# THE (IM)PRECISION OF SCHOLARLY CONSUMER BEHAVIOR RESEARCH

David Trafimow
Distinguished Achievement Professor of Psychology
New Mexico State University
Department of Psychology, MSC 3452
Box 30001
Las Cruces, NM 88003
Voice Phone: 575-646-4023
Email: dtrafimo@nmsu.edu

Michael R. Hyman (Corresponding Author)
Distinguished Achievement Professor of Marketing
New Mexico State University
College of Business
Box 30001, Dept. 5280
Las Cruces, NM 88003-8001
Voice Phone: 575-646-5238
Email: mhyman@nmsu.edu

Alena Kostyk
Lecturer in Marketing
University of Glasgow
Adam Smith Business School
University Avenue, Glasgow G12 8QQ
E-mail: Alena.Kostyk@glasgow.ac.uk



Declarations of interest: None.

**THE (IM)PRECISION OF SCHOLARLY CONSUMER BEHAVIOR RESEARCH**

**Abstract**

Knowing related empirical facts as precisely as possible is crucial to knowledge development. Does the sampling precision of published consumer research ensure it contributes meaningfully to marketing science? To answer this question, the sampling precision of four consumer-centric journals selected for their relative impact factors—*Journal of Consumer Research* (JCR), *Journal of Consumer Psychology* (JCP), *Journal of Consumer Marketing* (JCM), and *International Journal of Consumer Studies* (IJCS)—is compared. Based on a recently developed a priori procedure, analyses of articles published between 2000 and 2016 determined the precision of the reported sample means that estimate corresponding population means. Results show studies in all journals lack sufficient precision, with JCR and JCP studies half as precise as JCM and IJCS studies. Sampling precision's value to scientific advancement partly reflects its connection to replication probabilities. Given low replication rates, the most cited consumer-related studies, which strongly influence subsequent research, may be the most misdirecting.

**Keywords**: precision, sampling precision, a priori procedure (APP), replication, replication probabilities

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# THE (IM)PRECISION OF SCHOLARLY CONSUMER BEHAVIOR RESEARCH

## 1. Introduction

Understanding consumer behaviors, including the cognitive and affective bases for those behaviors, is scholarly consumer research's raison d'être. Like marketing research in general, consumer research often blends basic and applied scholarship. Consequently, empirical facts about consumers' behaviors energize both research spheres.

From a basic scholarship perspective, philosophical approaches to theory testing include verification (e.g., Hempel, 1965), falsification (e.g., Popper, 1959), and abduction (e.g., Haig, 2014). These approaches differ in two regards:

(1) Prescriptions about whether theory development should precede empirical fact-finding (e.g., predictive testing (Basmann, 1965; Bass, 1969)) or empirical regularities should precede theory development (e.g., double jeopardy (Ehrenberg, Goodhardt, & Barwise, 1990))

(2) Theory-fact orders responsive to additional issues, such as (a) auxiliary assumptions connecting non-observational theoretical terms (e.g., brand loyalty) to observational terms (e.g., repeat purchase frequency), (b) assumptions about the generalizability of obtained empirical facts (e.g., brand choice follows a beta-binomial/negative-binomial distribution (Ehrenberg, 1988); brand switching follows a zero-order Markov process (Bass, 1974)), and (c) relevant theoretical domains

Theory-testing approaches also differ by theory type, with prescriptions about theories as inductions from empirical facts, deductions based on speculations, inferences to best explanations, and the like. Despite philosophical differences, one crucial similarity among these approaches suggests the research presented here. Specifically, theory development and testing—

whatever the underlying theory's nature—eventually entails contact with empirically generated facts. At that juncture, *knowing empirical facts as precisely as possible is essential to scientific progress*.

The stakes are high. As is shown subsequently, imprecision implies irreproducibility. The *National Association of Scholars* has identified irreproducibility as its most crucial current concern across academic fields (Randall & Welser, 2018). Also, the lack of precise empirical facts militates against productive theorizing and useful applications (Trafimow, 2019a). The history of science shows basic and applied research in fields with more precise empirical facts (e.g., physics, chemistry, genetics) tend to outpace research fields with less precise empirical facts (e.g., psychology, management, marketing). Thus, the importance of precision in science generally or consumer research specifically cannot be overstated.

To control for journal prestige and publication standards, the sampling precision of findings (i.e., precision related to the empirical study's sample size) reported in four respected and similarly focused consumer research journals with diverse impact factors is investigated here. Although sampling precision problems in marketing scholarship likely extend beyond consumer research, conventional wisdom and limited data collection budgets make consumer research especially susceptible. For example, McQuitty (2004) discourages SEM-based studies with overly large samples to avoid "overrejection of tenable models" (p. 175). Because consumer research often relies on researcher self-funding, modest university grants, and low/no-cost student or crowdsourced (e.g., mTurk) samples, instead of corporate sponsorship or secondary corporate/government data, larger non-convenience samples often are cost-prohibitive (Ashraf & Merunka, 2017; Bloom & Milne, 1991; Espinosa & Ortinau, 2016; Peterson, 2001).

> Consumer research has grown and advanced without the benefit of significant amounts of funding….The vast majority of the studies…have been paid for out of either researchers' own pockets or their school's limited research funds. Big government or foundation grants to support programmatic research efforts have played a minor role (Bloom & Milne, 1991, p.255).

In essence, if a sampling precision problem exists in marketing scholarship, then it is most likely exists in consumer research, i.e., consumer research provides an excellent initial test case. Hence, the four subsequently examined research questions are as follows:

- Is precision in consumer research increasing or decreasing?
- Is the increasing number of experiments per article concomitant with decreasing precision per experiment?
- In consumer research articles, are a study's serial position (i.e., position within the sequence of summarized studies) and its precision related?
- What is the probability of successful replication for consumer research studies when desired sampling precision is determined a priori?

## 2. Sampling Precision

From an applied scholarship perspective, empirical fact precision is essential. Researchers have one of four goals for applied scholarship (Trafimow, 2016):

(1) <u>Develop applicable theory</u>. Precise empirical facts about the physical universe are as critical to theoretical physicists as precise empirical facts about consumers are to marketing scholars. Fraunhofer's discovery of his eponymous lines in the solar spectrum eventually led to Hoyle's theories about stellar nucleosynthesis (i.e., how nuclear

reactions within stars created elements heavier than helium) (Hoyle, 1957). Brahe's more accurate astronomical observations helped Kepler deduce his laws of planetary motion and provided compelling evidence for a heliocentric solar system (Stephenson, 1987). In marketing, low explained variance in published empirical studies on consumer purchases inspired Bass' theory of stochastic preference and brand switching (Bass, 1974).

(2) Search for subcategories of auxiliary assumptions to connect theories with applied domains. Applications of scientific theories or laws demand auxiliary assumptions, for which empirical fact precision is crucial for (dis)confirmation (O'Shaughnessy & O'Shaughnessy, 2002). Consider Newton's law of action and reaction, which alone is insufficient to guide rocket engine design. For internal combustion engines, rocket scientists also must make assumptions about thrust-producing flammable chemicals (e.g., hydrazine) and their storage. Without viable (i.e., sufficiently energy-dense and pre-ignition-stable) fuel, rocket scientists cannot connect Newton's law to escaping Earth's gravity. Thus, applied—rather than basic or theoretical chemistry research—is requisite to current spacecraft design.

Consider 'exchange of values', which many scholars have argued is marketing's core notion (Bagozzi, 1975; cf. Hyman, 2004; Kotler & Levy, 1969). Marketing scholars cannot apply the 'exchange of values' to marketing practice without making assumptions about ownership, rights, values, needs, and the like (Hyman, Skipper, & Tansey, 1991). Thus, marketers cannot apply the 'exchange of values' in a hypothetical assumption-free universe. Consumer researchers must deal similarly with the theory of reasoned action (TRA) or theory of planned behavior (Fishbein & Ajzen, 1975; 2010). The theory alone is insufficient to guide product development and selling efforts; instead, researchers must

rely on elicitation studies to uncover the beliefs and evaluations that motivate consumers' behaviors in certain situations.

Note that useful research need not be theoretical nor pertain strictly to a specific application. For example, research on metal fatigue suffered by alloys is worthwhile because it provides auxiliary assumptions pertinent to building large structures like suspension bridges, airplanes, and skyscrapers. Thus, applied research without immediate application can advance theory and practice.

(3) Combine theory with validated auxiliary assumptions to develop or use an application. Conducting the chemistry research requisite to synthesizing a viable rocket fuel—in fact, testing all auxiliary assumptions associated with spacecraft design, manufacture, and operation—is insufficient. Space missions require engineers to design launch and vehicle systems fabricated from sufficiently lightweight yet robust materials, fill fuel tanks, and plot courses to nonstationary celestial objects. This integrative effort may be formidable even when all mission-related components are known and accessible. In essence, understanding all spacecraft aspects differs from creating working flight systems.

When using TRA to create devices for redirecting consumer behavior (e.g., creating public service announcements to reduce unhealthful behaviors), understanding all model components individually does not guarantee their amalgamation will create an effective intervention. Some psychologists criticize TRA because interventions based on it fail despite combining only non-defective elements (Sniehotta, Presseau, & Araújo-Soares, 2014). Regardless, this integration distinguishes (3) from (2).

(4) Conduct non-theory-based research; for example, when Johann Vaaler invented the paper clip (Petroski, 1992). In such cases, the relevance of empirical fact precision is self-

evident, as facts represent the entirety of the researcher's contribution. In marketing, practice-driven studies to identify new product configurations, or social-media-based A/B tests meant to increase consumer conversion rates, fall into this category.

Regardless of the underlying philosophy of science, whether the research is theoretical or applied, or the degree theoretical and applied perspectives are blended, knowing empirical facts as precisely as possible is irreplaceable (unless trying to persuade by exaggerating, distorting, or fabricating the facts). For instance, radio stations charge per 1000 ad exposures, yet advertisers know the ratings data used to estimate total exposures by station and time slot are imprecise due to measurement limitations (e.g., Neilson's portable people meters and selective listener surveys only partly overcome the errors and omissions that plagued Arbitron's diary panels) (Fybush, 2015). Alternatively, audio podcasters can provide a more precise 'unique monthly audience' metric determined by web analytics; in response, podcast advertising revenues are increasing (Weissbrot, 2018). Without precise measurement of critical externalities, making optimal marketing decisions is impossible.

The three precision types are measurement precision, homogeneity-induced precision, and sampling precision (Trafimow, 2018a). Measurement precision, which relates inversely to an instrument's reliability, is an aspirational goal for consumer researchers, as measuring humans is constrained structurally and temporally (e.g., test-retest reliability for most consumer measures falls far short of 1.0; see, for example, Beatty & Kahle, 1988; Malhotra, 1981; Shimp & Bearden, 1982). Under homogeneity-induced precision, reduced variance attributable to querying similar respondents, either across or within groups, boosts the signal-to-noise ratio, thus easing focal effect discernment. Discouraged for 'effect application studies', which are trustworthy for estimating population effect sizes only for data from representative respondent samples, more

homogeneous respondent pools (e.g., undergraduate students) are acceptable for theory testing and scale development (Calder, Phillips, & Tybout 1981; Lynch 1982, 1983, 1999). Although these two precision types warrant further study, only sampling precision is studied here. Assuming random selection, the larger the sample size, the more precisely summary statistics estimate population parameters of interest (e.g., increased likelihood of a sample mean proximate to the population mean).

Why do consumer researchers query a subset of population members rather than a single member? As the former is more onerous and expensive to query than the latter, they must have accepted rationales. For example, statistically oriented researchers might claim 'as sample size increases, researchers are increasingly confident the sample mean approximates the population mean'. This rationale suggests two additional questions:

(1) How approximate is 'approximate'?

(2) How confident is 'confident'?[1]

Based on these two questions, Trafimow (2017) derived an equation for calculating the sample size $n$ to reach specified levels of proximity and confidence for a single mean. This equation provides the basis for the empirical study that follows.[2]

The a priori procedure (APP) in Trafimow (2017) requires researchers to specify (1) the targeted difference (in standard deviation units) between the sample and population means, and (2) the probability of being within that difference. A researcher decides the desired precision level by specifying the fraction of a standard deviation $f$ within which the sample mean should fall relative to the population mean. Also, researchers specify the desired confidence in obtaining a sample mean within $f$ of the population mean. Equation 1 is based on $f$ and the $z$-score $Z_C$ corresponding to the desired confidence level:

(1) $$n = \left(\frac{Z_C}{f}\right)^2.$$

For example, suppose researchers want a 95% probability of a sample mean within one-tenth of a standard deviation from the corresponding population mean. In that case, $Z_C = 1.96$ and $f = .1$, so $n = 384.16$. Hence, researchers must query 385 participants to meet precision and confidence specifications.

Although the focus in Trafimow (2017) is on using Equation 1 before determining the sample size needed to meet specifications, a posteriori use also is possible. Consider this simple algebraic rearrangement of Equation 1:

(2) $$f = \frac{Z_C}{\sqrt{n}}.$$

Now suppose researchers query 200 participants and want to be 95% confident in the sampling precision. The calculation is as follows: $f = \frac{1.96}{\sqrt{200}} = .14$. In general terms and at the conventional 95% level of desired confidence, Equation 2 gives the sampling precision. Because *f* represents proximity to the population mean in standard deviation units, lower (higher) *f* would indicate higher (lower) sampling precision.

The Trafimow (2017) equations are limited because they pertain to only a single mean rather than the multiple groups and factorial designs typically studied by social scientists. Consequently, Trafimow and MacDonald (2017) expanded the APP to include $k$ means; hence, researchers can calculate proximity to the mean, $f$, based on $k$ groups in an a posteriori fashion. Equation 3 is an algebraic rearrangement of an equation presented in Trafimow and MacDonald (2017), which included a proof:

(3) $$f = \frac{\Phi^{-1}\left(\frac{\sqrt[k]{p(k\ means)}+1}{2}\right)}{\sqrt{n}}.$$

### 2.1 Replication Probability and Its Relation to f, n, k, and p (k means)

To ensure the generalizability of reported findings, marketing thought leaders have repeatedly urged marketing researchers to replicate 'one-off' studies (Hunter, 2001; Park et al., 2015; Royne, 2018). Inter-study replications are tests of prior study results by independent researchers (Kwon et al., 2017). Disconnected teams ensure the inter-study replications vital to scientific advancement (Open Science Collaboration, 2015).

Notwithstanding the ongoing debate about research reproducibility in the social sciences (Center for Open Science, 2017), which has prodded some scholars to discourage practitioners from applying findings reported in marketing journals due to a dearth of convincing replications (Evanschitzky et al., 2007), history suggests marketing scholars generally continue to ignore calls for increased replication research. The number of published inter-study replications in marketing remains low despite extensive researcher and editorial office efforts to encourage replication study submissions (Evanschitzky & Armstrong, 2013; Kwon et al., 2017). Although some marketing scholars may welcome an increase in credibility-enhancing concordant replication studies, other scholars may question the contribution to the extant literature (i.e., many scholars believe 'confirming the known' is minimally informative). Conversely, few marketing scholars will appreciate replication studies discrediting their published work (Kerr, Schultz, & Lings 2016).

If most marketing scholars view replication research as dull or threatening, or resist it due to sociology-of-science-related factors (e.g., publication requirements for tenure; discomfort refuting colleagues in print), then neither goading by thought leaders nor restructuring doctoral student training (e.g., creating publication outlets for doctoral students' replication studies (Easley, Madden, & Dunn, 2000)) will boost replication rate. Hence, *the next best alternative is*

*to increase the likelihood of a study's confirmation had it been replicated*. In essence, marketing scholars can reduce erroneous published findings by boosting sampling precision (i.e., communitarian approach) rather than boosting replication rate (i.e., failed policing approach).

Whether called 'statistical replication' (Hunter, 2001) or 'pure replication' (Toncar & Munch, 2010)—using the same measures and procedures as the original researcher(s), including sampling the same population—is especially useful for verifying reports detailing novel behavioral effects (Lindsay & Ehrenberg, 1993). The success of such replication ties firmly to the sample analyzed in the original research (Anderson & Maxwell, 2017). Hence, researchers should attend to sampling precision, as low replication rate makes the probability of successful replication paramount (i.e., if replications are rare, then research trustworthiness is even more critical because confirmation is unlikely and erroneous findings are more likely to influence subsequent research and practice).

Equation 4 from Trafimow (2018b), where $p(k\ \text{Means} - \text{rep})$ denotes the probability all group means are within $f$ of their corresponding population means in the original and subsequent replication experiments, shows sampling precision and statistical replication probability are connected (subsequently, 'replication probability'). In Equation 4, $\Phi$ denotes the cumulative distribution function of the standard normal distribution.

(4) $$p(k\ \text{Means} - \text{rep}) = \left(2 \cdot \Phi\left(f\sqrt{n}\right) - 1\right)^{2k}.$$

Equation 4 has important implications for 'probability of sample means within $f$ of their corresponding population means' and replication probability. Figure 1, which assumes the sample size per group is constant at 100 and $f$ is constant at .1, shows the number of groups strongly influences both probabilities. However, these probabilities decrease as the number of groups increases from one to four; Panel A of Figure 1 shows the probability all sample means

approximate their corresponding population means decreases, and Panel B of Figure 1 shows the probability of replication also decreases. Placing the panels adjacently depicts the connection between both probabilities, although the latter is always less than the former. This connection provides an additional reason to know empirical facts as precisely as possible.

----- Insert Figure 1 about here ----

Figure 1 is limited because the sample size per group rather than the total sample size is constant. Figure 2 remedies that problem by keeping the total sample size constant at 200. Again, note the connection between the probability of sample means near the corresponding population means and replication probability.

----- Insert Figure 2 about here ----

Figures 1 and 2 are limited because they fail to depict what occurs to the two probabilities when total sample size $N$ and precision $f$ vary. In contrast, Figure 3 allows $N$ to vary along the horizontal axis, with separate curves representing increasing precision demands (.4, .3, .2, and .1) under a constant four groups. Figure 3 shows that increasing $N$ influences both probabilities and $f$ decreases (i.e., more precision is demanded) as either probability decreases. Again, note the close relationship between the two probabilities by comparing Panels A and B.

----- Insert Figure 3 about here ----

In summary, sample size, precision, number of groups, probability of meeting specifications, and replication probability are interrelated. Thus, knowing empirical facts as precisely as possible is critical to scientific advancement and magnified by the relatedness to replication probability.

## 3. Two Possible Counterarguments Against a Precision Focus

One counterargument consistent with conventional wisdom is 'the focus for consumer research is hypothesis testing rather than precise empirical knowledge'. Although hypothesis testing is essential, null hypothesis significance testing (NHST) is unsound. Because a comprehensive discussion of this claim would require its own lengthy article, instead consider the recent special issue of *The American Statistician* (*TAS*), entitled "Statistical Inference in the 21st Century: A World Beyond $p < 0.05$." This 43-article special issue included the following official statement by the American Statistical Association (ASA), which updated their previous statement about NHST.

> The ASA Statement on P-Values and Statistical Significance stopped just short of recommending that declarations of "statistical significance" be abandoned. We take that step here. We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term "statistically significant" entirely. Nor should variants such as "significantly different," "$p < 0.05$," and "nonsignificant" survive, whether expressed in words, by asterisks in a table, or in some other way (Wasserstein, Schirm, & Lazar, 2019, p. 1).

The argument that consumer researchers should stress hypothesis testing rather than precise empirical knowledge assumes that NHST is sound. However, the ASA statement, which reflects the judgment of the world's top statisticians, renders this assumption false. Any consumer researcher insisting on NHST is ignoring the state-of-the-art in statistical science.

In deciding whether to believe a hypothesis, inferential statistics cannot substitute for relevant topic expertise. Although such statistics can provide insights, valid conclusions about

hypotheses require expert judgment grounded in extensive substantive knowledge and precisely known empirical facts. To dramatize the latter's importance, imagine a scenario involving Laplace's omniscient demon. The demon claims that a set of sample means is unrelated to their corresponding population means. Because researchers should only trust data when the sample means provide precise estimates of corresponding population means, untrustworthy data cannot support or disconfirm a hypothesis. Alternatively, if the sample and corresponding population means are unrelated, then it is doubtful the findings are replicable. In turn, irreproducibility is problematic for researchers who rely on the sample means to accept or reject a hypothesis. Thus, precise empirical facts are a prerequisite for hypothesis testing.

Consider the following analogy. A researcher wants to estimate the height difference between two types of dogs: Great Danes and Corgis. Due to low sampling precision, the mean height of the Great Dane sample is highly inaccurate, so the researcher determines that the mean height of Great Danes is 1.5 feet. Now assume low sampling precision for the Corgis caused the researcher to determine their mean height is 2 feet. A statistical test comparing these two inaccurate means would indicate that Corgis are taller than Great Danes. Although an erroneous assessment of one population mean is unfortunate, the situation worsens when a mean comparison is involved because an erroneous conclusion about the mean difference can occur.

Under the typical assumption of a normal distribution and equal group sizes, the precision formula for a mean difference is identical to the formula for mean precision if the standard deviation is known (Trafimow, Wang, & Wang, 2020).[3] The only caveat: there are two groups with $n$ participants when applying Equation 2 to mean differences, whereas there is only one group with $n$ participants when applying it to single means. Focusing on mean differences rather than means themselves indicates slightly more pessimism. To depict this point, Figure 4 shows

how $f$ is a more favorable (lower) value for means than for mean differences at total sample sizes $N$ ranging from 10 to 500 and confidence constant at the standard 95% level. Hence, precision is more critical when it comes to testing the hypotheses involving mean differences, as opposed to single mean estimates.

----- Insert Figure 4 about here -----

Furthermore, good research entails more than mere hypothesis testing.[4] Focusing on mean differences is overly restrictive when there are issues associated with creating better hypotheses, creating better theories from which to derive better hypotheses, abduction to the best explanation, generalizability, and the like. One disadvantage of the hypothesis-testing craze, with concomitant NHST, is 'dichotomania', which is a tendency to think in a binary way (Greenland, 2017). 'Doing science' is more than binary decision-making about whether to accept or reject hypotheses, which is why philosophers and statisticians insist that researchers scrutinize all their data rather than perform simple NHST related to mean differences (Trafimow, 2019a).

A second counterargument against a precision focus involves power analysis. Consumer researchers are accustomed to performing power analyses to determine the sample sizes needed to meet a criterion probability for rejecting null hypotheses, assuming a particular population effect size. However, the ASA's repudiation of NHST suggests this status quo approach no longer pertains. Furthermore, power analysis, as currently performed, often causes imprecision. To accept this counterintuitive statement, consider how researchers typically perform power analyses. In most cases, researchers assume a medium-sized effect (e.g., Cohen's $d$ = .50) and perform a power analysis to find the number of participants needed for an 80% probability that NHST will result in rejecting the null hypothesis at an alpha equals .05 level; for example, 31

participants are required for the simplest case of a single mean. However, solving Equation 1 for precision with $n$ = 31 implies that $f$ = .35, which is unimpressive.

In summary, there are two strong reasons to emphasize precision over power. First, the ASA statement repudiating NHST implies that power analysis helps researchers to perform a procedure they should not perform. Second, power analysis often forces researchers into a lack of precision, which implies irreproducibility.

## 4. Method

*Journal of Consumer Research* (JCR: 2019 impact factor = 4.7; SJR ranking #4 in Marketing), *Journal of Consumer Psychology* (JCP: 2019 impact factor = 2.8; SJR ranking #13 in Marketing), *Journal of Consumer Marketing* (JCM; 2019 impact factor = 1.71; SJR ranking #59 in Marketing) and *International Journal of Consumer Studies* (IJCS: 2019 impact factor = 1.5; SJR ranking #61 in Marketing) were selected for analysis. Historically, these journals have focused on multidisciplinary consumer studies. Using Microsoft Excel's random number generator, twelve articles from each journal volume from 2000 to 2016 were selected. If a chosen article did not describe a quantitative analysis of means (e.g., a conceptual or qualitative analysis article), the immediately following or preceding (if no following available) article in the issue not previously included in the sample replaced it. This selection process yielded a sample of 204 articles from JCR, 190 articles from JCP, 143 articles from JCM, and 132 articles from IJCS.[5] Fewer articles published in JCM and IJCS involved means analysis, resulting in a relatively smaller sample (i.e., ultimately, a census of studies reporting mean analysis for the years 2000 through 2016). Sampled JCR and JCP articles predominantly reported results of field and laboratory experiments; in contrast, most JCM and IJCS articles summarized survey-based studies.

For each article, bibliographic information and the total number of empirical studies were recorded. For multi-study articles, every study with a means analysis was included. In total, 1386 studies from four journals were analyzed, with 590 studies from JCR, 466 studies from JCP, 192 studies from JCM, and 138 studies from IJCS. For each study, sampling precision calculations based on Equation (3) were performed, with a confidence level of 95% for all calculations. Also, replication probability at $f$ =.1, $f$ =.2, and $f$ =.3 was computed with Equation (4) for the first study in each of the 669 selected articles.

## 5. Results

The main finding is average proximity to the population mean (precision), expressed in fractions of standard deviation, is .51 (*SD* = .13) for JCR articles, .49 (*SD* = .14) for JCP articles, .29 (*SD* = .14) for JCM articles, and .26 (*SD* = .13) for IJCS articles. JCM (Cohen's *d* = 1.64) and IJCS (Cohen's *d* = 1.93) articles have superior precision compared to JCR articles, as illustrated by the smaller *f*-value (distance from the mean). Also, JCM (Cohen's *d* = 1.43) and IJCS (Cohen's *d* = 1.70) articles outperform JCP articles. However, those .29 and .26 figures for JCM and IJCS, respectively, are underwhelming in absolute terms.

Given the sampling precision is unimpressive in JCM and IJCS, and worse yet in JCR and JCP, is there reason to believe it is improving? The evidence is mixed. For JCR and JCP, the correlation between publication year and average proximity to the mean of studies published during that year is -.86 and -.71, respectively, which suggests strongly trending improvement. That is, sampling precision increases as years progress, which is exemplified by a low $f$ value (smaller distance from the mean). In IJCS, there is a weak trend for improvement, manifested as a correlation of -.34. In contrast, the correlation of .53 for JCM is substantially in the opposite direction.

Another longitudinal issue of interest is the proliferation of empirical studies per journal article. The correlations between the year and the average number of studies per article are .90 for JCP, .84 for JCR, .64 for JCM, and .18 for IJCS. Given limited financial support for scholarly data collection, this finding begs the question of whether encouraging more empirical studies per article constrains researchers to query fewer respondents per study, thereby decreasing sampling precision. In essence, do publication pressures force researchers into a Hobson's choice whereby they must boost the number of empirical studies per submitted manuscript at the expense of precision per study? Again, the evidence is mixed. The correlation is -.34 for JCR, which contrary to tradeoff pressures suggests articles presenting more empirical studies tend to achieve higher precision (smaller distance from the mean) per study. For JCP, the correlation is -.16, which suggests a weak trend in the same direction. In contrast, the correlation is .31 for JCM and .31 for IJCS, which suggests manuscripts submitted by lesser-funded scholars—most likely employed by less research-oriented universities with less stringent publication requirements—are competitively disadvantaged because the tradeoff pressures are insurmountable.

Furthermore, the serial position within the overall sequence of studies presented in an article is an issue. Primacy effects suggest the standard double-blind review process encourages researchers to conduct more precise initial empirical studies and less precise subsequent empirical studies for their multi-study manuscripts. For instance, studies described later in the sequence often include results of tangential tests "enforced by reviewer comments" (Fiedler & Prager, 2018, p.121). If researchers add empirical studies during multiple revise-and-resubmit iterations, perhaps they are conducting minimally acceptable empirical studies meant merely to appease reviewers. Thus, rational researchers may conserve limited data collection funding with lower precision studies that mollify reviewers. Although data from IJCS and JCM strongly

support this supposition (correlation of .88 and .69, respectively), data from JCR fail to support it ($r$ = -.26), and data from JCP show a strong trend in the opposite direction ($r$ = -.87).

Finally, the Table presents replication probability at different levels of precision ($f$ = .1, .2, and .3) for the four journals. Unsurprisingly, lowering the precision requirement increases the probability of a successful study replication for all journals. Notably, articles from JCM and IJCS, on average, outperform articles from JCR and JCP at each precision level.

----- Insert Table about here ----

## 6. Discussion

Precision was poor in general, although roughly twice as poor in JCR and JCP as in JCM and IJCS. The reasons the lower impact factor journals (JCM and IJCS) outperformed the substantially higher impact factor journals (JCR and JCP) are unclear. Study design specifics—for example, requiring an in-person manipulation (e.g., taste) or direct observation of consumer behavior—can impose limitations on sample size due to limited research funds and time. Additional investigation revealed that only 5% of JCM and IJCS articles relied on these study design features, yet 21% of JCP and 22% of JCR articles described direct behavioral observation or in-person manipulation.[6]

Computed across journals, the unweighted average precision is .39, and the weighted average precision is .45, which fall into the range Trafimow (2018b) termed 'poor precision'.[7] Given this adverse finding, it would be encouraging to report a trend towards higher sampling precision; however, the evidence is mixed, with substantial improvement in JCR and JCP, weaker improvement in IJCS, and substantial worsening in JCM. One supposition about these opposing trends is 'sample size recommendations differ for the quantitative methods used in studies typifying each journal' (e.g., McQuitty, 2004). However, the vast majority of sampled

JCR and JCP articles reported ANOVA/ANCOVA results (90% and 89%, respectively). ANOVA (58% and 59%) and t-test (32% and 31%) results dominated sampled JCM and IJCS articles. Because the data analysis methods used in articles sampled from the four journals are similar, SEM sample size guidelines are not the culprit.

Another possibility could be the accessibility of the sampled population. Perhaps the studies that exhibit lower levels of precision tend to investigate populations that are difficult to reach, and as a result, only a small sample is obtainable. However, 78% and 86% of studies in JCR and JCP, respectively, used a student sample; in contrast, 65% of studies in JCM and IJCS relied on a sample of the general population obtained via a survey or an online panel.[6]

Instead, the explanation may reside within the sociology of science. The differential promotion-and-tenure-related 'value' of an article in a first-tier versus second-tier journal has been diverging (Seggie & Griffith, 2009; Zamudio, Wang, & Haruvy, 2013). Thus, the objective function of most JCR and JCP submitters will strongly discourage them from skimping on sample size, as doing so would reduce publication likelihood. Alternatively, the 'value' of a JCM or IJCS article might be insufficient to warrant substantial investments of time and money, so the objective function of most JCM and IJCS submitters may only mildly discourage or even encourage them to skimp on their sample(s). In essence, researchers' expected value calculations may influence the statistical precision of their research.

The reduction in sampling precision for subsequent studies in JCM and IJCS but not JCR and JCP also may have sociology of science roots. Perhaps, as suggested previously, JCR and JCP submitters generally are better funded than JCM and IJCS submitters. As a result, JCM and IJCS submitters may quickly exhaust their project budget. In contrast, JCR and JCP submitters

are more likely to convince institutional gatekeepers to approve research budget increases for subsequent studies, thus eliminating their need to skimp.

One way to increase replication probabilities is to insist on greater sampling precision, which necessitates larger samples. As Figures 1 through 3 show in different ways, impressive replication probabilities are impossible sans impressive precision, unless researchers 'define away the problem' by using high a priori values for $f$. Larger sample sizes than those typically reported in JCR, JCP, IJCS, or JCM are mandatory for impressive precision and replication probabilities. However, if limited resources constrain many consumer researchers, so insisting on larger samples forces them to perform fewer studies, then a tradeoff between sample size per study and the number of studies is likely. Data from JCM and IJCS mildly support this tradeoff, whereas data from JCR mildly disconfirm it.

When inferential statistical thinking pertains, researchers should specify the population of interest or argue why generalizability is unimportant. The latter typically entails claims about attempting to generalize a theory instead of finding. Although such arguments seem reasonable, consider this fanciful example. Imagine researchers successfully tested a human-based consumer behavior theory on beings from an exoplanet. Although we could marvel that the theory pertains to alien lifeforms, we still might question its generalizability to untested human subpopulations. As most consumer researchers have limited their studies to WEIRD (i.e., Western, educated, industrialized, rich, and democratic) populations, the universality of consumer theories is suspect. Hence, researchers should specify the population of interest or admit it is implicit, even when conducting theory-testing research. When a population is implicit, researchers should explain why their theory should apply to other populations.

## 7. Conclusion

Any comprehensive history of science will indicate the importance of establishing relevant empirical facts. Such facts provide a foundation for theorizing and matter for verifying theories (Hempel, 1965), falsifying theories (Popper, 1959), and abduction (Haig, 2014). Regardless of a researcher's philosophical perspective, contact with empirical facts is a scientific necessity. As sampling precision and replicability are intimately related (Figures 1-3), replicability is an additional reason for sampling precision's importance, although there are additional benefits to obtaining precise empirical facts.

Marketing researchers generally and consumer researchers specifically need empirical facts to be as precise as possible and replicable. Hence, the lack of sampling precision in four well-respected consumer research journals is troubling. Moreover, evidence of improvement on this issue is mixed rather than unambiguous, thereby indicating greater concern about obtaining precise data is warranted. The link between sampling precision and replication is essential because more replicable research is preferred, especially when the replication rate is low (i.e., conducting more replicable studies partly precludes the need for replication research).

In an ideal academic world, less precise articles would be discounted, and thus cited less frequently and in less prestigious journals (hence affirming these journals' relative scholarly status). Unfortunately, two sociology-of-science-related realities dictate otherwise. First, full or partial inter-study replications in marketing remain a rarity despite repeated calls for them.[8] Concern about precision in general and sampling precision in particular derives mainly from the ongoing dearth of such studies. Second, the proliferation of academic publications means scholars must cope with increasing information overload, which they accomplish in part by relaxing their scrutiny of each publication they read. As a result, most scholars are, at best,

minimally sensitive to sampling precision. Instead, article citation incidence is more sensitive to factors such as journal reputation, author reputation, and the congruence between reported and newly discovered findings. Thus, academicians cannot rely on citation incidence to separate the empirical wheat from the empirical chaff.

Although many articles present different APP-related equations (see Trafimow, 2019b for a review), and the lead author has created an Excel spreadsheet (available upon request) for making APP calculations, only one article evaluating research in several psychology subdomains includes APP equations to assess research precision (Trafimow & Myüz, 2019). The present study extends consumer research meaningfully. Because psychology and consumer behavior are closely related fields, it is unsurprising that both are subject to problematic imprecision; in fact, the precision levels are similar in the two scholarly domains. Nonetheless, waiting until psychologists decide to address this problem will compromise the quality of current consumer research. Hence, the editors of marketing-related journals should begin requiring that consumer-centric studies rely on larger samples consistent with APP calculations concerning precision. By requiring authors to apply the APP before pre-registering their studies, which methodologists have touted as a way to eliminate problematic practices like p-hacking (Gonzales and Cunningham, 2015), journal editors would boost the trustworthiness of empirical findings.

## References


Anderson, S. F., & Maxwell, S. E. (2017). Addressing the 'replication crisis': Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research*, 52 (3), 305-324. doi: 10.1080/00273171.2017.1289361

Ashraf, R., & Merunka, D. (2017). The use and misuse of student samples: An empirical investigation of European marketing research. *Journal of Consumer Behaviour*, 16 (4), 295-308. doi: 10.1002/cb.1590

Bagozzi, R. P. (1975). Marketing as exchange. *Journal of Marketing*, 39 (4), 32-39. doi: 10.2307/1250593

Basmann, R. L. (1965). On the application of the identifiability test statistic in predictive testing of explanatory economic models. *Econometric Annual of the Indian Economic Journal*, 3, 387-423.

Bass, F. M. (1969). A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research*, 6 (3), 291-300. doi: 10.1177/002224376900600304

Bass, F. M. (1974). The theory of stochastic preference and brand switching. *Journal of Marketing Research*, 11 (1), 1-20. doi: 10.1177/002224377401100101

Beatty, S. E., & Kahle, L. R. (1988). Alternative hierarchies of the attitude-behavior relationship: The impact of brand commitment and habit. *Journal of the Academy of Marketing Science*, 16 (2), 1-10. doi: 10.1007/BF02723310

Bloom, P. N., & Milne, G. R. (1991). Historical perspectives on funding opportunities in consumer research. In R. H. Holman, & M. R. Solomon (Eds.), *NA - Advances in consumer research, volume 18* (pp. 255-261). Provo, UT: Association for Consumer Research.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, 8 (2), 197-207. doi: 10.1086/208856

*Center for Open Science* (2017). Our mission, https://cos.io/. Accessed 25 June 2018.

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics: Estimation, open science, and beyond.* New York, NY: Taylor and Francis Group.

Easley, R. W., Madden, C. S., & Dunn, M. G. (2000). Conducting marketing science: The role of replication in the research process. *Journal of Business Research*, 48 (1), 83-92. doi: 10.1016/S0148-2963(98)00079-4

Ehrenberg, A. S. C. (1988). *Repeat-buying: Facts, theory and applications, 2nd ed.* New York, NY: Oxford University Press.

Ehrenberg, A. S. C., Goodhardt, G. J., & Barwise, T. P. (1990). Double jeopardy revisited. *Journal of Marketing*, 54 (3), 82-91. doi: 10.1177/002224299005400307

Espinosa, J. A., & Ortinau, D. J. (2016). Debunking legendary beliefs about student samples in marketing research. *Journal of Business Research*, 69 (8), 3149-3158. doi: 10.1016/j.jbusres.2015.12.003

Evanschitzky, H., & Armstrong, J. S. (2013). Research with in-built replications: Comment and further suggestions for replication research. *Journal of Business Research*, 66 (9), 1406-1408. doi: 10.1016/j.jbusres.2012.05.006

Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60 (4), 411-415. doi: 10.1016/j.jbusres.2006.12.003

Fiedler, K., & Prager, J. (2018). The regression trap and other pitfalls of replication science—Illustrated by the report of the Open Science Collaboration. *Basic and Applied Social Psychology*, 40 (3), 115-124. doi: 10.1080/01973533.2017.1421953

Fishbein, M. A., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fishbein, M. A., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Fybush, S. (2015). Inconsistent Nielsen data vexes public radio stations and inspires a commercial solution. *Current: News for People in Public Media*, https://current.org/2015/04/inconsistent-nielsen-data-vexes-public-radio-stations-and-inspires-a-commercial-solution/. Accessed 29 June 2018.

Gillies, D. (2000). *Philosophical theories of probability*. London, UK: Taylor and Francis.

Gonzales, J. E., & Cunningham, C. A. (2015). The promise of pre-registration in psychological research: Encouraging a priori research and decreasing publication bias. *American Psychological Association*, https://www.apa.org/science/about/psa/2015/08/pre-registration. Accessed 14 January 2020.

Good, I. J. (1983). *Good thinking: The foundations of probability and its applications*. Minneapolis, MN: University of Minnesota Press.

Greenland, S. (2017). Invited commentary: The need for cognitive science in methodology. *American Journal of Epidemiology*, 186 (6), 639-645. doi: 10.1093/aje/kwx259

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. Cambridge, MA: MIT Press.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York, NY: Free Press.

Hoyle, F. (1957). *Frontiers of astronomy*. New York, NY: Signet Science Library.

Hunter, J. E. (2001). The desperate need for replications. *Journal of Consumer Research*, 28 (1), 149-158. doi: 10.1086/321953

Hyman, M. R. (2004). Revising the structural framework for marketing management. *Journal of Business Research*, 57 (9), 923-932. doi: 10.1016/S0148-2963(02)00502-7

Hyman, M. R., Skipper, R., & Tansey, R. (1991). Two challenges for the three dichotomies model. In T. Childers, et al. (Eds.), *AMA winter educators' conference proceedings* (pp. 417-422). Chicago, IL: American Marketing Association.

Kerr, G., Schultz, D. E., & Lings, I. (2016). 'Someone should do something': Replication and an agenda for collective action. *Journal of Advertising*, 45 (1), 4-12. doi: 10.1080/00913367.2015.1077492

Kotler, P., & Levy, S. J. (1969). Broadening the concept of marketing. *Journal of Marketing*, 33 (1), 10-15. doi: 10.2307/1248740

Kwon, E. S., Shan, Y., Lee, J. S., & Reid, L. N. (2017). Inter-study and intra-study replications in leading marketing journals: A longitudinal analysis. *European Journal of Marketing*, 51 (1), 257-278. doi: 10.1108/EJM-07-2015-0450

Lindsay, R. M., & Ehrenberg, A. S. C. (1993). The design of replicated studies. *The American Statistician*, 47 (3), 217-228.

Lynch, J. G. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9 (3), 225-239. doi: 10.1086/208919

Lynch, J. G. (1983). The role of external validity in theoretical research. *Journal of Consumer Research*, 10 (1), 109-111. doi: 10.1086/208949

Lynch, J. G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27 (3), 367-376. doi: 10.1177/0092070399273007

Malhotra, N. K. (1981). A scale to measure self-concepts, person concepts, and product concepts. *Journal of Marketing Research*, 18 (4), 456-464. doi: 10.1177/002224378101800407

McQuitty, S. (2004). Statistical power and structural equation models in business research. *Journal of Business Research*, 57 (2), 175-183. doi: 10.1016/S0148-2963(01)00301-0

*Open Science Collaboration* (Nosek, B. A., Aarts, A. A., Anderson, C. J., Anderson, J. E., & Kappes, H. B.) (2015). Estimating the reproducibility of psychological science. *Science*, 349 (6251). doi: 10.1126/science.aac4716

O'Shaughnessy, J., & O'Shaughnessy, N. J. (2002). Postmodernism and marketing: Separating the wheat from the chaff. *Journal of Macromarketing*, 22 (1), 109-135. doi: 10.1177/027467022001010

Park, J. Hwan, Venger, O., Park, D. Y., & Reid, L. N. (2015). Replication in advertising research, 1980-2012: A longitudinal analysis of leading advertising journals. *Journal of Current Issues & Research in Advertising*, 36 (2), 115-135. doi: 10.1080/10641734.2015.1023874

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28 (3), 450-461. doi: 10.1086/323732

Petroski, H. (1992). *The evolution of useful things: How everyday artifacts—From forks and pins to paper clips and zippers—Came to be true as they are*. New York, NY: Alfred A. Knopf.

Popper, K. R. (1959). *The logic of scientific discovery*. Oxford, UK: Routledge.

Randall, D., & Welser, C. (2018), The irreproducibility crisis of modern science: Causes, consequences, and the road to reform, *National Association of Scholars*, https://www.nas.org/reports/the-irreproducibility-crisis-of-modern-science/full-report. Accessed 14 January 2020.

Royne, M. B. (2018). Why we need more replication studies to keep empirical knowledge in check: How reliable is truth in advertising research? *Journal of Advertising Research*, 58 (1), 3-7. doi: 10.2501/JAR-2018-007

Seggie, S., & Griffith, D. A. (2009). What does it take to get promoted in marketing academia? Understanding exceptional publication productivity in the leading marketing journals. *Journal of Marketing*, 73 (1), 122-132. doi: 10.1509/jmkg.73.1.122

Shimp, T. A., & Bearden, W. O. (1982). Warranty and other extrinsic cue effects on consumers' risk perceptions. *Journal of Consumer Research*, 9 (1), 38-46. doi: 10.1086/208894

Sniehotta, F. F., Presseau, J., & Araújo-Soares, V. (2014). Time to retire the theory of planned behaviour. *Health Psychology Review*, *8* (1), 1-7. doi: 10.1080/17437199.2013.869710

Stephenson, B. (1987). *Kepler's physical astronomy*. Princeton, NJ: Princeton University Press.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press.

Trafimow, D. (2016). A taxonomy of applied research categories as an aid to research pertaining to aviation. *International Journal of Aviation Sciences*, 1 (1), 1-11.

Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics. *Educational and Psychological Measurement*, 77 (5), 831-854. doi: 10.1177/0013164416667977

Trafimow, D. (2018a). Confidence intervals, precision and confounding. *New Ideas in Psychology*, 50, 48-53. doi: 10.1016/j.newideapsych.2018.04.005

Trafimow, D. (2018b). An a priori solution to the replication crisis. *Philosophical Psychology*, 31 (8), 1188-1214. doi: 10.1080/09515089.2018.1490707

Trafimow, D. (2019a). Five nonobvious changes in editorial practice for editors and reviewers to consider when evaluating submissions in a post $p < .05$ universe. *The American Statistician*, 73 (sup1), 340-345. doi: 10.1080/00031305.2018.1537888

Trafimow, D. (2019b). A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics*, 7 (2), 1-14, https://www.mdpi.com/2225-1146/7/2/26. Accessed 9 November 2019.

Trafimow, D., & MacDonald, J. A. (2017). Performing inferential statistics prior to data collection. *Educational and Psychological Measurement*, 77 (2), 204-219. doi: 10.1177/0013164416659745

Trafimow, D., & Myüz, H. A. (2019). The sampling precision of research in five major areas of psychology. *Behavior Research Methods*, 51, 2039-2058, https://doi.org/10.3758/s13428-018-1173-x.

Trafimow, D., Wang, C., & Wang, T. (2020). Making the a priori procedure (APP) work for differences between means. *Educational and Psychological Measurement*, 80 (1), 186-198. doi: 10.1177/0013164419847509

Trafimow, D., Wang, T., & Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! *Educational and Psychological Measurement*., 79 (1), 129-150. doi: 10.1177/0013164418764801

Toncar, M. F., & Munch, J. M. (2010). Meaningful replication: When is a replication no longer a replication? A rejoinder to Stella and Adam (2008). *Journal of Marketing Theory and Practice*, 18 (1), 71-80. doi: 10.2753/MTP1069-6679180105

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Editorial: Moving to a world beyond "$p < 0.05$." *The American Statistician*, 73 (Supplemental), 1-19. doi: 10.1080/00031305.2019.1583913

Weissbrot, A. (2018, May 17). The podcasting ad opportunity is increasing – And so are its growing pains, http://adexchanger.com/digital-audio-radio/the-podcasting-ad-opportunity-is-increasing-and-so-are-its-growing-pains/. Accessed 28 June 2018.

Zamudio, C., Wang, Y. & Haruvy, E. (2013). Human brands and mutual choices: An investigation of the marketing assistant professor job market. *Journal of the Academy of Marketing Science*, 41 (6), 722-736. doi: 10.1007/s11747-013-0341-x

## Footnotes

1. Scholars could argue that these equations are nitpicky because worthwhile consumer research does not require extreme precision. Regardless, they are crucial to explaining the APP's rationale. Furthermore, precision is critical even when extreme precision is unnecessary.
2. Trafimow (2017) provides a proof of the equation.
3. Equation 1 assumes random sampling from a normally distributed population. Although Trafimow, Wang, and Wang (2019) have expanded the procedure to include skewed distributions, normal distributions are assumed hereafter.
4. Researchers have touted alternatives, such as confidence intervals (Cumming and Calin-Jageman, 2017) and Bayesian procedures (Gillies, 2000; Good, 1983), to the APP.
5. The full list of articles included in the final sample is available upon request.
6. The analysis of Study 1 in each sampled article served as the basis for this analysis.
7. In conjunction with mathematical simulations, Trafimow (2019a) suggests the following categories: 0 to .1 is excellent precision, .11 to .2 is good precision, .21 to .3 is moderate precision, and .31 to .4 is poor precision. However, Trafimow (2019a) stresses that these categories are informal and should be considered in that spirit.
8. Articles that incidentally revise/modify extant questionnaire scales, which do not entail "the reproducibility of the results of a previous study through the execution of a separate study conducted by different researchers in another investigative setting as an independent project," are excluded (Kwon et al., 2017).

**Table**

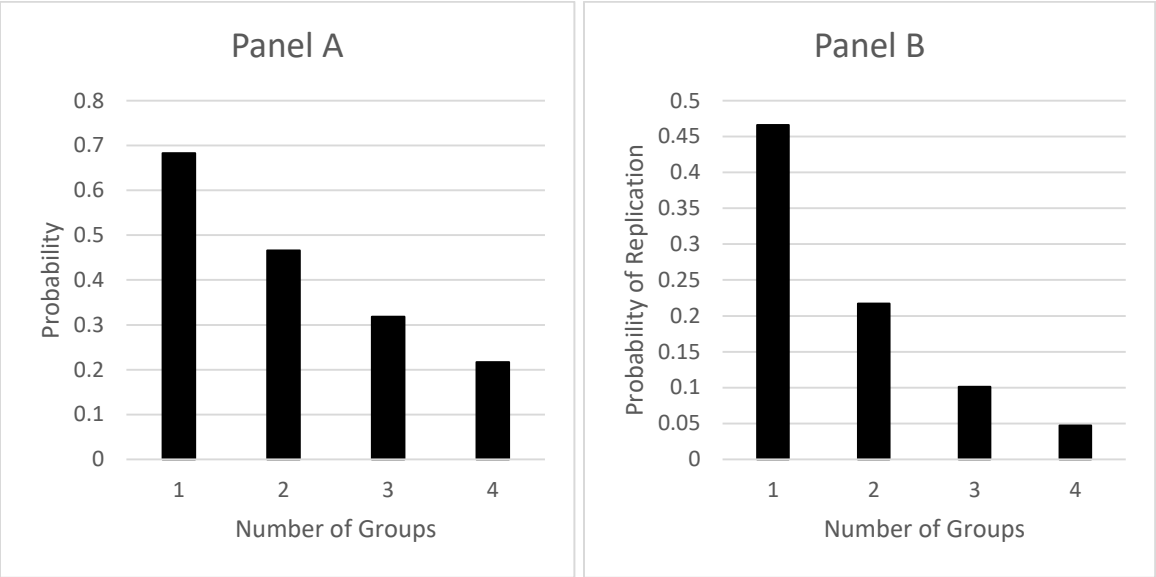
**Study 1: Average Replication Probability for Sampled Articles[a]**

| | ***f* = .1** | ***f* = .2** | ***f* = .3** |
|---|---|---|---|
| **JCR** | 1.6 % | 14.4 % | 39.9 % |
| **JCP** | 3% | 13% | 37% |
| **JCM** | 26.2 % | 60.1 % | 81.2 % |
| **IJCS** | 27% | 67% | 83% |

[a] *f* – desired distance from the population mean expressed as a fraction of standard deviation. Confidence is set at 95%.

**Figure 1**

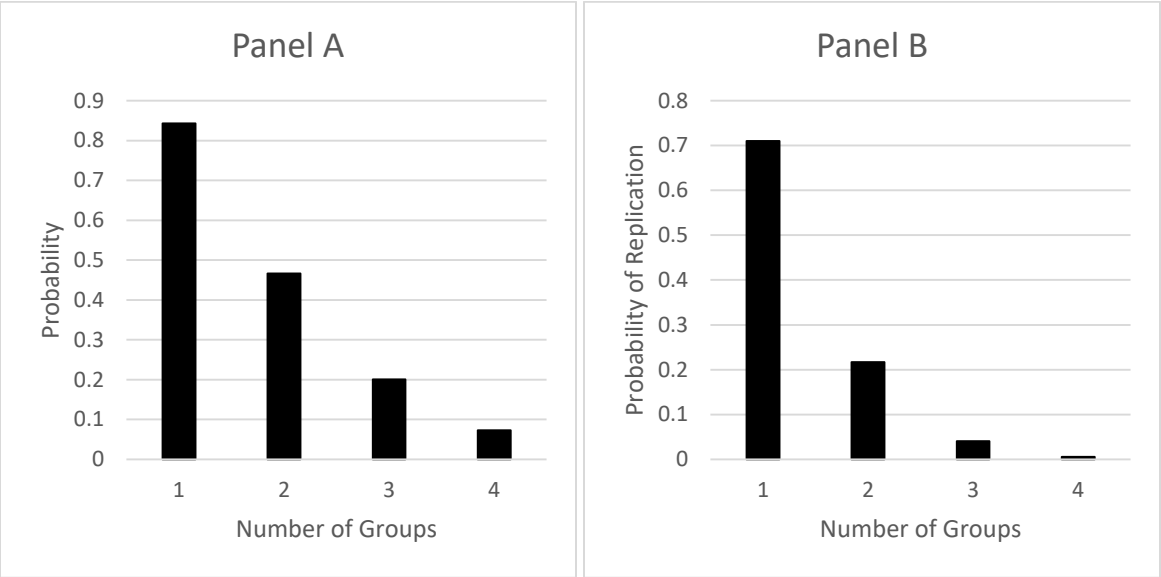
**Fixed Group Size: Effect on Precision and Replication Probability**




*Note*: The probability of obtaining sample means within .1 standard deviations of their corresponding population means (Panel A) and the probability of replication (Panel B), as a function of the number of groups, keeping the sample size per group constant at 100.

**Figure 2**

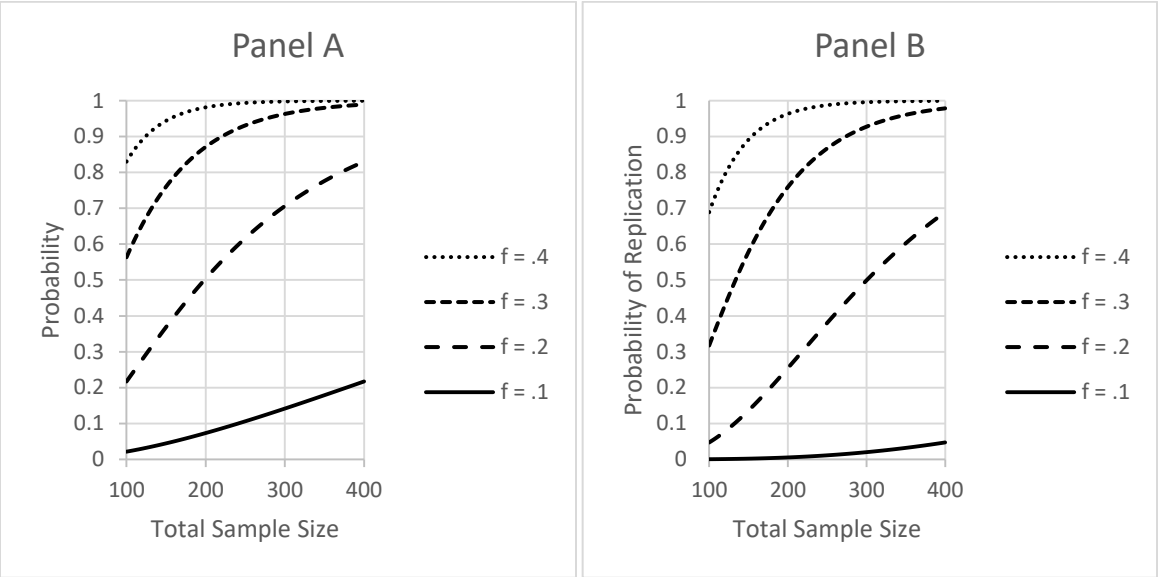
**Fixed Total Sample Size: Effect on Precision and Replication Probability**




*Note:* The probability of obtaining sample means within .1 standard deviations of their corresponding population means (Panel A) and the probability of replication (Panel B), as a function of the number of groups, keeping the total sample size constant at 200.

**Figure 3**

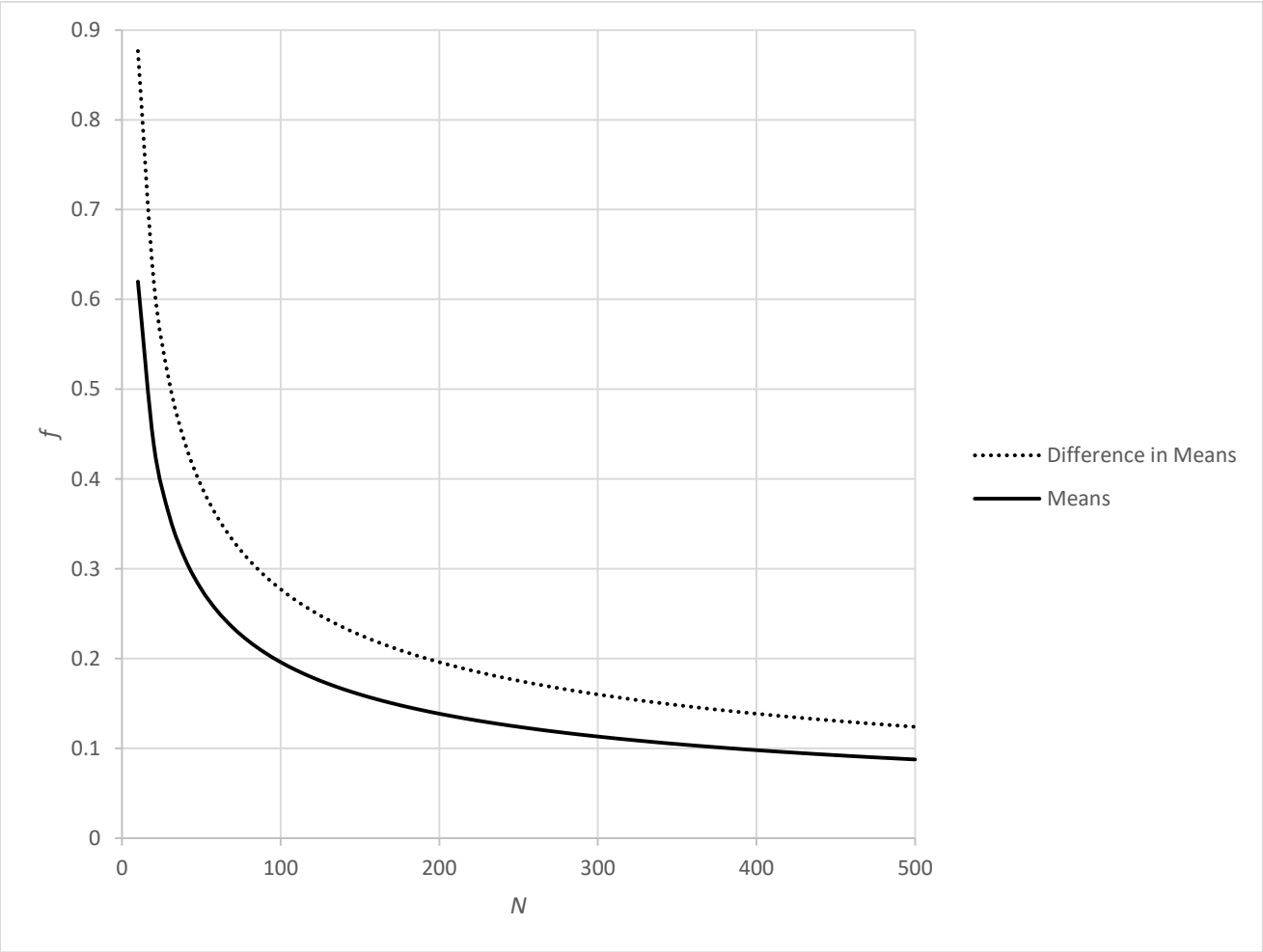
**Varying Total Sample Size: Effect on Precision and Replication Probability**




*Note*: The probability of obtaining sample means within prescribed distances (Panel A) and the probability of replication (Panel B) as a function of the total sample size and the distance criterion (.4, .3, .2, or .1).

**Figure 4**

**Sampling Precision: Means versus Means Differences**




*Note*: Precision $f$ is represented along the vertical axis as a function of total sample size N along the horizontal axis for means (lower curve) and differences in means (upper curve). Confidence is set at 95% throughout.