Journal of Business Research 125 (2021) 39–44

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

The harmful effect of null hypothesis significance testing on marketing research: An example

David Trafimow [a], Michael R. Hyman [b,*], Alena Kostyk [c], Cong Wang [d], Tonghui Wang [d]

[a] Distinguished Achievement Professor of Psychology, New Mexico State University, Department of Psychology, MSC 3452, Box 30001, Las Cruces, NM 88003, United States
[b] Distinguished Achievement Professor of Marketing, New Mexico State University, College of Business, Box 30001, Dept. 5280, Las Cruces, NM 88003-8001, United States
[c] University of Glasgow, Adam Smith Business School, University Avenue, Glasgow G12 8QQ, United Kingdom
[d] New Mexico State University, Department of Mathematics, Las Cruces, NM 88003-8001, United States

ARTICLE INFO

Keywords:
Apriori procedure
Excessive power
Precision
Confidence
Sample size

ABSTRACT

Null hypothesis significance testing (NHST) has had and continues to have an adverse effect on marketing research. The most recent American Statistical Association (ASA) statement recognized NHST's invalidity and thus recommended abandoning it in 2019. Instead of revisiting the ASA's reasoning, this research note focuses on NHST's pernicious peripheral effect on marketing research. One example of this problem is the well-known and influential recommendation against excessive power in McQuitty (2004, 2018). Instead, researchers always should prefer larger sample sizes because they always engender more precision than smaller sample sizes, ceteris paribus.

1. Introduction

In an article published in the *Journal of Business Research* roughly 15 years ago, McQuitty (2004) made recommendations about ideal sample sizes for structural-equation-model (SEM)-based studies. That article's influence has been impressive. In the author's words (2018, p. 273),

> I was pleased to hear from *Journal of Global Scholars of Marketing Science* (JGSMS) Editor-in-Chief Professor Arch Woodside that McQuitty (2004) is "ranked among the top 100 in all-time citation impact among JBR articles."

The quotation implies two notions. First, SEM-proficient marketing researchers value the article. (Collectively, we have cited McQuitty (2004) repeatedly in our SEM-based articles to justify the sample size.) Second, these researchers consider its recommendations sufficiently sound and influential for Arch Woodside (former *Journal of Business Research* editor) to solicit the author for a written commentary about it. Hence, McQuitty (2004) remains a persuasive article for marketing researchers.

McQuitty (2004) discusses the importance of power analysis and argues that researchers should strive for neither 'too little' nor 'too much' power (i.e., a Goldilocks ideal). To quote from the abstract (p. 175),

> Using articles from some leading business journals as examples, a survey finds that power tends to be either very low, implying that too many false models will not be rejected (Type II error), or extremely high, causing overrejection of tenable models (Type I error).

The article subsequently recommends (p. 181),

> If it can be demonstrated that power is too high unless an unreasonably small sample size is employed, it becomes necessary to permit greater latitude in the interpretation of fit statistics that are sensitive to sample size, while giving additional weight to those indices that are less sensitive to sample size. This remedy to the problem of excessive power is probably the most acceptable.

From the perspective that researchers should test their models using null hypothesis significance testing (NHST), this recommendation seems sensible. If the sample size (n) is 'too small', then there is insufficient power to detect an incorrect model; but if the n is 'too large', then every model will be rejected. Thus, the recommendation for an intermediate n—neither 'too small' nor 'too large'—appears sound. However, in light

* Corresponding author.
E-mail addresses: dtrafimo@nmsu.edu (D. Trafimow), mhyman@nmsu.edu (M.R. Hyman), Alena.Kostyk@glasgow.ac.uk (A. Kostyk), Cong960@nmsu.edu (C. Wang), twang@nmsu.edu (T. Wang).

https://doi.org/10.1016/j.jbusres.2020.11.069
Received 11 May 2020; Received in revised form 3 July 2020; Accepted 30 November 2020
Available online 13 December 2020

of the recent debate related to using NHST, we contend this recommendation is no longer valid. Hence, the two-fold goals here are (1) to question researchers' use of NHST, and (2) to argue that a larger n is always better than a smaller n, ceteris paribus. If the argument is valid, then the intermediate n recommendation is unsound, which in turn exemplifies the main point: NHST induces faulty analytical thinking, and thus scholarly marketing journals should abandon it.

2. NHST

The flaws of NHST, such as causing exaggerated effect sizes (Grice 2017; Hyman 2017; Kline 2017; Locascio 2017a, 2017b; Marks 2017), are well-documented (see Hubbard 2016; Ziliak and McCloskey 2016 for highly cited reviews). In response, resistance to discontinuing its use is waning; for example, *Basic and Applied Social Psychology* banned it in 2015 (Trafimow and Marks 2015). The 2019 special issue of *The American Statistician*, which contains 43 critical articles and an editorial statement by the American Statistical Association (ASA) recommending immediate abandonment of NHST, best exemplifies the attitudinal change toward NHST. That statement includes the following crucial paragraph.

> The [2016] 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).

Accordingly, medical journals such as the *New England Journal of Medicine* have changed their statistical policies.

Despite the ASA's recommendation, marketing journals have yet to adjust their statistical policies. If marketing scholars believe that statisticians' criticisms of NHST are esoteric, then rehashing these criticisms is pointless for motivating change. Hence, the goal here is to show how NHST-related thinking fosters poor thinking and deleterious consequences for marketing research. In this vein, the suggestion that SEM-proficient researchers avoid large samples that cause "overrejection of tenable models (Type I errors)" is revisited (McQuitty 2004, p. 175).

3. Sampling precision

It is cliché among statisticians that the larger the sample, under typical assumptions about random and independent sampling, the more it resembles the population. In essence, larger samples make researchers more confident that relevant sample statistics approximate their corresponding population parameters. Ceteris paribus, larger samples imply higher precision.

Invoking Laplace's omniscient Demon can dramatize precision's importance. Suppose the Demon reported that the sample values used to estimate a structural equation model were unrelated to the population values. This discrepancy would discourage scholars from accepting that model, as high confidence in the correspondence between model values and population values is a ubiquitous assumption. In essence, SEM researchers desire *confidence* in the *precision* of the values they calculate.

Recently, Trafimow and his colleagues (e.g., Trafimow, 2017; Trafimow & MacDonald, 2017; Trafimow, Wang, & Wang, 2019; see Trafimow, 2019 for a review) developed the a priori procedure (APP), which is an inferential method that specifies the answers to two questions:

- How closely do researchers want sample statistics and population parameters to correspond? (Precision)
- With what probability do researchers want to fulfill a precision criterion? (Confidence)

For example, assume Dr. Smith specifies the desired levels of precision and confidence before collecting data and then applies an appropriate APP equation to calculate the n for meeting or exceeding those levels. In a simple case, Dr. Smith plans to draw a random and independent sample from a normally distributed population. She wants a sample large enough to meet her precision and confidence specifications for the sample mean. Trafimow (2017) provides a derivation of Equation (1):

$$n = \left(\frac{z_c}{f}\right)^2; \tag{1}$$

where

- n is the sample size required to meet her specifications,
- f is the fraction of a standard deviation she deemed sufficiently precise,
- z_c is the z-score that corresponds to her desired level of confidence (e.g., 1.96 for the conventional 95% confidence level).

Suppose Dr. Smith wants to be 95% confident that a sample mean will be within 0.1 of a standard deviation of the population mean. To calculate the minimum required n, she can use Equation (1) as follows: $n = \left(\frac{1.96}{0.1}\right)^2 = 384.16 \approx 385$ participants. Thus, Dr. Smith must query 385 people to meet her desired confidence and precision levels. Once Smith collects the data, she computes the descriptive statistics of interest. Dr. Smith requires nothing additional, as she a priori designed her study to meet targeted precision and confidence levels.

To show the importance of sample size for precision, suppose instead that $n = 25$. In that case, the precision is only 0.4, which differs dramatically from the case of $n = 385$ and a resulting precision is 0.1.

Because researchers often create more sophisticated study designs, Trafimow and his colleagues created new APP equations. For example, they have published equations that can handle

- k means (Trafimow and MacDonald 2017);
- differences between means for matched or independent samples (Trafimow et al., 2019);
- skew-normal, instead of more limited normal, distributions (Trafimow et al., 2019; Wang, Wang, Trafimow, & Chen, 2019);
- standard deviations (assuming normal distributions) and scales (assuming skew-normal distributions) (Wang, Wang, Trafimow, and Zhang 2019); and
- estimating skewness (Wang, Wang, Trafimow, and Myüz 2019).

Although there are no published APP equations for SEM analyses, there are APP equations for correlations, which under the typical bivariate normality assumption is sufficient here. In simple models, with a dependent variable that is a function of a single independent variable, path coefficients are identical to correlation coefficients. In complex models, correlation coefficients are readily convertible into path coefficients. For example, the path coefficients for a mediation model that involves three variables, where variable A is exogenous, variables B and C are endogenous, and paths AB, AC, and BC exist, are related to the correlation coefficients as follows:

$$\begin{cases} AB = r_{AB} \\ AC = (r_{AC} - r_{BC}{}^{*}r_{AB})/(1 - r_{AB}^2) \\ BC = (r_{BC} - r_{AC}{}^{*}r_{AB})/(1 - r_{AB}^2) \end{cases} \tag{2}$$

Now consider the relationship between the APP and power analysis. Although the APP and power analysis seem similar, as researchers could

use either one to determine n, they differ philosophically and mathematically (Trafimow & Myüz, 2019). Philosophically, researchers use the APP to determine the n required to meet specified precision and confidence levels; in contrast, they use power analysis to determine the n needed for a good chance (e.g., 80%) of obtaining a statistically significant p-value. Mathematically, the desired precision level influences the APP but not power analysis, and the expected (or required) effect size influences power analysis but not the APP. If researchers abandon NHST, the power analysis is futile because its raison d'etre is to facilitate NHST.

For example, suppose Dr. Smith wants to calculate a single mean from normally distributed data and to detect a medium effect size (Cohen's $d = 0.5$) with 80% power at alpha $= 0.05$. A power analysis using these conventional values recommends $n = 31$. However, solving Equation (1) for precision with $n = 31$ implies an unimpressive $f = 0.35$ (i.e., a 95% probability that the sample mean is within 0.35 standard deviations of the population mean). In contrast, precision at the more impressive $f = 0.10$ at 95% confidence would require 385 participants. Or for a typical two-condition experiment with random assignment of participants to either an experimental or control group, a power analysis indicates that 63 participants per group ($n = 126$) would meet conventional specifications. Yet, precision is again an unimpressive $f = 0.35$, whereas 770 participants per condition ($n = 1540$) are needed to achieve precision at the more impressive level of $f = 0.10$ (see the calculator at https://app-normal.shinyapps.io/N_TwoSamples_EstimateMean/ and Li et al., 2020 for details).

Now imagine a simple correlation case. Assume Dr. Smith believes that latent variables X and Y are canonically correlated, and their indicator variables are normally distributed. The relevant coefficient is the correlation between X and Y, which r_{XY} symbolizes at the sample level and ρ_{XY} symbolizes at the population level. The goal: r_{XY} should be a sufficiently precise estimate of ρ_{XY}. The larger the sample size, the higher the sampling precision (see Fig. 1). Consider these three effects.

- Regardless of the precision level, the required n is larger when researchers desire greater confidence (e.g., 95% versus 90%).
- The required n decreases (increases) as the precision level decreases (increases). Increasing n increases precision level, and increased precision level ensures that r_{XY} provides a good estimate of ρ_{XY}.
- Precision level and confidence interact. That is, the required n is very large at stringent precision and confidence levels, but becomes smaller quickly and in a non-additive manner as either specification becomes less stringent.

For example, the difference in the necessary n between $f = 0.1$ and $f = 0.2$ is larger for a confidence level of 95% rather than 90%. If Dr. Smith insists on what Trafimow (2018) termed 'excellent' precision at the 95% confidence level, she needs an n of 802 despite her model's simplicity. In contrast, a power analysis for a simple correlation coefficient implies an n of only 29 people (see https://www.masc.org.au/stats/PowerCalculator/PowerCorrelation for the power calculator used here). If a researcher takes the scenario involving Laplace's Demon seriously and wants the sample correlation coefficient to be a precise estimator of the corresponding population parameter, the power-analysis-recommended n is woefully small. To the present point, 802 participants would qualify as representing 'excessive power' according to McQuitty (2004, p. 181), and yet Fig. 1 renders visible the precision advantage.

4. Implications

As the opening quote from McQuitty (2018) states, overestimating the influence of McQuitty (2004) would be difficult. To justify a modest n, many researchers have cited the admonition in McQuitty (2004) about excessive effect sizes. Nevertheless, whether this warning has benefitted or harmed subsequent marketing research is unknown.

To answer this question, imagine Dr. Smith can collect data from human subjects without cost or effort. Given that precision is essential, and a larger n implies greater precision, she ought to draw a large sample. Nonetheless, if she follows the sample size advice in McQuitty (2004) to avoid "overrejection of tenable models" (McQuitty 2004, p. 175), she compromises her study's precision. In discussing MacCallum, Browne, and Sugawara (1996), McQuitty (2004, p.181) acknowledges this issue.

> An alternative approach to the problem of excessive power and small sample sizes is offered by MacCallum et al. (1996), who observe that "…whereas a moderate N might be adequate for achieving a specified level of power for a test of overall fit, the same level of N may not necessarily be adequate for obtaining precise parameter estimates.'' (p. 144).

MacCallum, Browne, and Sugawara (1996) suggest that Dr. Smith

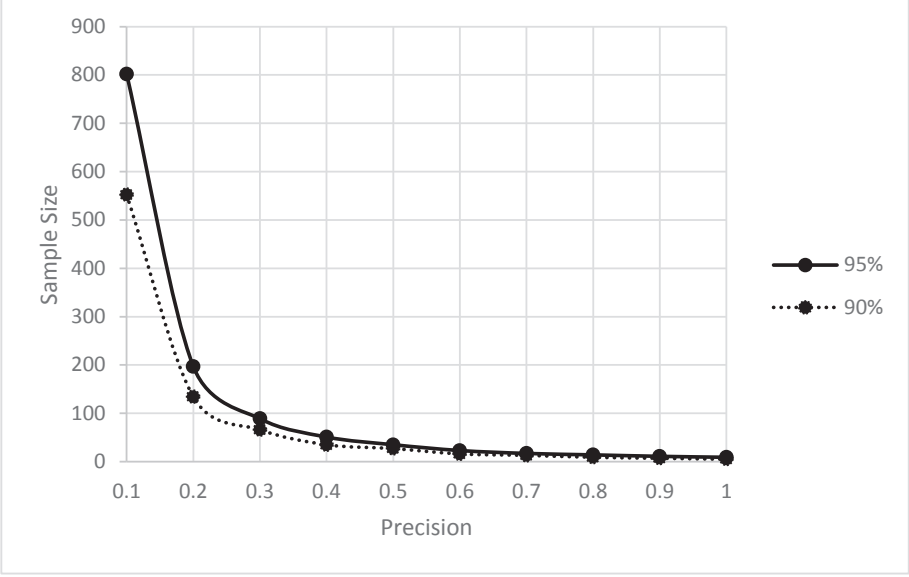

Fig. 1. Sample Size versus Sampling Precision. The n necessary to meet the specifications ranges along the vertical axis as a function of the desired precision along the horizontal axis and on the desired confidence (95% and 90%).

should first estimate and then test her model. If her initial estimation effort goes well, she should re-estimate her model but vary the n to achieve the desired power level. Although the input (covariance or correlation) matrix is the same in both cases, she should use a predetermined n to assess model fit during the second stage.

Ponder this recommendation. Although parameter estimates should rely on all collected data, researchers should assess model fit with partial data only! This strange yet seemingly plausible recommendation illustrates how NHST yields lousy research advice, with a profound negative effect on knowledge development. Given the often non-negligible cost and effort related to the data collection and the scientific duty to present the most accurate evidence, disregarding portions of collected data when assessing a model seems ill-advised.

5. Considering all assumptions, not only the null hypothesis

Now consider the following fact about NHST that substantive researchers often overlook. Specifically, NHST does not result in rejecting only the null hypothesis, but rather rejecting a broad set of auxiliary assumptions containing the null hypothesis. The set contains so many assumptions related to significance testing that some researchers have proposed that assumption taxonomies are needed (e.g., Bradley and Brand 2016; Trafimow 2019). Although such taxonomies are not applicable here, they suggest the following: it is unlikely that all the assumptions within a substantial set of assumptions will be true. For example, the present authors are unaware of a published social science study in which the underlying assumption of random and independent sampling was met (Berk and Freedman 2003). Similarly, Likert-type data used in much marketing research is known not to be continuous, nor to be distributed normally, despite the assumption of indicator variable continuity or normalcy in SEM (Bentler and Chou 1987; Westland 2010). Thus, one problem with NHST is that researchers use it to test a set of assumptions that is almost certainly wrong.

Is it sensible to avoid collecting large samples to avoid overrejecting tenable models? What does McQuitty (2004) mean by 'tenable'? McQuitty (2004, 2018) suggest a reasonable synonym is 'correct'. However, the many substantive and statistical assumptions that researchers make when testing their models suggest that at least one of their assumptions is false (i.e., the whole model cannot be correct). If all models are at least partially wrong, then they cannot be overrejected.

6. Box and Draper quotation

A counter-argument could focus on replacing the word 'tenable' (i.e., correct) with a phrase like 'good enough', 'not too bad', or 'useful.' In this context, a famous quotation by Box and Draper (1987) is worth considering: "Essentially, all models are wrong, but some are useful" (p. 424). Perhaps McQuitty (2004) should have argued that a large n can cause overrejection of useful models.

However, this argument also is problematic, as low p-values provide disconfirming evidence for a null model, but high p-values do not provide confirming evidence. For example, suppose the p-value is a very high 0.99. Its correct interpretation is 'strong evidence against the model is lacking' rather than 'strong evidence for the model exists'. In essence, a valid conclusion is precluded rather than the model is correct or useful. Because many excellent researchers have misunderstood this point, Greenland (2017) recommended converting p-values to an index of contrary information about a model. Based on the information theory developed in Shannon (1948), Equation (3) allows researchers to enact Greenland's recommendation.

$$bits\ of\ information\ against\ the\ null\ model = -log_2(p) \tag{3}$$

Fig. 2 shows the implications of transforming p-values into bits of information against the null model. When the p-value is low, such as 0.001, it represents 9.57 bits of information against the null model. When the p-value is at the conventional significance level (0.05), it represents 4.32 bits of information against the null model. Finally, when the p-value is the largest possible (1.00), it represents zero bits of information against the null model. Zero bits of information against the null model is not translatable validly into a positive statement about model correctness, sufficiency, or usefulness.

Typical fit indexes, based on p-values or statistics that can yield p-values, often perform poorly and lead to nonsensical results. Imagine another Demon scenario in which the Demon claims that the population mean in an experiment is 10.00, and the population standard deviation is 8.00. Researcher A queries ten people and obtains a mean of 9.93, whereas Researcher B queries 1,000,000 people and obtains a mean of 9.95. Researcher B's sample mean provides a better population mean estimate than Researcher A's sample mean. However, the calculation of a p-value, a t-value, or any other test statistic would suggest the opposite.

Like skilled writers, excellent researchers know when to flout convention. For example, researchers typically assume the

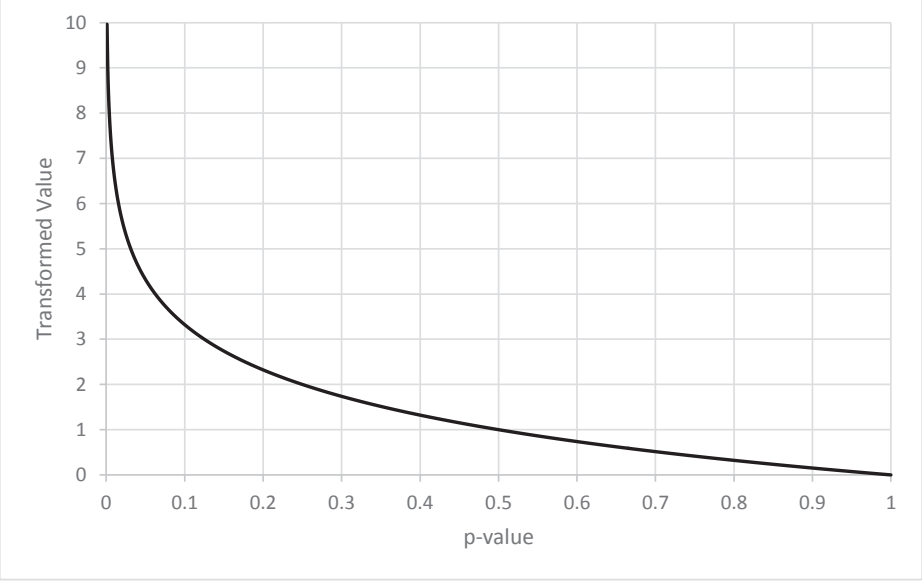

Fig. 2. Transforming p-values into Bits of Information versus the Null Model. The bits of information against the null model range along the vertical axis as a function of the p-value ranging across the horizontal axis.

trustworthiness of fit indexes based on statistics that can yield *p*-values. However, such statistics become suspect when they depend on a superabundance of assumptions that practically guarantee at least one is incorrect, thus rendering the model 'incorrect' (i.e., provides evidence that can refute but not support the complete model).

How can Dr. Smith determine if the data fit her model well? She could base the model on strong prior theorizing. Consider again the theory that *X* causes *Y*, which then yields an empirical hypothesis that there is a canonical correlation between *X* and *Y*, both operationalized as a set of indicator variables. Although simple, this theory is vague. How much should *X* correlate with *Y*? Assume a more robust theory from which Dr. Smith could predict that the correlation between *X* and *Y* is 0.40. In the latter case, it is easier to determine model fit. To the extent that the sample correlation is close to 0.40, the fit is good, whereas to the extent that the sample correlation is far from 0.40, the fit is poor. Note that more study participants are preferred because a larger *n* increases the precision with which the sample correlation coefficient estimates the population correlation coefficient. When NHST-induced thinking does not cloud the model-fitting process, it becomes much more straightforward, and there is no need to excise some data to determine model fit. Dr. Smith should use all collected data to determine the sample correlation coefficient as an estimate of the population correlation coefficient, and to determine the model fit. The more similar the sample correlation coefficient is to the population correlation coefficient, the better the data test the theoretical prediction. If the sample correlation coefficient is far from its theory-suggested value, then Dr. Smith should revisit either the theory or her derived model if the *n* is sufficient to engender good precision. Dr. Smith also might collect additional data to test if the sample correlation coefficient is reliable.

In summary, the Box and Draper (1987) quotation stands, but not within an NHST context. Nor did Box and Draper argue that it should. However, marketing researchers are so accustomed to NHST that they fail to recognize that the statistics used to perform NHST and associated thinking are problematic, and might offer a false sense of providing useful insights (see Trafimow, Hyman, and Kostyk, 2020).

7. Is NHST never useful in SEM contexts?

We thank an anonymous reviewer for suggesting the possibility that NHST may be appropriate for simple SEM models, with assumptions more likely to be correct, than for complex models, with assumptions less likely to be correct. Although a reasonable possibility, the entire set of assumptions for a simple SEM model is partly incorrect. For example, the ubiquitous assumption of random and independent sampling is incorrect because all random selection procedures are imperfect. If even one assumption in the statistical model is partly incorrect, it follows that the statistical model is wrong. In turn, if the statistical model is wrong, then performing NHST to reject it is redundant.

Relative to simple models, more complex models rely on numerically more problematic assumptions, ceteris paribus. Because all models lack complete rightness, they are wrong a priori and not due to significance test results. NHST cannot indicate a model's proximity to truth, which is a matter for expert judgment.

8. The big picture

Although the latest ASA statement has repudiated NHST, and respected scholarly journals are beginning to desk-reject manuscripts that include it, the problem extends beyond using an invalid procedure for testing models. NHST contaminates statistical thought by encouraging dichotomous thinking: a finding either exists or does not exist based on NHST results (Greenland 2017). Of course, the finding always exists in the sense that the effect size never equals zero. Thus, the issue is whether the *n* is sufficient for obtaining a satisfactory estimate of the population effect size. The larger the *n*, the more the obtained effect size is trustworthy as an estimate of the population effect size. Ceteris paribus, a larger *n* is better. Unfortunately, this commonsensical conclusion becomes questionable after researchers accept a flawed procedure such as NHST and its associated 'lousy thinking'.

The argument presented here—that a logically valid conclusion can derive from a false premise—is not limited to McQuitty (2004, 2018). Any conclusion derived from an argument with a false premise or conflicting premises is untrustworthy (Skipper and Hyman 1987). If NHST were sound, then the recommendation in McQuitty (2004) would be logical and valid. Unfortunately, NHST is problematic, thereby rendering this recommendation unsound despite its logical validity.

In closing, consider this waggish argument.

- Major premise: If the moon has gravity, then all people have ten arms.
- Minor premise: The moon has gravity.
- Conclusion: All people have ten arms.

Despite human experience indicating otherwise, the conclusion follows logically from the major premise. Similarly, NHST represents a false major premise. Empirical social science research is only as trustworthy as its weakest aspect (Hyman and Sierra 2012), and a false initial premise often yields an untrustworthy conclusion. The 'large *n*' admonition in McQuitty (2004, 2018) is one such conclusion. If marketing researchers should eschew the NHST-related thinking that produces problematic conclusions, then marketing journals should encourage them by banning NHST and NHST-inspired thinking from its future articles.

References

Berk, R. A., & Freedman, D. A. (2003). Statistical assumptions as empirical commitments. In T. G. Blomberg, & S. Cohen (Eds.), *Law, punishment, and social control: Essays in honor of Sheldon Messinger* (2nd ed., pp. 235–254). New York, NY: Aldine de Gruyter.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Socio Meth & Res, 16*(1), 78–117.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York, NY: John Wiley & Sons.

Bradley, M. T., & Brand, A. (2016). Significance testing needs a taxonomy: Or how the Fisher, Neyman-Pearson controversy resulted in the inferential tail wagging the measurement dog. *Psych Rep, 119*(2), 487–504. https://doi.org/10.1177/0033294116662659.

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

Grice, J. W. (2017). Comment on Locascio's results blind manuscript evaluation proposal. *Basic and Applied Social Psychology, 39*(5), 254–255. https://doi.org/10.1080/01973533.2017.1336093.

Hubbard, R. (2016). *Corrupt research: The case for reconceptualizing empirical management and social science*. Los Angeles, CA: Sage Publications.

Hyman, M. R. (2017). Can 'results blind manuscript evaluation' assuage 'publication bias'? *Basic and Applied Social Psychology, 39*(5), 247–251. https://doi.org/10.1080/01973533.2017.1350581.

Hyman, M. R., & Sierra, J. J. (2012). Adjusting self-reported attitudinal data for mischievous respondents. *International Journal of Market Research, 54*(1), 129–145. https://doi.org/10.2501/IJMR-54-1-129-145.

Kline, R. (2017). Comment on Locascio, results blind science publishing. *Basic and Applied Social Psychology, 39*(5), 256–257. https://doi.org/10.1080/01973533.2017.1336093.

Li, H., Trafimow, D., Wang, T., Wang, C., & Hu, L. (2020). User-friendly computer programs so econometricians can run the a priori procedure. *Frontiers Management Business, 1*(1), 2–6. https://doi.org/10.25082/FMB.2020.01.002.

Locascio, J. (2017a). Results blind publishing. *Basic and Applied Social Psychology, 39*(5), 239–246. https://doi.org/10.1080/01973533.2017.1336093.

Locascio, J. (2017b). Rejoinder to responses to "results blind publishing". *Basic and Applied Social Psychology, 39*(5), 258–261. https://doi.org/10.1080/00031305.2018.1505658.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psych Meth, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130.

Marks, M. J. (2017). Commentary on Locascio 2017. *Basic and Applied Social Psychology, 39*(5), 252–253. https://doi.org/10.1080/01973533.2017.1350580.

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

McQuitty, S. (2018). Reflections on "Statistical power and structural equation models in business research". *Journal of Global Scholars of Marketing Science, 28*(3), 272–277. https://doi.org/10.1080/21639159.2018.1434806.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27(3/4), 379–423*, 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Skipper, R. B., & Hyman, M. R. (1987). Evaluating and improving argument-centered works in marketing. *Journal of Marketing, 51*(4), 60–75. https://doi.org/10.2307/1251248.

Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics. *Edu Psych Measure, 77*(5). https://doi.org/10.1177/0013164416667977.

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

Trafimow, D. (2019). 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.

Trafimow, D., & MacDonald, J. A. (2017). Performing inferential statistics prior to data collection. *Edu Psych Measure, 77*(2), 204–219. https://doi.org/10.1177/0013164416659745.

Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology, 37*(1), 1–2. https://doi.org/10.1080/01973533.2015.1012991.

Trafimow, D., & Myüz, H. A. (2019). The sampling precision of research in five major areas of psychology. *Behavior Research Methods, 51*(5), 2039–2058. https://doi.org/10.3758/s13428-018-1173-x.

Trafimow, D., Wang, T., & Wang, C. (2019). From a sampling precision perspective, skewness is a friend and not an enemy! *Edu Psych Measure, 79*(1), 129–150. https://doi.org/10.1177/0013164418764801.

Trafimow, D., Hyman, M. R., & Kostyk, A. (2020). The (im) precision of scholarly consumer behavior research. *Journal of Business Research, 114*, 93–101. https://doi.org/10.1016/j.jbusres.2020.04.008.

Wang, C., Wang, T., Trafimow, D., & Chen, J. (2019). Extending a priori procedure to two independent samples under skew normal settings. *Asian Journal of Economics and Banking, 3*(2), 29–40.

Wang, C., Wang, T., Trafimow, D., & Myüz, H. A. (2019). Desired sample size for estimating the skewness under skew normal settings. In V. Kreinovich, & S. Sriboonchitta (Eds.), *Structural changes and their economic modeling* (pp. 152–162). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-04263-9.

Wang, C., Wang, T., Trafimow, D., & Zhang, X. (2019). Necessary sample size for estimating the scale parameter with specified closeness and confidence. *Intl J Intel Tech and App Stat, 12*(1), 17–29. https://doi.org/10.6148/IJITAS.20190312(1).0002.

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. https://doi.org/10.1080/00031305.2019.1583913.

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476–487. https://doi.org/10.1016/j.elerap.2010.07.003.

Ziliak, S. T., & McCloskey, D. N. (2016). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: The University of Michigan Press. https://doi.org/10.3998/mpub.186351.

Dr. David Trafimow is a Distinguished Achievement Professor of Psychology at New Mexico State University, a Fellow of the Association for Psychological Science, and Executive Editor of the *Journal of General Psychology* and *Basic and Applied Social Psychology*. He received his Ph.D. in psychology from the University of Illinois at Urbana-Champaign in 1993. His current research interests include attribution, attitudes, cross-cultural research, ethics, morality, philosophy, philosophy of science, methodology, potential performance theory, and the a priori procedure.

Dr. Michael R. Hyman is a Distinguished Achievement Professor of Marketing at New Mexico State University in Las Cruces, New Mexico. His more than 100 academic journal articles, 60 conference papers (12 which received a 'best paper' award), four co-authored/co-edited books, 30 other academic contributions, and 50 non-academic works, attest to his writing compulsion. He has served on 16 editorial review boards and as a journal co-editor. Currently, he is a *Journal of Business Ethics* section editor and a *Journal of Marketing Theory and Practice* associate editor. His research interests include marketing theory, marketing ethics, consumer response to advertising, survey research methods, philosophical analyses in marketing, and marketing futurology. Now a loyal New Mexican, he splits his time between Las Cruces and Cloudcroft with his wife, four sons, three dogs, and three cats.

Dr. Alena Kostyk is a Lecturer in Marketing at the University of Glasgow in Glasgow, UK. Her work has appeared in academic outlets such as the *Journal of Business Research*, *European Journal of Marketing, Journal of Consumer Behaviour, International Journal of Market Research,* and *Journal of Research in Interactive Marketing*. Before entering academia, Dr. Kostyk spent nearly ten years working in the private sector. Her primary research projects focus on human decision making in futuristic environments and marketplaces, both from the theoretical and empirical perspectives. Her secondary research area lies within innovative research methodology.

Dr. Cong Wang is an Assistant Professor of Mathematics at the University of Nebraska in Omaha, Nebraska. She was a Graduate Student in the Department of Mathematics at New Mexico State University in Las Cruces, New Mexico. Her recent scholarly work has appeared in journals such as *New Ideas in Psychology, Educational and Psychological Measurement, Asian Journal of Economics and Banking, Communication in Statistics—Simulation and Computation*, and *International Journal of Intelligent Technology and Applied Statistics.*

Dr. Tonghui Wang is a Professor of Mathematics at New Mexico State University in Las Cruces, New Mexico. His recent scholarly work has appeared in journals such as *Communications in Statistics–Simulation and Computation, Asian Journal of Economics and Banking, Educational and Psychological Measurement, International Journal of Intelligent Technologies and Applied Statistics, New Ideas in Psychology, Journal of Systems Science and Complexity, International Journal of Innovative Science and Modern Engineering*, and *Journal of Uncertain Systems*.