Meta-analysis techniques are widely used to combine results from different studies to estimate the true effect size in the population and to explain the variability of estimates found in primary studies. These methods have changed over the years and continue to evolve, taking progressively into account the critics raised. This paper reviews applications of these methods in marketing research published in several journals to assess the relevance of such use. We discuss important issues related to meta-analysis methods and how marketing researchers handled them: issues related to studies’ selection and screening prior to model estimation, correction for artefacts, publication bias, estimation methods used to summarize effect sizes, and those related to the statistical power of meta-analysis. In light of the results found, recommendations are made for future meta-analysis practice in marketing.

Key words: marketing, meta-analysis, publication bias, statistical power.
Introduction. Empirical generalizations constitute the basis for knowledge accumulation in science. In 1995, the review Marketing Science devoted a special issue to this theme in marketing. Bass and Wind (1995) defines empirical generalisation as „a pattern or regularity that repeats over different circumstances and that can be described simply by mathematical, graphic or symbolic methods“ (page G1). According to this author, there are four main approaches to synthesising results, with a view to generalize, from empirical studies on a common scientific issue: meta-analysis, literature review, content analysis, and clustering and informal methods. Nowadays, meta-analysis is the most widely used (and recommended) approach in searching for generalizable results (Cooper and Hedges, 1994). It constitutes a rigorous alternative to narrative literature reviews (Farley et al., 1995). Indeed, compared to narrative literature reviews, meta-analysis permits determining not only the sign of the relation between two variables but also its magnitude. Moreover, it allows testing statistically various hypotheses. Besides, it has been shown that literature reviews does not account for sample size (the number studies included in meta-analysis and the number of observation used in each study), have low statistical power, and give rise to biased results (Hedges and Olkin, 1985). As to the remaining methods, namely informal and content analyses, they are seldom used in practice. This paper therefore focuses on empirical generalizations in marketing by the way of meta-analysis.

The purpose is to examine applications of meta-analysis for the accumulation of knowledge relevant to marketing research. The following two questions are addressed: What are the practices in this domain? Do those practices conform to the usual recommendations for performing a meta-analysis? The answers to these questions are of primary importance to assess the validity of the results obtained in the applications to marketing published so far. As underlined by Wolf (1990), „there is need for more reviews (meta-analytic or otherwise) of the meta-analyses that have already been conducted in order to better understand the implications of the varying philosophical, methodological, and statistical practices that have been used“ (p. 151). In marketing, none of the previous surveys has examined how meta-analysis procedures have been used and how threats to the validity of a meta-analysis have been dealt with by researchers.

It is to be noted that our work differs from the review carried out by Franke (2001), which focused on the substantive results of meta-analysis applications in marketing. Our research takes place upstream, concentrating on issues of methodology rather than substantive results. We propose a critical review of procedures used in these applications. The principal objectives of the present research are:

i) a descriptive analysis of the characteristics of the published studies such as the number of effects estimated per study, the size of the sample used, the length of the period for which the effect size is estimated, and so on. These results are useful in carrying out a simulation (which has to reflect the real practices) or in calculating the power of the meta-analysis;

ii) an assessment of the quality and appropriateness of the meta-analytical techniques used in practice. In other words, it is relevant to verify whether the statistical assumptions of the methods used are met.

In light of our findings, we identify problem areas and suggest avenues for improvement and discuss future perspectives of meta-analysis use.
Methodology. Over the last fifteen years, important methodological improvements have been achieved in conducting meta-analyses. They relate to criteria for inclusion of studies, to the choice of the method of estimation, to tests of publication bias, and to problems linked to correlated estimates. These developments allow improving the robustness of meta-analysis results and palliating the deficiencies recorded in the literature: the „apples and oranges“ and „garbage in, garbage out“ problems, where it is argued that studies are too dissimilar to make comparisons relevant (Eysenck, 1978); „the file drawer“ problem which arises from failure to obtain a representative sample of the population of studies on some domain of interest (Rosenthal, 1979); and other secondary critics (see Sharpe, 1997; Matt and Cook, 1994). Procedures for performing a meta-analysis. The statistical literature on meta-analysis discuss numerous alternative procedures and methods for assessing the validity of a generalization (Bangert-Drowns, 1986). In consequence, it is difficult to propose a best way to carry out a meta-analysis, all the more because some recommendations are recent and have not been subjected to an evaluation. Nevertheless, it is possible to present the main lines common to the different approaches that will be used as guidelines in assessing applications of meta-analysis in marketing.

A widely accepted definition of meta-analysis is as „the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating findings“ (Glass, 1976, p. 3). Over the last 25 years, meta-analysis has become a popular and methodologically sophisticated technique for quantitatively summarizing findings from a large body of replication studies.

A number of text-books on meta-analysis are now available (Hedges and Olkin, 1985; Farley and Lehman, 1986; Wolf, 1986; Hunter and Schmidt, 1990; Cooper and Hedges, 1994). Meta-analysis methods are widely used in medicine and psychology, and more and more researchers in marketing have recourse to these techniques to summarize findings from various studies and to explain the variability of estimates.

A major contribution of meta-analysis is the increased statistical power that can be obtained by combining information of primary empirical researches. This increased power is crucial to detect a small effect size and to resolve the problem of conflicting findings of independent studies. A second contribution is the potential precision of the overall effect estimate. Indeed, a large sample of meta-analysis (compared to that of a single study) yields a decrease in the standard error of the estimated effect size (Hunter and Schmidt, 1990; Cooper and Hedges, 1994). A third contribution is that it allows achieving a greater generalizability of findings than any single study does because it accounts for numerous moderator variables that could explain the variability of the effect sizes found in various imperfect replication studies (Farley and Lehman, 1986; Hunter, 2001). A fourth contribution of meta-analysis is to examine complex theoretical models through research synthesis or to create an empirical model with the existing data. Indeed, none of the primary studies can by itself examine the whole relations of a complex theoretical model (Farley et al., 1994; Geyskens et al., 1999). Often, each of them is confined to the examination of only some of these relations. The interest in meta-analysis lies in its capacity to organise the whole set of conclusions from these studies to get the most accurate fit of the theoretical model. Meta-analysis can thus contribute to theory testing by providing data to be analysed with other methods such as bayesian regression (Vanhonacker and Price, 1992) and structural equation models (Viswesvaran and Ones, 1995; Cox et al., 1994). Thus, beside innovative or creative articles, meta-analysis can contribute to theory development.

In marketing, meta-analysis is increasingly used to summarize findings and to set the agenda for future research (Farley et al., 199, 1998; Franke, 2001). It has been applied to a variety of research topics in marketing. Obviously, the accuracy of these applications

2 Studies measuring different things, manipulating different concepts ("apples and oranges" threat), and/or being of different quality ("garbage" threat).
depends largely on the correct use of the recommended procedures, in conformity with the underlying hypotheses.

**Issues related to selective inclusion of studies and coding.** The criteria for selecting studies to be included in the meta-analysis need to be considered with particular attention so as to prevent the „apples and oranges“ threat and selection bias. The criteria of selection, such as the definition of research topic, have to be specified a priori in the protocol of the meta-analysis in order to well delimit the field of potential studies to include. The results of two meta-analyses on the same research topic can diverge if the inclusion criteria are different. For example, the inclusion or exclusion of certain papers may modify substantially the conclusions of the meta-analysis (Grégoire et al., 1995).

**Literature search:** Once the research area is defined, relevant studies are identified by searching journals, books, conference proceedings, theses, as well as numerous electronic bibliographic databases (EBSCO, ABI inform, Psychlist, ERIC, Econlit), professional databases etc. Another important source consists of consulting colleagues and/or experts in the domain under study. Of course, when selecting studies one must ensure that each study is somewhat a replication of the topic of interest (Hunter, 2001).

However, in practice, some constraints such as the language of publication and the fact that relevant articles may be unpublished, limit the scope of study identification (selection bias). As a consequence of this non-exhaustiveness, the studies included in the meta-analysis may be non-representative and can thus lead to biased conclusions. Grégoire et al. (1995) showed that including non-English studies in their meta-analysis affects findings.4

**Quality criteria:** The choice of quality criteria is problematic (Cook et al., 1992). Wortman (1994) argued that quality is determined by relevance (construct and external validity) and acceptability (internal and statistical validity). However, there is no consensus on this subject (Sharpe, 1997). Although the procedure for mixing up results from studies of different quality was criticized (Wachter, 1988), most meta-analyses in the social sciences include all of the studies treating the topic of interest because of the lack of a rule (see Cooper and Hedges, 1994).

**Coding schema:** The characteristics of the studies are coded for sample size, effect size and potential moderator variables defined a priori. Nevertheless, the latter does not easily lend themselves to codification because several definitions and measures may coexist for the same variable. In addition, imprecision and lack of information in some studies make their coding difficult. At last, the limited number of observations for some moderators leads to aggregate them to increase the number of observation. In consequence, the comparability of findings in several imperfect replications is not always straightforward (Orwin, 1994). Two independent authors may code differently the same study. Hence, there is advantage in having the primary studies coded by several authors and searching for a consensus about contradictory coding so as to obtain a high inter-rater agreement coefficient.

**Issues related to measurement of effect size. Common effect size metric:** To be compared, primary studies findings must have a common measure. This common effect size refers to the magnitude of effect observed across studies. There are numerous effect size indicators or metrics for measuring the strength of relationship or the magnitude of

---

3 Choice and definition of the dependent and independent variables of interest. As for any eligibility criterion, there exists a risk of excluding relevant empirical studies simply because they do not meet the definition adopted.

4 These authors shows that in at least one of the 36 meta-analyses they reviewed, the exclusion of papers for linguistic reasons produced results different from those which would have been obtained if the exclusion criteria had not been used.

5 To deal with the „garbage“ threat, some authors advocated exclusion of poor studies (see Kraemer And Yesavage, 1998).
difference between variables: differences between means (Cohen’s d and Hedge’s g), measures of association (Pearson’s r and explained variance indicators: \( \omega^2, \eta^2, \epsilon^2 \)), regression slopes (\( \beta \) coefficients) and combined probability (p-value). The formulas for calculation and conversion as well as the properties of all of the indicators are discussed by Fern and Monroe (1996). Cohen (1988) provides some guidelines for interpreting effect sizes.

**Correction for artefacts:** The variability of effect size across studies can partly be due to methodological and statistical artefacts such as sampling error, measurement error, and range restriction. These artefacts can induce a large error\(^6\) in the estimation of the effect size so as to create spurious findings (Hunter and Schmidt; 1990). According to the work of Hunter and Schmidt (1996), the proportion of variance due to artefacts in I/O psychology amounts on average to 80 % of variance of effect size. This statistical and methodological variability across effect size estimates can lead researchers to incorrectly conclude that it is due to moderating variables. Therefore, the bias introduced by these artefacts should be controlled and corrected for in order to obtain accurate effect size.

However, some researchers argued that the implementation of corrections for artefacts is not always desirable when methodological artefacts and substantive situational moderators are correlated (James et al., 1992). In this case, the correction for artefacts may be eliminate a substantive moderator.

**Exploratory analysis:** Data screening is a prerequisite step for model estimation. An exploratory statistical analysis effect sizes adjusted for artefacts (stem-and-leaf plot, box-plot, funnel plot, normality tests) has to be performed for, on the one hand, detecting outliers and assessing normality,\(^7\) and, on the other hand, discovering patterns to make hypothesis development and refinement (Light et al., 1994; Behrens, 1997).

**Estimation and testing procedure issues. Weighted effect sizes:** Since the primary studies use different sample sizes, the studies with large samples should be given more weight, as they produce more precise effect sizes. Therefore, the estimates of effect size should be weighted by the sample size or by the inverse of the variance, (Hedges and Olkin, 1985; Hunter and Schmidt, 1990). Other authors advocate the unweighted meta-analysis (Osburn and callender, 1992). However, Fuller and Hester (1999) showed that some of the advantages of the unweighted average overall effect size predicted by Osburn and Calender’s (1986) simulation did not hold.

**Homogeneity tests:** Moderator variables are another source of variability of results. Indeed, primary studies addressing the similar questions will be expected to vary in a number of ways: definition of outcomes, methods used, situational factors, and so on. Therefore, any meta-analysis should check for a possible heterogeneity of findings. Several authors noted that heterogeneity of results is the rule rather than an exception in the real world data (Hunter and Schmidt, 1990, 2000; Field, 2003; Erez et al., 1996; RC, 1992). Formal statistical tests for assessing the degree of variability of the effect size corrected for artefacts across studies are available (Hedges and Olkin, 1985). Such tests, in spite of their low statistical power for samples of small size can be used to decide whether the primary studies are considered as sharing a common effect size (Field, 2001, Hedges and Pigott, 2004).

**Moderator variables analysis:** Sources of heterogeneity (moderating effects) should be investigated and quantified (Sagie and Koslowsky (1993). Two approach have been proposed:

- **Sub-group analysis:** This investigation is accomplished by breaking out the full set of studies into subsets ones based on a hypothesised moderator variable. However, sub-group analysis tends to increase the number of analyses and hence increase the probability to have significant results by capitalization on chance (type I error inflated). Let’s

---

\(^6\) The effect size has a systematic downward bias.

\(^7\) The methods of estimation of the effect size require the assumption of multivariate normality.
illustrate the problem of capitalization on chance in the following example. Suppose that the primary studies’ results vary according to the estimation method (OLS, GLS, ML)\(^8\) and data characteristics (annually, monthly, weekly). Then, we can conduct more than 15 subgroup analyses: OLS, IV, GLS, OLS-annual, OLS-weekly, and so on. If we add other variables such as environmental characteristics and model specification, the number of possible combinations is so high that some „positive“ results can arise by chance with no plausible explanation (Matt and Navarro, 1997). In addition, the number of sub-group analyses may be limited by the number of observations necessary to obtain a sufficient statistical power. Finally, a sub-group analysis is univariate and thus does not allow controlling for simultaneous factors that may interfere in the relation tested (existence of confounding factors).

- **Meta-regression analysis:** For the reasons cited above, several authors have criticised sub-group analysis (Berkey et al., 1995; Olkin, 2004). They advocate the use of a multiple regression model to account for heterogeneity among studies. In contrast to conventional meta-analysis (sub-group analysis), meta-regression relates the size of effect (regression coefficient) to one or more moderator variables of the primary studies involved. So, this approach is more suitable for measuring the impact of independent variables on the dependant variable (Farley et al., 1995). While often, sub-group analysis focuses only on the percentage of variance explained.

Another advantage of multiple regression is that it is possible to carry out various tests: homoscedasticity, outliers detection, normality, overall goodness of fit. Furthermore, there exist numerous methods of estimation suited to different hypotheses on the linearity of the relations, to the normality assumption and to the existence of outliers (robust estimators). Finally, when the moderator variables are continuous (not categorical) only regression analysis can be carried out. Therefore, meta-regression is flexible enough to enjoy generality (Farley and Lehman, 1986, Olkin, 2004). However, meta-regression has some limitations due to the possibility of multicollinearity among the characteristics of primary studies and aggregation bias because the scarcity of observations for certain characteristics may force the researcher to group them with other ones. In some cases, procedures to solve these problems consist in not taking into account the whole set of relevant information. The estimated model may be hence miss-specified.

**Statistical models for inference:** There are four principal statistical models in meta-analysis for inferring the magnitude of the effect size from a sample of independent results: the fixed-effects (FE) and the random effects (RE) models for a sub-group analysis, and the fixed regression and the random regression models for a meta-regression analysis (see appendix B). These models imply different statistical assumptions. The FE approach assumes that the primary studies included in the meta-analysis are sampled from populations with the same or constant effect size (homogeneous case) or when researchers regard the studies in their meta-analysis as the entire universe of studies of interest rather than the broader task of estimation of the population effect sizes for the given topic as a whole\(^9\) (Hedges and Vivea, 1998, Hunter and Schmidt, 2000). Under these assumptions the FE model has an appropriate level of Type I error (Overton, 1998, Hedges and Vivea, 1998; Hunter and Vivea, 1998). However, if there is evidence of heterogeneity among the population effects\(^10\), then the RE model should be used (Hedges and Vivea, 1998; Hunter and Vivea, 1998). In other words, the conclusions of the meta-analysis describe only the findings of the studies included and cannot be generalized to the entire domain of interest.

---

\(^8\) OLS, GLS and ML stand for Ordinary Least Squares, Generalized Least Squares and Maximum Likelihood, respectively.

\(^9\) Hedges and Vivea (1998) call this case a conditional inference. In other words, the conclusions of the meta-analysis describe only the findings of the studies included and cannot be generalized to the entire domain of interest.

\(^10\) Effects sizes vary randomly from one study to another. In others words, each primary study reviewed comes from a population that is likely to have a different effect size to any other primary study included in the meta-analysis.
and Schmidt, 1990, 2000; Field, 2003). The RE approach assumes that the studies included in the meta-analysis are a sample of all possible studies that could be done on a area of interest. In this case, effect size estimates collected on different local contexts constitute a super-population of random effect sizes characterised by its mean and standard deviation. Thus, the RE model takes into account two sources of variation: within-studies, due to error sampling (i.e., like the FE model), and between-studies, due to differences among studies. Due to its larger confidence intervals the RE model is then less subject to Type I error bias than the FE model is (Field, 2003). For all these reasons, several authors recommend systematic recourse to the RE model (Hunter and Schmidt, 1990, 2000; Hall and Brannick, 2002; Overton, 1998; Field, 2003; Erez et al., 1996). Fixed regression and random regression models are some extensions of, the FE model and the RE model, respectively (see appendix B). Estimation methods of these models are described in several text books (Cooper and Hedges, 1994; Hunter and Schmidt, 1990) and articles (see also appendix B).

Independent samples: The majority of classical estimation methods used in meta-analysis make the assumption that effect sizes are obtained from independent samples. However, in practise the same study with the same sample can report multiple effects. For instance, different procedures can be used on a same sample to estimate a price elasticity. In consequence, these outcomes are potentially correlated and the assumption of independence is violated. Ignoring these inter-correlations among effect sizes conducts to an inflated Type I error rate (Raudenbush et al., 1988; Bijmolt and Pieters, 2001).

One approach, rarely used, to remedy this problem is to restrict the domain of testable hypotheses and to conduct a separate analysis for each effect. The recourse to this procedure is limited by the scarcity of data. Another way is to average out (or to compute the median of) the multiple outcomes for each study. However, this method does not fully account for all the available information and is only appropriate when the multiple outcomes are parallel measures of a single domain or construct (Hedges and Olkin, 1985, Hall, 1994). Several authors (Raudenbush et al., 1988; Gleser, 1994; Sohn; 2000) proposed an alternative method based on weighted multivariate regression for dealing with studies that use multiple effects (see appendix B).

Sensitivity analysis issue. Because the best way to conduct a meta-analysis does not exist and different methods can lead to divergent results, the robustness of the conclusions of the particular meta-analysis should be examined in a thorough sensitivity analysis. Various analyses must be performed to assess the robustness of findings: comparison of the overall effects calculated according to different methods, carrying out various meta-analyses by including and excluding of some primary studies in order to estimate the overall effect with the same method, use of robust methods (Brown, Homer, Inman, 1998), examination of temporal patterns (Kayande and Bhargava, 1994), and publication bias. Finally, meta-analysis applications use various tests for which the calculation of the statistical power is recommended.

Assessing publication selection effects: Publication and selection biases\(^{11}\) occur in a meta-analysis when the effect size estimates are observed in only a subset of the studies that were actually conducted (Hedges, 1992). The dissemination of findings in several journals, laboratories (working papers), conferences and countries modifies the probability that a study is included in a meta-analysis (selection bias). In addition, studies reporting „positive“ results are those that are likely to be published and therefore to be included in a meta-analysis, which may introduce a „positive“ bias (publication bias). Various methods have been developed to identify and remedy publication bias (Rust et al.,

\(^{11}\) A more restrictive definition states that publication bias arises when only studies reporting statistically significant or „plausible“ results are being published. Rosenthal (1979) called this the „file drawer problem“. So, a meta-analysis based on published studies (biased sample) may result in bias in favour of significant or „positive“ findings.
Recent methods for adjusting the meta-analysis for publication bias use weighted distribution theory based on the premise that a study is included in the analysis with a probability determined by the outcome (e.g., p value) (Dear and Begg, 1992). Another simple method to assess publication bias is to calculate a fail-safe \( N \) statistic or “file drawer” \( N \) to estimate the number of unpublished studies with an effect size of zero that would have to exist so as to render the overall effect size insignificant (Rosenthal, 1979).

**Statistical power:** The statistical power is recognised as an important indicator for assessing the validity of results from a research study. Cohen (1988) defines the power of a statistical test of a null hypothesis as “the probability that it will lead to the rejection of the null hypothesis, i.e., the probability that it will result in the conclusion that the phenomenon exists.” Therefore, the power of a meta-analysis is the probability that it will lead to a statistically significant result. In medicine, Flather et al. (1997) showed that invalidation of results of meta-analyses by larger clinical trials for various topics can be explained to a large extent by the fact that meta-analyses are underpowered. Although researchers should conduct power analysis prior to a meta-analysis in order to avoid an underpowered meta-analysis (Flather et al., 1997; Hedges and, 2001; Cohn, 2003; Munzer et al., 2003), it is not always possible to obtain beforehand the necessary data for carrying out this analysis. Generally, it can be only conducted once the meta-analysis is performed.

**Application of meta-analysis in marketing**

**Data.** First, we identified a set of journals which are likely to publish applications of meta-analysis in marketing from 1980 to 2003. Second, the ABI/inform, EBSCO and Sciencedirect databases were used for any articles in other journals that contain applications of meta-analysis in marketing with key terms „Meta-analysis“, „integrative review“, „quantitative review“, „empirical generalizations“, and including key terms used in marketing: marketing, consumer, price, advertising, product, brand, channel, promotions, trust, sales-force and so on. We have also consulted French journals and the English abstracts of the main German and Spanish journals, and identified two articles translated from English. For an application to be included, it had to meet one criterion: it must use meta-analytical techniques. Studies using conventional vote-counting methods were excluded (p.e. Lancaster and Lancaster, 2003; Chetty and Hamilton, 1993; Souza, 2004). As noted by Hedges and Olkin conventional vote counting procedures are very simple, „inherently flawed and likely misleading” (1985, p. 48) (see also Bushman, 1994). The relative contradiction between the results of Andrews and Franke (1991) and those of Lancaster and Lancaster (2003) who used vote-counting approach to study the determinants of cigarette consumption is an illustration. In total, among the 98 applications identified, 68 passed our screening criterion.

Table 1 gives the number of meta-analyses published during the period 1980 to 2003 by journal. It shows that few leading journals in marketing account for the majority of applications and Journal of Marketing Research (JMR) represents about 24 % of the total. The diversity of reviews that published at least one meta-analysis in marketing is encouraging. However, this diversity did not involve an increase in the number of meta-analyses each year. Indeed, the Figure 1 shows that this number remained stable.

---

12 However this statistic offers no guidance when the differences in the file drawer studies are significant but in the opposite direction to the results from the studies included in the meta-analysis (see Bangert-Drowns, 1986).
13 The probability to detect certain differences in treatment effect is low.
14 Principally, the leading journals in marketing.
15 The inclusion criteria used to select primary studies also explain these conflicts results.
over the period. This is confirmed when the number of applications is regressed on the linear and quadratic effects of time. The results show that the linear trend is insignificant. To examine whether the introduction of the meta-analysis approach may be considered a successful innovation in marketing research, we used the discretised version of the Bass diffusion model (1969). We regressed the number of applications per year on the number of cumulative applications up to previous year and the square of the latter term. When several papers are due to the same author(s), only the first one is retained, in conformity with the hypotheses of the Bass diffusion model.

Table 1

<table>
<thead>
<tr>
<th>Title of Journal</th>
<th>Number of applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal of Marketing Research (JMR)</td>
<td>16</td>
</tr>
<tr>
<td>Journal of Consumer Research (JCR)</td>
<td>7</td>
</tr>
<tr>
<td>Journal of Public Policy &amp; Marketing (JPPM)</td>
<td>6</td>
</tr>
<tr>
<td>Journal of Business Research (JBR)</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Marketing (JM)</td>
<td>4</td>
</tr>
<tr>
<td>International Journal of Marketing Research (IJRM)</td>
<td>4</td>
</tr>
<tr>
<td>Marketing Science (MS)</td>
<td>4</td>
</tr>
<tr>
<td>Journal of Advertising research (JAR)</td>
<td>3</td>
</tr>
<tr>
<td>Journal of the Academy of Marketing Science (JAMS)</td>
<td>3</td>
</tr>
<tr>
<td>Psychology &amp; Marketing (P&amp;M)</td>
<td>3</td>
</tr>
<tr>
<td>Others*</td>
<td>13</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>68</strong></td>
</tr>
</tbody>
</table>

* Each of these journals published only one or two meta-analyses.

The OLS estimation method shows that both parameters of the Bass model have expected sign, magnitude and order between them ($p = 0.017 < q = 0.11$) but neither the innovation coefficient ($p$) nor the imitation coefficient ($q$) were significant, indicating that meta-analysis methods are weakly diffused among researchers in marketing$^{17}$.

Figure 1

$^{17}$ For comparison, by using the Bass model, Baumgartner and Homburg (1996) found that the diffusion of Structural Equation Models in marketing is successful.
Two main reasons can be put forward to explain this result:

– Lack of replication\(^{18}\) studies in marketing to perform a meta-analysis. Hubbard and Vetter (1996) noted that replications constitute less than 5% of published empirical work in the management and marketing fields. Unfortunately, these authors noted that the replications of existing studies are depreciated and often rejected by scientific reviews, the latter preferring innovative articles generating new paradigms or methods (see also Wright, 1998). Yet, as underlined by Hunter (2001), \textit{“we need a total sample size of }N = 153,669\textit{ to estimate a causal effect to two digits... If the average sample size were as high as }N = 200,\textit{ we would need over 700 replication studies“} (p. 149); and

– Lack of training or lack interest of researchers in these statistical tools. Of the 68 meta-analyses in marketing considered here, a great number is due to a few authors.

Other relevant statistics for the applications are summarized in the Table 1. It appears that the average sample size reported by applications is of 306 observations. When large studies are excluded, the average sample size decreases to 206. These values are quite high compared to average samples used in other research areas in the social sciences such as psychology (\(N = 100\)) (Hunter and Schmidt, 1992). We also calculated the quartiles of sample sizes, which are robust (less sensitive) to extreme values. The median is of about 195 observations.

Moreover, the median number of articles included in a meta-analysis is high (37 articles). Since the statistical power of a meta-analysis is a function of the sample size as well as of the number of studies (Hedges and Pigott, 2001), one may expect a large statistical power for meta-analyses in marketing.

Finally, the meta-analysis can be used to summarize the results of several studies or to test theories. Most published applications belong to the first category (84.6%). Meta-analyses devoted to theory testing represent 15.4%, a satisfactory proportion though rather low.

\textbf{Issues related to selective inclusion of studies and coding.} It is encouraging that among the applications identified according to some criteria defined to narrow the research domain, a majority (61.7%) covers both published and unpublished articles in English. However, the applications including only studies published in English represent 89.7% of the total. Therefore, it is likely that relevant works written in languages other than English (either published or not) are not included in these meta-analyses. As a consequence, their results can be biased and/or not generalizable for every setting.

Regarding the difficulties related to the coding of the characteristics of studies, there are only a few applications which contain enough information to have a precise idea about inconsistencies in coding and procedures to resolve them. Nevertheless, the few applications (16.2%) reported a high interrater agreement coefficient, about 95% agreement between authors coding the same study and only one meta-analysis reported a low coefficient. Though several authors excluded studies for failure to report sufficient information to calculate effect sizes and/or to carry out necessary tests (statistical validity), no quality criterion other than that of publication status\(^{19}\) was used in the selection of studies.

\textbf{Issues related to measurement of effect size. Common effect size metric:} The choice of indicators and their interpretation are very little discussed by the authors who merely mention the definition of the metric. Table 2 presents the different effect size indicators used in the applications. It appears that regression slopes (\(\beta\)’s) and measures of association are used by 55.9% and 48.5% applications, respectively. Surprisingly, 7.4% of applications reported an overall p-value indicator (significance level) but did not produce neither the magnitude of the effect size nor its variability across primary studies.

\(^{18}\) Hubbard and Armstrong (1994) discuss several reasons for the paucity of replication.

\(^{19}\) Some authors limited their investigations to published articles.
**Table 2**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>Q1(25%)</th>
<th>Q2(50%)</th>
<th>Q3(75%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period of estimation (years)</td>
<td>21,5</td>
<td>12,25</td>
<td>19</td>
<td>27,5</td>
</tr>
<tr>
<td>Number of studies included</td>
<td>42,8</td>
<td>17,00</td>
<td>37</td>
<td>59,0</td>
</tr>
<tr>
<td>Number of measurements of the Effect size per study</td>
<td>5,7</td>
<td>2,14</td>
<td>4,12</td>
<td>7,7</td>
</tr>
<tr>
<td>Sample size (number of observations)</td>
<td>355</td>
<td>133</td>
<td>197</td>
<td>371</td>
</tr>
<tr>
<td>Only published articles</td>
<td>61,7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effect size metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohen’s d and Hedge’s g Pearson’s r</td>
<td>39,7</td>
<td>0,27</td>
<td>0,16</td>
<td>0,26</td>
</tr>
<tr>
<td><strong>Magnitude (absolute value)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indicators (ω², η², ε²)</td>
<td>8,8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression slopes</td>
<td>52,9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>7,4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Screening data and correction for artefacts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sampling error</td>
<td>13,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>measurement error</td>
<td>13,2</td>
<td>0,28</td>
<td>0,082</td>
<td>0,144</td>
</tr>
<tr>
<td>Exploratory analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Statistical methods used for combining and inference</strong></td>
<td>22,1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted effect sizes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent samples</td>
<td>63,7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Homogeneity tests</strong></td>
<td>7,3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Subgroup analysis with</td>
<td>38,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed -Effect model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random-Effect model</strong></td>
<td>42,6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Meta-regression analysis with:</td>
<td>19,1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed-Regression model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Random-Regression</strong></td>
<td>52,8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimator used</td>
<td>0,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GLS</strong></td>
<td>38,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ML</strong></td>
<td>16,2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodness-fit R²</td>
<td>0,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity analysis</strong></td>
<td>38,2</td>
<td>48,2</td>
<td>37,0</td>
<td>48,3</td>
</tr>
<tr>
<td>Statistical power</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail-safe Number</td>
<td>0,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection bias test</td>
<td>20,5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend</td>
<td>0,0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robust regression</td>
<td>8,8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Some applications reported more than one metric.
Although this combined probability method is useful to calculate the so-called fail-safe N number (Rosenthal, 1979), it is not very informative to summarize the results of studies (Hedges and Olkin, 1985; Hunter and Schmidt, 1990). Finally, although effect size measures based on the proportion of variance accounted for are inherently non-directional and can have identical values for two studies with conflicting results or patterns (opposite values), they are still used by 8.8% of applications.

**Correction for artefacts:** The effect sizes observed across studies can be affected by sampling error, measurement error and other artefacts (see Hunter and Schmidt, 1990). While sampling error cannot be corrected for in a single effect size indicator (un-systematic error), measurement error must be (systematic error). Most of the studies did not discuss this point: only 13.2% of them corrected for measurement error and sampling error whereas a great majority of them manipulated latent variables. However, the authors who corrected for measurement errors point out the lack of information in primary studies to compute reliabilities.

To evaluate the importance of the bias in the effect size due to measurement error, we calculated an overall reliability of the applications that reported this index (13.2% of applications). It is equal to about 0.7920 (with s. e. = 0.10). Thus, on average the magnitude of the effect size is systematically reduced by about 21%. Therefore, after correction the average $r$ (not corrected for artefacts) reported across applications, arose from 0.2721 to 0.34 with 95-percent confidence interval [0.078; 0.602]. Finally, the proportion of variance due to artefacts (observed variance/sampling error) represents only to 28% of the observed variance. Moreover, the sampling variance estimators often are biased negatively (Aguinis, 2001). So, our findings indicated that some extent more than 70% of the total variation is associated with other factors (moderators variables).

Thus, if we add the bias due to sampling error of the variance observed across primary findings as well as measurement error, it appears that the magnitude and precision of the effect sizes reported in the majority of applications and not corrected for these artefacts have to be interpreted with caution.

**Exploratory analysis:** 22.1% of the applications carried out an exploratory analysis (see Table 2). This analysis is often limited to a graphical representation of the data. Only 5.9% conducted an analysis to identify outliers and/or to assess the normality of variables. This result is similar with those obtained by Baumgartner et al. (1996) on the applications of structural equation models (8%). It seems then that this step prior to estimation is largely neglected by researchers.

**Estimation and testing procedure issues.** **Weighted effect size:** Encouragingly, 63.2% of the applications used the weighted average of the effect size estimates (see Table 2). However, this proportion decreases drastically when restricting to authors who used regression models. Only 27.8% of meta-regressions used the GLS estimator despite the fact that the primary studies used different sample sizes. The recourse to OLS regression (72.2%) of application used regression model) may threaten the validity of applications’ findings when the assumption of homoscedasticity is violated (Chandrashekaran and Walker, 1993). Sometimes, the authors justify the use of OLS by arguing that the results obtained with the GLS method are similar.

**Homogeneity:** The homogeneity test is of great importance to decide about the choice of the estimation method. For example, the fixed-effects model assumes the assumption of homogeneity. When the effect distribution is heterogeneous, a random-

---

20 For comparison, meta-analysis carried out by Churchill and Peter (1984) on the reliability of rating scale reported an average reliability equal to 0.75 (s. e. = 0.156).

21 Calculated from 402 effects reported by applications which used measures of association to estimate the effect size. It is interesting to note that the observed effect size (0.27) is roughly similar to that obtained by a meta-analysis of effect sizes in consumer behaviour experiments (Peterson et al., 1985).

22 The weighted estimator tends to be unbiased.
effects model is required. Yet, only 38.2% of the applications carried out a homogeneity test to assess differences across studies (see Table 2). Another interesting result is that the vast majority of these tests are significant (assumption of homogeneity rejected), implying that the heterogeneity of the results reported by empirical studies is rather the rule in marketing.

**Moderator variables analysis:** Several approach are available to explain the variability of estimates (see Table 2). However, we observed that the vast majority of meta-analyses used a fixed-effects model (sub-group analysis) or a fixed-effects regression model for any period of publication. The systematic recourse to the fixed-effects model is disconcerting since the homogeneity hypothesis is seldom verified. So, this procedure is only adequate when there is one substantive moderator variable (Schmidt, 2000). In the case where there are more than one moderator variable, the subsets of studies must be internally heterogeneous.

In addition, in the case of sub-group analysis, little attention is given to control for Type I error when conducting multiple tests (capitalization on chance). Unfortunately, only 19.1% of the applications used the RE model recommended by Hunter and Schmidt (1990), of which three recent applications made explicit reference to the RE model (Farrell and Hakstian, 2001; Cano et al., 2004; Jaramillo et al., 2004). Thus, it seems that most of the authors paid little attention to this issue, which again questions the validity of the results obtained. Yet, the use of the FE model for correlation coefficients instead of the RE model in the heterogeneous case, lead to overestimate the effect size by about 15% to 45% depending on the sample size used and to fail to control for Type I error for the associated significance tests for meta-analyses including less than 15 studies (see Field, 2001). Moreover, the probability of detecting small effects was low (Field, 2001, 2003).

**Independent samples:** The vast majority of the applications report multiple measurements of the effect size from the same sample. The median number of measurements of the effect size reported per application is about 4 (see Table 2). These measures are potentially correlated and the assumption of independent findings made in the applications can be violated. Although some researchers are aware of this problem (13.2%), the procedure used to remedy it consists of merely averaging out (or computing the median of) the multiple outcomes for each study. This solution is far from being sufficient. We recommend the use of the new statistical methods which deal with the correlated measurements within studies (Raudenbush et al., 1988b; Gleser and Olkin, 1994; Bijmolt and Pieters, 2001; Sohn, 2000).

**Selection bias test:** Although the methods to deal with the selection bias exist, none of the applications used them whereas the majority of applications included only published English language studies. Yet, as noted in other domains, the inclusion of studies written in other languages can modify the conclusions of a meta-analysis (Grégoire et al., 1995). A possible explanation for the absence of these tests is that the recent techniques to test for publication bias, to the exception of graphical procedures, are mathematically sophisticated.

**Sensitivity analysis:** Few studies conducted sensitivity analyses of results (see Table 2). Thus, none of the applications provided the statistical power of tests used in the meta-analysis. As already noted by Cohen (1988), researchers concentrate too often on the control for Type I error, with little attention being paid to the control for Type II error (Hunter and Schmidt, 1996). As regards, robust methods of estimation of the effect size, they were neglected by almost all the authors. Only one study used robust methods (Brown, Homer and Inman, 1998) to explain the variability across primary results.

---

23 Applications published before 1990 and after 1990 were compared. We found that the practises are similar.
Lastly, only a few authors explored the stability of the overall size time (8.8%) and the number of filed studies - Fail-safe N (20.5%).

In summary, for any period examined, the reviewed applications paid little attention to publication bias, correction for artefacts, stochastic dependencies among effects sizes, adequate methods of estimation and control for Type I error when conducting a large number of tests. However, in overall, it is not clear whether the findings are biased and, if so, in which direction. Indeed, some threats (e.g. publication bias, FE approach) tend to overestimate the magnitude of the effect size while others (unreliability of measures used) tend to underestimate it.

Discussion and perspectives. In view of all the insufficiencies discussed above, it is legitimate to question the validity of the empirical generalisations drawn from the applications of meta-analysis in marketing. To resolve this important issue, we evaluated the statistical power of the different tests used in the applications. We used the procedures proposed by Hedges and Pigott (2001, 2004) to compute the statistical power levels for various tests used by researchers.

Power of significance tests of measures of association: About 55.8% of the „true“ effect sizes computed in the applications are measures of association. In total, among 402 effects size reported in the applications, we found 358 for which there is sufficient information to calculate the statistical power. We converted these measures into a common metric (Pearson’s r). The computation of the power $p$ of the two-tailed test for an effect size $\rho = \rho_0$ is given by:

$$
p = 1 - \phi(C_{\alpha/2} - \lambda) + \phi(-C_{\alpha/2} - \lambda),
$$

where $\phi$ is the standard normal cumulative distribution function, the parameter $C_{\alpha/2}$ is the 100(1-α/2) percent point, $\sqrt{v}$ is the standard error of $\rho$, $\rho$ is the correlation coefficient and $\lambda$ is the mean of the normal distribution of $z = (\rho - \rho_0)/\sqrt{v}$ when the null hypothesis is false ($\rho \neq \rho_0$). The value of $v$. depends on the model used to estimate the overall effect size $\rho$: $v. = 1/\sum w_j (w_j = 1/\sigma_i^2)$ the case of the FE model (homogenous case) and $w_j = (\sum_i \sigma_i^2 + \tau^2)^{-1}$ in the RE model (heterogeneous case). When the between-studies variance component $\tau^2$ was not reported, to compute the statistical power of significance test the RE model we adopted the convention suggested by Hegdes and Pigott (2001), namely that $\tau^2 = 0.33 \sum_i \sigma_i^2$ is a small degree heterogeneity, $\tau^2 = 0.66 \sum_i \sigma_i^2$ (k is the number of studies reviewed) is a medium degree heterogeneity and $\tau^2 = \sum_i \sigma_i^2$ is a large degree heterogeneity.

Illustration Geyskens et al. (1999) carried out a meta-analysis of satisfaction in marketing channel relationships. For each relationship examined, they computed the coefficient of correlation (effect size), the number of independent studies $k$, and the number of observation of each study. Unfortunately, these authors did not reported the variance of the effect size, so to calculate the power for Geyskens et al. (1999) review, we use the Fisher’s $z$-transform of the observed correlation where $z = 1/2ln[(1 + r)/(1 – r)]$. Indeed, the variance of $z$-transform is $\sigma_z^2 = 1/(n_j - 3)$ which depends only on sample size, $n_j$, which is often reported in the applications. In addition, we suppose the values of $\sigma_i$ are

---

24 As most of the reviewed applications used the FE model, $\tau^2$ is rarely reported.
identical to obtain an estimate of $\nu$, which depends on $n_i$ and the number of studies $k$. Here, we examine only the relationship between economic satisfaction and trust. Geyskens et al. (1999) reported an average sample per study $n = 94.4$, the number of studies $k = 5$ and average effect size $r = 0.393$ ($z = 0.412$). If we suppose that the values of $\sigma_i$ are approximately identical ($n_i$ constant across studies) then $\sigma^2 = (94.4 - 3)^{-1} = 0.0109$. Therefore, an estimate of $\nu$ can be computed as $\nu = \frac{\sigma^2}{k} = 0.002188$.

The power of the two-tailed test at $\alpha/2 = 0.025$ for $z = 0$ (or $\rho = 0$) inferred by the FE model requires the computation of $\lambda = \frac{0.415 - 0}{0.002180.5} = 8.79$. Then the power test to reject the null hypothesis ($\rho = 0$) is given by $p = 1 - \Phi(1.96 - \lambda) + \Phi(-1.96 - \lambda) = 1$. If the effect size is inferred by RE model, to compute statistical power, we need the values of $\rho$ and $\tau^2$. However, Geyskens et al. (1999) did not use the RE Model, so $\tau^2$ is unknown. Then, following the recommendations of Hedges and Pigott (2001), we posit different value for $\tau^2$: a small degree of heterogeneity ($\tau^2 = 0.33 \sigma^2_i$), medium degree of heterogeneity ($\tau^2 = 0.66 \sigma^2_i$) and large degree of heterogeneity ($\tau^2 = \sum_i \sigma^2_i$).

Besides, under heterogeneous hypothesis of effect size, Field (2001) shows that the FE model for correlation coefficients tended to overestimate effect sizes by about 15% to 45% (mean = 30 %). Then, we posit $r = \frac{0.393}{1 + 30\%} = 0.302$ as the value inferred by the RE model. A moderate amount of heterogeneity the power is $p = 0.999$. A large heterogeneity the power is $p = 0.997$. Thus, given our assumptions, under FE model or RE model, we have very satisfactory power to detect at least a correlation of 0.3 between trust and economic satisfaction.

We applied these procedures for the tests reported by applications. Encouragingly, for any the model used (FE and RE models) in the applications for inference about effect size, we have reasonable power level to detect a significant population effect size (reject the null hypothesis $\rho = 0$). Only 18% of the applications that used the FE model have a low power (< 0.8); the proportion increases to 28% for those applications that used the RE model (Table 3).

**Power of goodness-fit of the regression model:** The lack of necessary information did not allow us to calculate the power of significance tests of the $\beta$ slope index. However, we computed the statistical power of the test of goodness of fit of the regressions. The weighted sum squares about regression line is used as a test of Goodness of fit\(^{25}\). As for the procedure referred to above, we computed the statistical power of this test for fixed-regression model and random-model under three assumptions: small, medium and large heterogeneity between effect sizes (Hedges and Pigott, 2004). The results obtained (see Table 3) show that the statistical power is low in the case of a large heterogeneity between effect sizes, indicating that the models estimated were not correctly specified. However, the latter result is to be interpreted with caution because the parameters used to calculate the statistical power were estimated by means of the fixed-regression model instead of the random-regression model (as required).

**Power of homogeneity tests:** More than one third of the applications used a test of homogeneity of the effect sizes. We computed the statistical power of these tests under three assumptions: small, medium and large heterogeneity between effect sizes (Hedges and Pigott, 2001, 2004). Table 3 shows that these tests have low power to detect small and medium heterogeneity for a population of effect sizes. Consequently, the homogeneity

\(^{25}\) This test compares the null hypothesis which assumes that the estimated model holds with the alternative hypothesis which assumes that at least one of the parameters of the model is not significant in the linear relation.
hypothesis can be wrongly accepted when differences between effect sizes are of small or medium magnitude. Fortunately, all of the homogeneity tests reported in the applications are generally significant, indicating a strong heterogeneity of the primary findings. This result also shows that the use of the FE model in these applications was not adequate.

### Table 3

<table>
<thead>
<tr>
<th>Significant tests used for</th>
<th>% of applications with statistical power p:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p &lt; 0.5</td>
</tr>
<tr>
<td><strong>The mean effect size (Pearson’s r) in:</strong></td>
<td></td>
</tr>
<tr>
<td>FE model</td>
<td>11</td>
</tr>
<tr>
<td>RE model with</td>
<td></td>
</tr>
<tr>
<td>small degree heterogeneity</td>
<td>17</td>
</tr>
<tr>
<td>medium degree heterogeneity</td>
<td>20</td>
</tr>
<tr>
<td>large degree heterogeneity</td>
<td>19</td>
</tr>
<tr>
<td><strong>Goodness-fit of regression model in</strong></td>
<td></td>
</tr>
<tr>
<td>Fixed regression model</td>
<td>0</td>
</tr>
<tr>
<td>Random-regression model</td>
<td>0</td>
</tr>
<tr>
<td>small degree heterogeneity</td>
<td>0</td>
</tr>
<tr>
<td>medium degree heterogeneity</td>
<td>37</td>
</tr>
<tr>
<td>large degree heterogeneity</td>
<td></td>
</tr>
<tr>
<td><strong>Homogeneity with</strong></td>
<td></td>
</tr>
<tr>
<td>* Only for subgroup analyses involving at least four studies.</td>
<td></td>
</tr>
<tr>
<td>small degree heterogeneity</td>
<td>90</td>
</tr>
<tr>
<td>medium degree heterogeneity</td>
<td>58</td>
</tr>
<tr>
<td>large degree heterogeneity</td>
<td>28</td>
</tr>
</tbody>
</table>

All these results are encouraging, the power levels of the tests used in the applications being satisfactory. Thus, one can assume that when reported estimates of the effect size are considered to be significant they really are. But, a doubt subsists as to their precision (level and range). Indeed, the procedures and models used in the applications involve estimation bias of effect size.

Meta-analysis is recognised as a powerful procedure for summarizing, exploring moderator variables and generalizing conflicting empirical findings from various studies. As stated by Hunter and Schmidt (1992), “meta-analysis is not merely a new way of doing literature reviews. It is a new way of thinking about the meaning of data, even our views of the individual empirical study and perhaps even our views of the basic nature of scientific discovery” (p. 1173). Unfortunately, only 68 applications of meta-analysis in marketing were published over the period under review (1980–2004). Therefore, it is advisable to encourage researchers in marketing to use these techniques and journal editors to publish replication and extension studies in order to improve the primary findings because the generalisation of results from a small number of studies cannot be conclusive (Lindsay and Ehrenberg, 1993; Hubbard and Vetter, 1996; Hunter, 2001).

However, meta-analytical techniques are not easy to use. The lessons drawn from the applications reviewed in this paper show that numerous lacunas subsist, threatening the validity of the conclusions. The most important, in our view, is that researchers focus mainly on the substantive difference between studies while little interest is given to problems related to errors of measurement and to the sampling scheme. A second threat to the validity of generalization is the lack of interest in screening the data to, for instance, identify outliers or assess the assumption of normality. A third limitation is the
predominant use of unweighted regression analysis to assess moderator variables, which ignores differences in precision. A fourth threat to the validity of the conclusions resides in the potential publication bias. Finally, the excessive use of to the fixed-effects model also constitutes a limitation. Several simulation studies showed that the random effects model is better suited to real data (Field, 2003; Hunter, 2001).

To protect against these threats, we thus recommend the following nine-steps procedure:

1) delimit precisely the research area for identifying the relevant literature;
2) pre-specify which covariates are going to be examined;
3) assess by two ore more reviewer in order to check the relevance and the quality of each article retrieved, and to test the reproducibility of coding decisions;
4) screen the estimates obtained from primary studies;
5) correct each estimate of the primary studies for artefacts;
6) test the overall homogeneity of study-level effect size estimates;
7) use meta-regression (fixed or random) when two ore more moderator variables are investigated. It is recommended to use a random-regression;
8) compute the statistical power of the meta-analysis;
9) conduct a sensitivity analysis.

This procedure has not been applied to the applications examined in this article because the data used by the researchers are not available. Therefore, we recommend that journal editors adopt policies requiring all authors using meta-analysis to provide with the data used in their manuscripts26. Such practice is relevant in as it allows: verifying the validity of the results in light of new methodological improvements of the meta-analysis, and re-examining these results as new studies become available.

**Conclusion.** In fact the conflicting results reported by empirical studies on the same topic disappoint not only researchers but also managers and policymakers. Meta-analysis is one of the methods used to explain and to handle the variability findings across studies. This article is aimed at providing a state of the art of the applications of meta-analytical techniques in marketing research, particularly how researchers deal with threats to validity of meta-analysis (the „apples and oranges“ problem, selection bias, estimation bias). This article shows that although many researchers are aware of at least one of these validity threats, most of them continue to use straightforward methods, despite of numerous insufficiencies. Consequently, the conclusions of the great majority of the applications reviewed are to be interpreted with caution, particularly the precision of magnitude of effect sizes reported. Moreover some threats to the validity of inference about the existence of a relationship between variables, such as unreliability of coding, stochastic dependencies among effect sizes and bias of publication, have not been examined in the present research, because of the lack information reported in the reviewed applications.

In future, we hope that the „complexity“ of recent methods such as random effects models, publication bias tests and robust regressions should not discourage researchers from using them in performing a meta-analysis. Some of those statistical tools are now available in well-known statistical software packages such as STATA and SAS. All the more, several authors showed their efficacy, though they did not take into consideration for all the critics raised in the literature about the use of meta-analytical methods.

Finally we recommend that replication and extension of existing empirical studies should be in order to build up a large database of facts on each particular research topic, thereby increasing the number of application of meta-analysis in marketing. Indeed, as underlined by Hubbard and Vetter (1996), „many empirical findings in the business literature are isolate and fragile, as they have been largely immune from examinations designed to assess their reproducibility and generalizability“ (p. 62).

---

26 The data may be made available upon request or through the Internet.
Bibliography

22. Flather Marcus D. Strengths and Limitations of Meta-Analysis: Larger Studies May Be


44. **Matt George E.** What Meta-Analyses Have and Have Not Taught Us About Psychotherapy
Appendix A
Applications of meta-Analysis reviewed


Appendix B

Models for inference in meta-analysis

The problem is to estimate the „true“ value of the unknown effect size $\beta$ observed from $k$ independent studies. We denote $\hat{\beta}_i$ the population effect size observed in the $i$th study, $i = 1, 2, \ldots, k$. Four somewhat different statistical models can be used for inference about the „true“ effect size $\beta$. The choice of a statistical method is crucially determined by the assumptions made on the statistical process followed by this parameter.

Fixed-effects model

The true parameter $\beta$ is assumed to be the same in the different studies. So, the estimates $\hat{\beta}_i$ could be set as follows:

$$\hat{\beta}_i = \beta + \varepsilon_i \tag{A-1}$$

Assuming $\hat{\beta}_i \sim N(\beta; \sigma)$ and the sampling error $\varepsilon_i \sim N(0; \sigma_i)$, then the estimator of $\beta$ that minimizes the variance is
\[
\hat{\beta} = \frac{\sum_{i} \hat{\beta}_i w_i}{\sum w_i} \tag{A-2}
\]

with \( w_i = 1/\hat{\sigma}_i^2 \), \( \hat{\sigma}_i \) is an approximately unbiased estimate of \( \sigma_i \).

**Fixed-effects regression model**

A natural extension of the fixed-effects model is to suppose that the true effect is depending on a set of study characteristics \( X(X_{i1}, \ldots, X_{im}) \), where \( X_{i1} \) is a row vector that contains the values of the covariates for study \( i \) and \( \alpha \) is a column vector of regression coefficients.

So, equation (A-1) becomes

\[
\hat{\beta} = X\alpha + \varepsilon. \tag{A-3}
\]

OLS can be used if \( \hat{\sigma}_i \) across studies are approximately equal (homoscedasticity assumed), otherwise a weighted-least-squares (WLS) procedure is appropriate. So, the estimator of \( \alpha \) is \( \hat{\alpha} = (X'VX)^{-1}X'V\hat{\beta} \), with \( V \) being the covariance matrix. This matrix is diagonal when the estimates are independent. When the estimates are correlated, \( V \) is a full symmetric matrix. \( V \) has diagonal terms of the form \( \sigma_i^2 + \tau^2 \) and sub-diagonal terms are the covariance between estimates of the same study (correlated effect sizes).

**Random-effects model**

The random effects model assumes that the studies are a random sample of a much larger population of studies and that their results have different true effects. In this case, the true parameter \( \beta_i \) follows a random process \( \beta_i = \beta + \nu_i \), with \( \beta_i \sim N(\beta; \tau^2) \), where \( \tau^2 \) is the between-study variance or the variance of the population from which \( \nu_1, \ldots, \nu_k \) are sampled. The random term \( \nu_i \) gives the measurement of the specificity of the \( i \)th study.

It is convenient to decompose the observed effect into fixed and random components. Indeed, as \( \hat{\beta}_i = \beta_i + \varepsilon, \) \( \hat{\beta}_i^* = \beta + \nu_i + \varepsilon. \) Assuming \( \varepsilon \) and \( \nu_i \) are not correlated, then \( V(\hat{\beta}_i^*) = V(\varepsilon_i + \nu_i) = \sigma_i^2 + \tau^2. \) So \( \hat{\beta}_i \sim N(\beta, \sigma_i^2 + \tau^2) \) and the estimate of the „true“ effect size becomes:

\[
\hat{\beta} = \frac{\sum_{i} \hat{\beta}_i w_i}{\sum w_i} \tag{A-4}
\]

with \( w_i = (\sigma_i^2 + \tau^2)^{-1} \).

**Random-effects regression model**

An extension of the random model is to suppose that the true effect is depending on a set of study characteristics \( X(X_{i1}, \ldots, X_{im}) \) plus a random term: \( \beta = X\alpha + \nu. \) Then,

\[
\hat{\beta} = X\alpha + \nu + \varepsilon. \tag{A-5}
\]

In other words, equation (A-5) assumes that only part of variability in the true effects is explainable by studies characteristics \( (X\alpha) \). In contrast, the fixed-effects regression model supposes that the study characteristics account completely for variation in the true effect sizes (\( \tau^2 = 0 \)).

The residual variance \( V(\hat{\beta}) \) will be heteroscedastic as long as \( \nu \) varies across studies. Clearly, it would be inappropriate to use OLS to estimate both the unknown values \( \alpha \)
and $\tau^2$. In large samples the ML estimates are efficient (Hedges, 1992; Berkey et al. 1995, Sohn, 2000). For other alternative estimators (see Sohn, 2000).

Note that the equations (A-3) and (A-5) are sometimes called the meta-regression.

UDC 330.3+338.1 (477)

Dr. Sc. (Economics), prof. Nataliia Stukalo,
Dr. Sc. (Economics), prof. Nikolay Derkach

Oles Honchar Dnipropropetrovsk National University (Ukraine)

GLOBAL AND LOCAL DETERMINANTS OF ECONOMIC GROWTH AND COMPETITIVENESS OF UKRAINE

The article is devoted to investigation of key factors that influence Ukrainian economy further development. The main determinants of economic growth and competitiveness improvement of Ukraine are considered. These determinants were divided into local (corruption, bureaucracy, ineffective infrastructure and legislation, political instability etc.) and global (global financial crisis, geopolitical and geoeconomical situation, integration processes level etc.).

Key words: economic policy, economic growth, competitiveness, global and local determinants.

Introduction. The complex concept of national economy competitiveness includes main success determinants of the country development. The high level of nation competitiveness is characterized by strong and flexible economy, innovative business environment, investment potential, dynamic market growth, effective companies and social institutions. However current conditions and level of country’s competitiveness is under strong impact of both global and local determinants. It’s essential to consider them in order to form optimal policy of economic growth and competitiveness improvement. The issue has been investigated by many famous researchers including Timo J. Hämäläinen, S. Demurger, J. Sachs, W. T. Woo, S. Bao, G. Chang, D. Evans, and others.

However global financial crisis as well as some other current changes caused changes of key determinants of countries’ economic growth and competitiveness.