Journal of World Business 52 (2017) 125–126

Contents lists available at ScienceDirect

Journal of World Business

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

Editorial

From the editors: New directions in the reporting of statistical results in the Journal of World Business

1. Introduction

Statistics and hypothesis testing have advanced understanding in the field of international businesses. Statistical analysis has allowed international business researchers to state that they either reject or fail to reject a prediction advanced as a hypothesis derived from theory. The common decision rule for procedure involves the calculation of *p*-values for parameters and comparing them to standardized thresholds. Yet, despite the longstanding practice of using the *p*-value as a default decision rule, there has also been an extensive history of criticism surrounding the use of this decision rule (e.g., Edwards, Lindman, & Savage, 1963, Berger & Sellke, 1987; Cobb, 2015; Demidenko, 2016; Meehl, 1978).

For many years the statistical community's response to the criticisms of *p*-values was rather muted. However, in 2015 the editorial leadership of *Basic and Applied Social Psychology* banned the publication of *p*-values in the journal, stating the null hypothesis significance testing procedure behind them was invalid (Trafimow & Marks, 2015). This decision was reached as the editors felt the bar of $p < 0.05$ was set too low, allowing weak results to pass off as established facts. The banning of *p*-values at *Basic and Applied Social Psychology* sent reverberations through academic and applied communities that are involved with data analytics.

The world's largest community of statisticians, the American Statistical Association (ASA), formed a committee to respond to the controversy by re-examining the usage of *p*-values in statistical reporting. Lengthy discussions ensued (see Wasserstein & Lazar, 2016) but the general consensus of the committee was that "the current culture of statistical significance testing, interpretation, and reporting has to go" (Ziliak, 2016, p. 2). The discussions led to a statement on *p*-values by the ASA. Ronald Wasserstein, the executive director of the ASA, stated "the *p*-value was never intended to be a substitute for scientific reasoning" (American Statistical Association, 2016, p. 1). Rather it was now the intention of the ASA "to steer research into a 'post $p < 0.05$ era"'.

Space precludes a reprint of the entire text of the ASA's statement but the following points are worth stating for the readership of JWB (Wasserstein, 2016, pp. 131–132):

1 Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold.
2 Proper inference requires full reporting and transparency.
3 A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result.
4 By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis.
5 Researchers should recognize that a *p*-value without context or other evidence provides limited information. For example, a *p*-value near 0.05 by itself offers only weak evidence against the null hypothesis. Likewise, a relatively large *p*-value does not imply evidence in favor of the null hypothesis; many other hypotheses may be equally or more consistent with the observed data. For these reasons, data analysis should not end with the calculation of a *p*-value when other approaches are appropriate and feasible.

One might ask what simple decision rule will replace the $p < 0.05$ criterion. The ASA has demurred on this issue and at the time of writing, there is no generally held consensus on what decision rules should be used. Various approaches have been proposed, including the use of replication studies (Sawyer & Peter, 1983; Singh, Ang, & Leong, 2003), the use of effect size estimates and confidence intervals (Aguinis et al., 2010; Ely, 1999; Hubbard and Meyer, 2013; Lin, Lucas, & Shmueli, 2013), and the use of Bayesian methods (Hahn, 2014; Sawyer & Peter, 1983), Bayes factors or likelihood ratios, and decision-theoretic modeling.

Nonetheless, the days of simple decision rules for inference and theory testing are passing due to the increasingly recognized limitations of these rules in the scientific community. Changes in the practice of statistical analysis as a means for further advancing knowledge are upon us. To this end, the editorial leadership of JWB has decided to be one of the first to incorporate these revisions to practice.

Academic journals have an important role to play in discipline-specific knowledge advancement by articulating (often tacitly but sometimes explicitly) what criteria are to be used to assess whether results from data represent new knowledge in terms of established or emerging theories. For decades *p*-values have been the dominant criterion. This leads to a situation where hypothesis testing based on *p*-values is adopted by authors so as to improve the likelihood of journal receptiveness, and where future scholars such as graduate students are taught these methods so as to better enhance their prospects of contributing to knowledge advancement. As a result, a reinforcing cycle regarding statistical decision making based on *p*-values is created and becomes increasingly solidified. We can therefore see that academic journals also have

http://dx.doi.org/10.1016/j.jwb.2016.12.003

the capability to help foster a more diversified environment for data-based decision making. This Perspectives piece gives an explicit broadening of the knowledge-assessment criteria beyond *p*-values to both its readers and to potential authors. Moreover, instead of a single criterion, the journal now encourages efforts toward multiple streams of evidence reinforcing claims of new knowledge generation.

2. Changes at JWB regarding statistical reporting

The threshold values of $p < 0.05$ and $p < 0.01$ provide a rough yardstick by which researchers and consumers of research implicitly assess the confidence of rejecting the null hypothesis. If the actual value of p is unknown, however, it is unclear if the actual p is indeed 0.049 or 0.011, which would numerically make a difference on paper. As such, reporting only whether a threshold has been reached (often designated with an asterisk such as * for $p < 0.05$) results in a loss of information (Aguinis et al., 2010). Accordingly JWB will now require authors to provide either standard errors or exact *p*-values (without asterisks) or both. JWB will also require authors to report the sample size in all statistical tables. If multiple analyses based on different sample sizes appear within one table, all sample sizes will need to be reported in such a way that it is clear which sets of results arise from which sample size.

In the absence of a consensus on statistical decision rules, JWB will now require authors to provide a statistical reporting narrative as a separate document accompanying submissions. In this narrative, authors will describe their efforts to ensure that empirical results are replicable and meaningful. For example, authors may wish to provide effect size or power calculations. The authors may conduct replication analyses or split the sample and use one half for theory testing and the other for replicability investigation. Authors should also reflect on whether their statistically significant results also possess substantive significance. When appropriate, descriptive statistics such as confidence intervals and total variance accounted for by a given variable, and the use of charts can help with this narrative. If there is a space constraint, authors are allowed to put this additional information in a separate document accompanying submissions. Further, effect sizes or equivalent that illustrate the strength of the relationships investigated should be part of the reporting of findings. These should be interpreted and discussed on their implications for theory and practice.

Given the replicability of results come from good research designs and processes, including good research decisions and judgment calls along the way, it would be important to have the research design well documented as far as possible. Decisions and choices made in the stages of defining population, sampling frame, and the actual sample and its representativeness need to be clearly justified and articulated. The extent to which extensive modeling efforts were made but not reported should also be documented in the narrative. For example, if various forms of models were considered in the process of the research but only one was eventually adopted, this should be documented for the purposes of transparency. Estimations with alternative versions of variables (logarithmic transformations or alternative factor analytic rotations) should also be disclosed.

Considerable attention should also be given to the process of data collection and sampling. It has to be clear that the ultimate sample units for analysis are representative of the intended population. This means that the intended population can be articulated and that all elements of said population have a known probability of selection into the sample. The implications for the work have to be discussed with this intended population in mind.

3. Summary

At present most empirical articles in JWB rely heavily on null-hypothesis significance testing and *p*-values to uncover knowledge and advance the discipline. The point of this perspectives article is that entrenched procedures for statistical reporting will be changing, and JWB will be changing now instead of later. In general JWB seeks to move toward a more pluralistic perspective on empirically-based research as opposed to the current *p*-value monoculture. Alternative approaches for establishing the meaningfulness and replicability of results (such as the example methodologies described above) are encouraged in statistical reporting at JWB.

References

Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*(3), 515–539.

American Statistical Association (2016). *American Statistical Association releases statement on statistical significance and p-values [press release]*. . . [Retrieved from] http://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of *p* values and evidence. *Journal of the American Statistical Association*, *82*(397), 112–122.

Cobb, G. (2015). More renovation is too little too late. *The American Statistician*, *69*(4), 266–282.

Demidenko, E. (2016). The p-value you can't buy. *The American Statistician*, *70*(1), 33–38.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193–242.

Ely, M. (1999). The importance of estimates and confidence intervals rather than p-values. *Sociology*, *33*(1), 185–190.

Hahn, E. D. (2014). *Bayesian methods for management and business: Pragmatic solutions for real problems*. NJ: Wiley Hoboken.

Hubbard, R., & Meyer, C. K. (2013). The rise of statistical significance testing in public administration research and why this is a mistake. *Journal of Business and Behavioral Sciences*, *25*(1), 4–20.

Lin, M., Lucas, H. C. Jr., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*, *24*(4), 906–917.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Sawyer, A. G., & Peter, J. P. (1983). The significance of statistical significance tests in marketing research. *Journal of Marketing Research*, *20*, 122–133.

Singh, K., Ang, S. H., & Leong, S. M. (2003). Increasing replication for knowledge accumulation in strategy research. *Journal of Management*, *29*(4), 533–549.

Trafimow, D., & Marks, M. (2015). *Editorial. Basic and Applied Social Psychology*, *37*(1), 1–2.

Wassersteinon, R.L., Editor, on behalf of the American Statistical Association Board of Directors, ASA Statement on statistical significance and p-values, The American Statistician, 70, 2, 2016, 131–133.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context process, and purpose. *The American Statistician*, *70*(2), 129–131.

Ziliak, S. T. (2016). *The ASA's statement on p-values: Context, process, and purpose*. . . [Available at Online supplement to Wasserman, R.L. and Lazar, N.A. (2016)] http://amstat.tandfonline.com/doi/suppl/10.1080/00031305.2016.1154108.

Eugene D. Hahn*
Salisbury University, Information & Decision Sciences, 1101 Camden Ave., MD 21801, United States

Siah Hwee Ang
Victoria University of Wellington, Wellington, New Zealand

* Corresponding author.
E-mail address: edhahn@salisbury.edu (E. Hahn).

Available online 7 January 2017