which procedures should be used to summarize and integrate the empirical results?	Identify and apply the procedures to integrate the results and test the differences between results across studies.
6. Interpret the results of the synthesis	What conclusions can we draw from the results of the meta-analysis?	Summarize the results of the meta-analysis.
7. Present the research methods and results	Which data should be presented in the summary report?	Identify and apply editorial rules in order to highlight the most significant results.

Table 2: Examples of moderating effects in Peterson and Jolibert (1995)

Characteristics of the study	Perceived quality				Purchase intention			
	mean	Standard deviation	N	Sig.	mean	Standard deviation	N	Sig.
Type of respondent - students - consumers - professionals	 0.28 0.30 0.32	 0.24 0.25 0.31	 139 704 122	0.44	 0.05 0.28 0.28	 0.05 0.32 0.22	 218 129 209	0.00
Sample size - less than 260 - more than 260	 0.28 0.32	 0.24 0.26	 577 387	0.00	 0.16 0.27	 0.21 0.26	 411 145	0.00
Number of countries studied - less than 10 - more than 10	 0.30 0.30	 0.25 0.23	 566 398	0.96	 0.14 0.28	 0.23 0.22	 347 209	0.00

Source: adapted from Peterson and Jolibert (1995).

Table 3: List of studies selected by Szymanski, Troy and Bharadwaj (1995) for their meta-analysis

Authors/year	Mean sample size	Number of effects reported	Number of statistically significant effects reported	Mean effects reported	Scope of effects reported
Bernt *et al.* (1994)	163	3	1	0.22	0.05 to 0.49
Brown and Lattin (1994)	85	4	2	0.24	0.09 to 0.38
Huff and Robinson (1994)	95	5	5	0.54	0.45 to 0.66
Kalyanaram (1993)	346	12	12	0.96	0.36 to 1.33
Kekre and Srinivasan (1990)	Nr	2	2	8.41	8.21 to 8.61
Lambkin (1988)	144	4	2	10.08	2.64 to 19.61
Lambkin (1992)	2746	2	2	12.27	10.27 to 14.27
Mascarenhas (1992)	112	2	2	0.26	0.17 to 0.34
Mitchell (1991)	76	8	7	4.69	-0.12 to 9.87
Reddy, Holak and Bhat	660	2	2	0.06	0.06 to 0.07
(1994)	1209	1	1	-14.5	Na
Robinson (1988)	119	6	3	5.49	3.81 to 6.97
Robinson (1990)	371	1	0	-2.11	Na
Robinson and Fornell (1985)	45	2	1	9.2	7.9 to 10.5
Sullivan (1992)	106	2	2	0.49	0.49 to 0.49
Urban *et al.* (1986)	255	8	8	7.42	1.89 to 14.9
Vanhonacker and Day (1987)					

Source: Szymanski, Troy and Bharadwaj (1995, p. 19)

Table 4: List of studies selected by Trappey (1996) for his meta-analysis

Study	Number of estimations per study	Measure of the effect (*d*)	Sample size	Significativity
Byrne (1959)	1	0.2482	105	-
Champion and Turner (1959)	1	0.1978	38	-
Hawkins (1970)	1	0.0977	20	-
George and Luther (1975)	1	0.4109	37	-
Cuperfain and Clark (1985)	2	0.5737	30	-
		0.5660	32	-
Kilbourne *et al.* (1985)	2	0.0326	99	-
		-0.2942	355	+
Gable *et al.* (1987)	4	0.0898	425	-
		-0.3094	425	-
		-0.4093	425	-
		0.2196	425	+
Caccavale *et al.* (1981)	10	0.0000	106	-
		0.2630	106	-
		0.0736	106	-
		0.0481	106	-
		0.0761	106	-
		0.2877	106	-
		0.4067	106	-
		0.2013	106	-
		0.0879	106	-
		0.3902	106	+
Weinstein *et al.* (1986)	1	0.1082	89	-

Note: a plus sign in the last column indicates a statistically significant result.

Table 5: Results of Trappey's meta-analysis (1996)

	Complete sample	Reduced sample
Number of studies	**23**	21
Sample size	**3565**	2715
Hedges' d_t	**+0.0045**	+ 0.1173
95% confidence interval	**-0.061 + 0.070**	+ 0.042 to + 0.193
Cochran's Q_t test	**63.5****	Not significant

** significant at a 1% threshold (Chi-square test)

Table 6: Excerpt from the data table Szymanski, Troy and Bharadwaj (1995) used to search for sources of heterogeneity across results

	N	t	ES	Assortment size	Advertising expenditures	Price	FMCG	SBU	Relative MS	Order of entry
Bernt *et al.* (1994)	118	0.32	0.054	0	0	1	1	0	1	1
	118	0.68	0.116	0	0	1	1	0	1	1
	255	49.20	0.492	0	0	1	1	0	1	1
Brown and Lattin (1994)	129	5.88	0.410	0	1	0	1	0	0	1
	129	1.26	0.121	0	1	0	1	0	0	1
	129	8.10	0.380	0	1	0	1	0	0	1
	40	0.36	0.039	0	1	0	1	0	0	1
Huff and Robinson (1994)	95	4.01	0.660	1	1	0	1	0	1	1
	95	5.13	0.500	1	1	0	1	0	1	1
	95	5.45	0.480	1	1	0	1	0	1	1
	95	6.54	0.590	1	1	0	1	0	1	1
	95	2.40	0.420	1	1	0	1	0	1	1
	95	1.49	0.290	1	1	0	1	0	1	1
Lambkin (1988)	187	1.96	0.070	1	1	1	0	1	0	0
	187	2.31	0.519	1	1	1	0	1	0	0
	129	1.62	0.095	1	1	1	0	1	0	0
	129	2.31	0.355	1	1	1	0	1	0	0
(…)										

Table 7: Analysis of sources of heterogeneity (ANCOVA) conducted by Szymanski, Troy and Bharadwaj (1995)

	Supposed effects	Unweighted Model		Sample Size Weighted Model (SSW)	
		n	Regression coefficients	n	Regression coefficients
Constant (effect size)		64	3.18 (2.19)	61	4.21** (1.79)
Explanatory variables omitted					
Product Line Breadth	+	39	7.77** (2.47)	36	4.56** (1.53)
Marketing expenditures	+/-	48	-4.21** (2.33)	46	3.95** (1.74)
Price	+/-	33	3.05** (1.41)	31	0.09 (1.12)
Characteristics of the sample					
Consumer Business	+	34	4.19** (2.15)	34	-0.48 (1.44)
Strategic level (SBU)	+/-/0	31	8.07** (2.27)	29	7.10** (1.86)
Characteristics of scales					
Relative MS	+/-	32	-0.24 (1.19)	30	0.96 (0.98)
Actual order of entry	+/-	34	-4.81** (1.72)	32	-3.94** (1.38)
R^2 (R^2 adjusted)			**0.43 (0.36)**		**0.84 (0.82)**
Fisher F-test (*p-value*)			**6.06 (<0.001)**		**38.81 (<0.001)**
Max VIF			**5.10**		**8.74**

Source: Szymanski, Troy and Bharadwaj (1995, p. 23)

Figure 1: Sample representation of values distributed around the mean effect size (combined) produced with MetaWin®

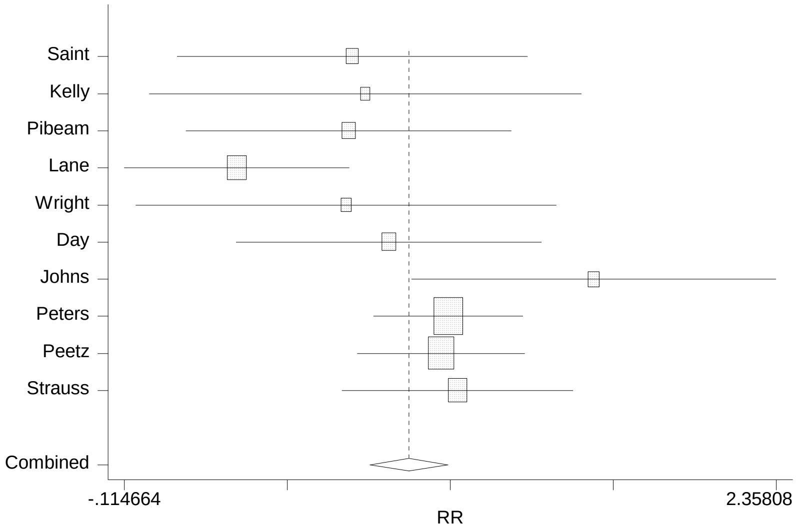
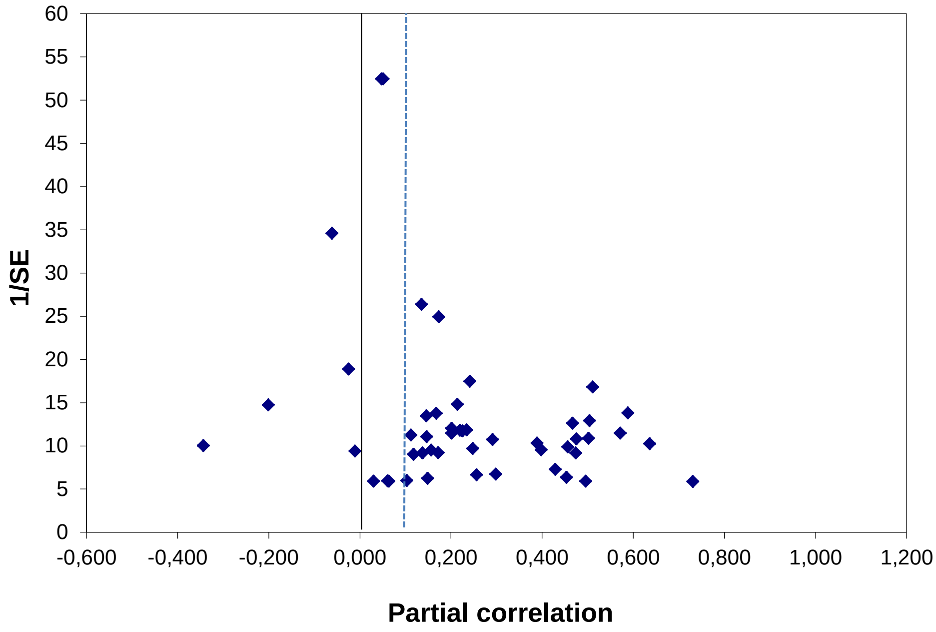


Figure 2: Funnel plot with data from Szymanski, Troy and Bharadwaj (1995)



Box 1: Definitions and fundamental principles of meta-analytical procedures

• Definitions:

The ***effect size*** or ***amplitude of effect*** measures the intensity of the relationship between two variables of interest. The most common methods for calculating effect size are: Pearson's correlation coefficient, Cohen's *d*, which measures the standardized difference between two means, and the odd ratio, which compares the likelihood of an event in two groups. The choice of one or another depends on the field of research, the nature of the data collected and its statistical processing.

The ***statistical power*** of a study measures its capacity to demonstrate the effect of a variable of interest, providing this effect does indeed exist. In other words, the statistical power of a test is its aptitude (in terms of probability) to produce statistically significant results if the measured effect is real. The power is equal to 1-β, where β is a Type 2 risk, the risk of not revealing an effect that does indeed exist. The power of the test increases, in particular, with the size of the sample and the importance of the sought-after effect.

Sampling error or fluctuation occurs when a researcher observes only part of the entire sample. Thus, all estimations based on a sample can be subject to sampling error.

Measurement error corresponds to an error in the scales used to produce the results. This depends on the quality and accuracy of a measurement tool in providing an indication close to the real value.

• Fundamental principles:

The general principle of meta-analysis is based on the hypothesis that ***the amplitude of the link between two variables (or effect size) is a constant and each study dedicated to this relationship measures this constant.*** Therefore differences in results observed across individual studies cannot be attributed to sampling fluctuation. The real value of the effect size remains unknown as existing studies only offer estimations subject to sampling error and fluctuation. Consequently, the purpose of a meta-analysis is to obtain the best possible estimation of the common metric.

This approach is based on three main principles:
- an exhaustive search for existing studies
- rigorous and well-argued selection of studies
- an estimation of the common effect size

Box 2: Principal metrics used in meta-analyses

For results from experimental research, the effect size is often expressed by Hedges' *g*. This is calculated for each study with the following formula:

$$g = \frac{\overline{Y}^E - \overline{Y}^C}{s}$$

where $\overline{Y^E}$ is the mean of the experimental group, $\overline{Y^C}$ is the mean of the control group and *s* is the aggregated intra-group variance. The variance is calculated as follows:

$$s = \sqrt{\frac{(n^E - 1)(s^E)? + (n^C - 1)(s^C)?}{n^E + n^C - 2}}$$

where n^E and s^E represent the size and standard deviation of the experimental group and n^c and s^c are the size and standard deviation of the control group. However, *g* is a biased estimator, especially for small samples. It is therefore better to use the unbiased estimator *d*, which is calculated in the following manner:

$$d = \left(1 - \frac{3}{4N - 9}\right) g$$

Example 1: In a study concerning age and brand awareness, the mean age of consumers in the control group is $Y^C = 41$ with a standard deviation of 4.5 years. In the experimental group (made up of people aware of a particular brand) $Y^E = 38$ and $s^E = 4$. The size of the control group is $n^C = 158$ and the experimental group is $n^E = 155$. An estimation of the common standard deviation is obtained as follows:

$$s = \sqrt{\frac{(155 - 1)4^2 + (158 - 1)4{,}5^2}{158 + 155 - 2}} = 4{,}26$$

The effect size is estimated by using the previous formulas:

$$g = \frac{38 - 41}{4{,}26} = -0{,}70 \qquad d = \left(1 - \frac{3}{4 \times 313 - 9}\right) \times (-0{,}70) = -0{,}69$$

For results from correlational type research, the regression coefficients, Pearson's correlation coefficients or even estimated elasticities in each study can be transformed into *r* partial correlation coefficients (Wolf, 1986, p. 35).

Example 2: The following table illustrates how to calculate the standardized effect from data produced by 4 regressions:

Study	Sample size	Standardized (or not) regression coefficient	*t* value associated with the regression coeff.	*k* number of explanatory variables in the regression model	*r* partial correlation coefficient
A	118	-0.054	-0.32	7	-0.030
B	85	-0.410***	-5.88	4	-0.467
C	95	-0.660***	-4.01	4	-0.389
D	98	0.122	0.10	9	0.011

$$r = \sqrt{\frac{t^2}{(t^2 + dof)}} = \sqrt{\frac{0.32^2}{\left(0.32^2 + (118 - 7 - 1)\right)}} = 0.030$$

where t is the Student's t-test associated with the principal explanatory variable and dof is the number of degrees of freedom associated with the regression equation.

Furthermore, when researchers use tests to compare two groups (for example Student's t-test or Fisher's F-test for continuous variables) it is still possible to calculate *r* based on these different tests Wolf (1986, p. 35) or Rosenthal (1991, p. 19) propose conversion tables for passing from one statistic to another.

Box 3: Combining standardized effect sizes

In the case of effect sizes calculated using a regression coefficient or *r* correlation, the mean of the coefficients is calculated to produce an estimation of the effect size within the population. The procedures of Hunter, Schmidt and Jackson (1982) and of Rosenthal and Rubin (1978) propose weighting the different r according to sample size in order to avoid overestimating the effect of studies focused on small populations. Therefore the estimated effect size of the population is calculated by weighting the effect size according to sample size *N* for each study using the following formula:

$$\bar{r} = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i}$$

where N_i is the number of individuals in the sample and r_i is the effect size of each study.

This formula applies not only to effect sizes measured with *r* but also those measured with, among others, Fisher's Z transformation and the standardized mean difference *d*. The variance of *r* is calculated in the following manner and is used to present the formula for the 95% confidence interval ($r - 1{,}95\sqrt{v_r} \leq r \leq r + 1{,}95\sqrt{v_r}$):

$$v_r = \frac{\left(1 - r^2\right)^2}{n - 1}$$

In the case of a meta-analysis using Cohen's *d* or Hedges' *g* as a common metric, it is advisable to calculate the weighting coefficient w_i first, which corresponds to the inverse of the variance associated with each estimator *d*. It can also be calculated more directly with the following formula:

$$w_i = \frac{2\left(n_{i1} + n_{i2}\right) n_{i1} n_{i2}}{2\left(n_{i1} + n_{i2}\right)^2 + n_{i1} n_{i2} d_i^2}$$

where n_{i1} and n_{i2} = the number of observations in group 1 and group 2 of study *i*.

Once the weighting coefficient has been calculated the weighted mean can be obtained in the following manner:

$$\bar{d} = \frac{\sum_{i=1}^{k} d_i w_i}{\sum_{i=1}^{k} w_i}$$

and the formula for the 95% confidence interval around this mean is:

$$CI_{d.95\%} = \bar{d} \pm \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}$$

Box 4: The procedure in Hunter, Schmidt and Jackson (1982)

The procedure in Hunter, Schmidt and Jackson (1982) does not necessarily use statistical tests to determine homogeneity of effects but seeks, in an initial stage, to evaluate the variance of effects linked to sampling errors. Thus, in order to determine the variance of effect sizes within a population, they calculate the sum of the squared differences between each effect size and the estimated effect size within the population. Each difference is then weighted according to the sample sizes of each study. This corresponds to the variance observed within the population.

$$s_r^2 = \frac{\sum_{i=1}^{k} N_i \left(r_i - \bar{r}\right)^2}{\sum_{i=1}^{k} N_i} \qquad \text{[Observed variance]}$$

Then the share of variance of effects linked to sampling errors is calculated as follows:

$$s_{er}^2 = \frac{k\left(1 - \bar{r}^2\right)^2}{\sum_{i=1}^{k} N_i} \qquad \text{[Variance linked to sampling errors]}$$

where *k* is the number of studies in the sample.
Finally, the observed variance is subtracted from the variance linked to sampling errors and the difference corresponds to the residual variance.

$$S_{pxy}^2 = S_r^2 - S_{er}^2 \qquad \text{[Adjusted or residual variance]}$$

If the residual variance is less than 25% of total variance, effect sizes are considered homogeneous. If the contrary is true, it is necessary to determine the moderating variables (Hunter, Schmidt and Jackson, 1982). However this rule is not sufficient for testing the homogeneity of small study samples. Thus, an additional test designed to confirm a lack of homogeneity across studies, based on consideration of the variance, is provided for in the new procedure by Hunter and Schmidt (1990). This is a non-parametric test that follows a Chi-square law and is interpreted like Cochran's Q test:

$$c_{k-1}^2 = \frac{N}{\left(1 - r^2\right)^2} S_r^2$$

In the end, the emergence of moderating variables, in Hunter and Schmidt's procedure (1990), must be based on a series of indicators:

(1) the 75% residual variance rule
(2) the homogeneity test based on a Chi-square law
(3) the 95% confidence interval, while considering that if the zero value is included on this interval, we accept the hypothesis of a coefficient $r_{xy} = 0$.

Box 5: The procedure in Hedges and Olkin (1985)

Hedges and Olkin (1985) proposes a homogeneity test of effect sizes in response to the following fundamental question: *Is variation in effect sizes really caused by moderating variables or simply sampling errors?* Total homogeneity across studies is tested with Cochran's Q, which can be used to test the null hypothesis according to which all effect sizes are equal (Hedges and Olkin, 1985 ; Gurevitch and Hedges, 1993). The total heterogeneity of a sample, Q_T, is calculated in the following way:

$$Q_T = \sum_{i=1}^{k} w_i \left(d_i - \bar{d}_t\right)^2$$

where $w_i = 1/v_i$ is the inverse of the sample's variance, d_i is the effect size of study i and $\bar{d}_t$ is the estimation of the mean effect size within the population.

The value obtained is distributed like a Chi-square. If Q_T is close to 1, the variation in results is due to sampling. If the contrary is true, the variation is due to sampling fluctuation. This procedure can be used to determine the existence of any groups of studies with homogeneous effect sizes.