

7-29-2020

Why and How to Replace Statistical Significance Tests with Better Methods to Evaluate Hypotheses


Andreas Schwab

Iowa State University, aschwab@iastate.edu

William H. Starbuck

University of Oregon

Follow this and additional works at: https://lib.dr.iastate.edu/management_pubs

 Part of the [Business Analytics Commons](#), [Management Information Systems Commons](#), [Management Sciences and Quantitative Methods Commons](#), and the [Technology and Innovation Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/management_pubs/58. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Conference Proceeding is brought to you for free and open access by the Management at Iowa State University Digital Repository. It has been accepted for inclusion in Management Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Why and How to Replace Statistical Significance Tests with Better Methods to Evaluate Hypotheses

Abstract

This symposium will introduce and discuss how scholars can improve upon statistical significance tests, which continue to constrain the production of knowledge in management science. The extensive use of these tests in quantitative research has led to the accumulation of "statistically significant" results that are both too small to be practically relevant and so small that they are unlikely to replicate. A field that aspires to provide useful advice to managers needs to focus on practically important and robust effects. The proposed symposium introduces and discusses alternative approaches to overcome the limitations of statistical significance tests - such as, effect size measures, confidence intervals, graphs, meta-analyses, baseline modeling and the implications of these approaches for the accumulation of scientific knowledge. A final "Question and Answer" session will offer additional opportunities for further discussions, advice and recommendations.

Disciplines

Business Analytics | Management Information Systems | Management Sciences and Quantitative Methods | Technology and Innovation

Comments

This accepted proceedings is published as Schwab, A., Starbucks, W.H., Why and How to Replace Statistical Significance Tests with Better Methods to Evaluate Hypotheses. *Academy of Management Proceedings*, 2020(1); doi: [10.5465/AMBPP.2020.14228symposium](https://doi.org/10.5465/AMBPP.2020.14228symposium). Posted with permission.

This conference proceeding is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/management_pubs/58

Submission Number: 14228
Panel Symposium Proposal

**Why and How to Replace Statistical Significance Tests with
Better Methods to Evaluate Hypotheses**

Presenters and Panel Members

William H. Starbuck
University of Oregon

Andreas Schwab
Iowa State University

ABSTRACT

This symposium will introduce and discuss how scholars can improve upon statistical significance tests, which continue to constrain the production of knowledge in management science. The extensive use of these tests in quantitative research has led to the accumulation of "statistically significant" results that are both too small to be practically relevant and so small that they are unlikely to replicate. A field that aspires to provide useful advice to managers needs to focus on practically important and robust effects. The proposed symposium introduces and discusses alternative approaches to overcome the limitations of statistical significance tests -- such as, effect size measures, confidence intervals, graphs, meta-analyses, baseline modeling and the implications of these approaches for the accumulation of scientific knowledge. A final "Question and Answer" session will offer additional opportunities for further discussions, advice and recommendations.

Search Terms: Statistical Significance, Practical Significance, Effect Sizes, Confidence Intervals, Effect Size Distributions, Baseline Models, Research & Publication Norms.

Primary Sponsors
Organizational Behavior Division (OB)

January 13, 2020

SYMPOSIUM OVERVIEW

Statistical significance testing has been criticized by methodologists on various grounds and for a long time (Wasserstein et al., 2016; Schwab et al., 2011; Starbuck, 2006; Gigerenzer, 2004; Cortina & Folger, 1998; Schmidt, 1996; Cohen, 1994). This criticism suggests that the extensive use of statistical significance tests in quantitative management research has led to the accumulation of deceptive findings. For example, by gathering enough data even meaninglessly small effects are sufficient to disconfirm any related point null hypothesis. Consequently, management journals are full of "statistically significant" results that are too small to be practically relevant and unlikely to replicate in other studies.

In a field that aspires to provide useful advice to managers, we need to focus on practically important effects that are robust across a wide variety of settings. To identify practically meaningful findings, methodologists have recommended the reporting of effect sizes, confidence intervals, graphs, and baseline models as ways to move beyond statistical significance tests in the interpretation of empirical data (Cumming, 2012; Hubbard, 2015; Schwab & Starbuck, 2013). The *American Statistician* devoted an entire special issue to the topic of how to move to "a world beyond $p < 0.05$ " (Wasserstein et al., 2019) and the *Strategic Management Journal* no longer accepts publications that only apply fixed .05 statistical significance tests (Bettis et al., 2016). The application of alternatives to statistical significance tests, however, has been limited by both a lack of sensitivity regarding the inherent problems of statistical significance tests and a limited familiarity with alternative approaches.

SYMPOSIUM PRESENTATIONS

This symposium will include presentations by panel members on key topics as well as open discussions of these topics among symposium participants. After a brief topic introduction,

the following short presentations will create the base for a subsequent open discussion of relevant topics with the symposium participants.

In the first presentation, Bill Starbuck illustrates how the use of statistical significance tests has fostered an over-reliance on binary true-false interpretations of research findings and therefore journals' refusing to publish studies that do not report statistically significant findings. Researchers who believe their careers depend on publishing in top journals make tortured efforts to extract statistical significance through data manipulations and numerous unreported statistical analyses. Many researchers have become cynical about the purposes of research, treating hypotheses, data, and analyses as nothing more than means to obtain tenure or prestige. We must put a stop to such corrupting influences.

In the second presentation, Andreas Schwab summarizes key limitations of statistical significance tests. The presentation outlines the risks and likelihoods of both false-positive and false-negative empirical findings due to sample size sensitivity, dichotomized p-values, p-hacking, HARKING and other problematic practices. This presentation provides a systematic conceptual evaluation of the limitations of statistical significance tests. In addition, it introduces and discusses approaches on how to improve these currently prevalent problematic practices. The recommended improvements include focusing on effect sizes, confidence intervals, graphs capturing uncertainty of effects, and meta-analytic investigations.

In the third presentation, Andreas Schwab will further elaborate on the current state-of-the-art in effect size calculation by introducing various standardized and unstandardized effect size measures and their respective merits and limitations. He will illustrate the application of effect size evaluations in the context of an actual empirical study using data from a study of the training of nurses. He shows how statistical significance test can provide misleading results and

how effect size and confidence interval evaluations promise a much deeper and inherently more meaningful interpretation of empirical observations.

In the final presentation, William H. Starbuck introduces several approaches to overcoming the corrupting biases in current research practices. He endorses the explicit reporting of effect sizes and the uncertainty associated with effect sizes as an important step forward. More generally, researchers should acknowledge that theories have many, partially inconsistent properties. Comparing preferred theories with baseline models can foster stepwise improvements in theories. Robust regression techniques make more accurate estimates of parameters than does ordinary least squares regression. Theories allowing for only a few independent variables are more likely to generalize to new data, especially data about different contexts. Each research study ought to support generalizations to new data or to data about similar but different situations. Empirical experiments have shown that researchers who believe more than one theory is plausible can make more accurate predictions by averaging the predictions made by the alternative theories. Since each theory captures a part of the complex reality, averaging several alternative theories corrects in part for the deficiencies of individual theories. After the presentations by the panel members, symposium participants will have the opportunity to follow-up on any questions during a moderated Q&A session.

SYMPOSIUM SCHEDULE (Total length 90 min)

(1) Introduction (2.5 min)

(2) Panelist Presentations (60 min)

- How bias against null findings corrupts research into a careerist game: Torturing data to make them obey
(Bill Starbuck, 15 min)
- What is wrong with statistical significance tests and how to evaluate hypotheses instead
(Andreas Schwab, 15 min)
- Effect size and effect size distributions: Alternative ways to evaluate hypotheses

(Andreas Schwab, 15 min)

- Useful alternatives to statistical significance

(Bill Starbuck, 15 min)

(3) Moderated Panel Discussion and Q&A Segment (25 min)

(4) Final Remarks (2.5 min)

TARGET AUDIENCE

We observe that an increasing number of scholars are critical regarding how researchers use statistical significance tests in management research (Hubbard, 2015; Nuzzo, 2014; Schwab et al., 2011; Schwab & Starbuck, 2017; Goldfarb & King, 2016; Orlitzky, 2012). Some leading journals, such as *SMJ* (Bettis, et al., 2016), and the APA Manual (2010) now recommend that authors move beyond statistical significance tests. Initiatives by the American Statistical Association (Wasserstein et al., 2019), the Center for Open Science (<https://cos.io/>) and Responsible Research in Business and Management (RRBM; <https://rrbm.network/>) now promote related advancements of research methods. The focus of this symposium is to push the related debate from recognizing limitations of statistical significance tests toward discussing specific ways to improve upon them (Schwab, et al., 2011; Aguinis et al., 2009). We believe that a deeper understanding and debate about approaches to improve on statistical significance tests represent a crucial next step toward a change in methodological practices in the field of management research. Based on the wide-spread use of statistical significance tests across all AOM divisions, the proposed symposium addresses issues relevant across research streams and we anticipate strong interest across all divisions.

The ultimate objective of this symposium is to contribute to the improvement of research quality in the field of management. The recognition of such methodological opportunities will enable researchers to implement improved approaches in their individual studies. The symposium, however, also has a broader and more long-term objective. We hope it will

contribute to changes of methodological conventions related to training, reviewing and publishing in our field. These collective changes require broader agreement among management scholars. This symposium intends to stimulate and support related discussions among scholars. These discussions are a necessary foundation for any successful collective methodological change initiatives in the future.

SYMPOSIUM-RELATED WEBSITE

We have already created a website to facilitate communication with and among symposium participants after the symposium (<https://sites.google.com/site/nhstresearch/>). This website will ultimately provide access to all the presentation slides. It already contains references and links to additional more detailed and comprehensive information on each of the topics and issues covered by the proposed symposium. This website will enable continuing communication with scholars after the symposium to support their efforts to apply some of the introduced methodological alternatives in their research. We believe such continuing communication efforts are crucial for supporting more fundamental long-term collective change in methodological conventions in the field of management research. Hence, we will promote the website during the symposium and familiarize participants with its features.

In summary, we consider the proposed 2020 AOM symposium an important part of a much broader currently ongoing initiative to raise the awareness of AOM members about the severe limitations of statistical significance tests and to discuss and develop alternative approaches that address these limitations. Based on feedback from related activities at prior AOM conferences and the increasing attention to related issues in the management literature, we consider the topic of alternatives to statistical significance tests of high interest to AOM conference participants. Given the intensive application of statistical significance tests across all

AOM divisions, we consider this symposium relevant to quantitative researchers in the RM, HR, OB and all divisions.

REFERENCES

- American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: American Psychological Association.
- Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2009). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organization Research Methods*, 13(3), 515-539.
- Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (Forthcoming). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cortina, J.M., & Folger, R.G. (1998). When is it acceptable to accept a null hypothesis: No way, Jose? *Organizational Research Methods*, 1, 334-350.
- Cumming, G. (2011). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, New York.
- Cumming, G. & Finch, S. (2002). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, 61, 532-575.
- Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis and the interpretation of research results*. Cambridge, UK: Cambridge University Press.
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33, 587-606.
- Goldfarb, B., & King, A.A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, 37(1), 167-176.
- Hubbard, R. (2015). *Corrupt Research: The Case for Reconceptualizing Empirical Management and Social Science*. SAGE Publications.
- Kline, R. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychology Association.
- Nuzzo, R. (2014). Statistical errors. *Nature*, 506(7487), 150-152.
- Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods*, 15(2), 199-228.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, 1, 115-129.

- Schwab, A., Abrahamson, E., Fidler, F., & Starbuck, W. H. (2011). Researchers should make thoughtful assessments instead of null-hypothesis significance tests. *Organization Science*, 22(4), 1105-1120.
- Schwab, A., & Starbuck, W. H. (2013). Why baseline modelling is better than null-hypothesis testing: Examples from international business research. In T. M. Devinney, T. Pedersen & L. Tihanyi (Eds.), *Philosophy of Science and Meta-Knowledge in International Business and Management* (pp. 171-195): Emerald.
- Schwab, A. & Starbuck, W. H. (2017). A Call for Openness in research reporting: How to turn problematic practices into helpful tools. *Academy of Management Learning & Education*, 16(1), 125-141; [DOI: 10.5465/amle.2016.0039](https://doi.org/10.5465/amle.2016.0039).
- Starbuck, W. H. (2006). *The production of knowledge: The challenge of social science research*. Oxford University Press, New York.
- Wasserstein, R.L. & Lazar, N.A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.
- Wasserstein, R.L., Schirm, A.L. & Lazar, N.A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73, 1-19.

PANELISTS AND PRESENTERS

William H. Starbuck

William Starbuck is courtesy professor in residence at the Lundquist College of Business of the University of Oregon and professor emeritus at New York University. He received his M.S. and Ph.D. in industrial administration at Carnegie Institute of Technology, after receiving an A.B. in physics at Harvard. He has also been awarded honorary doctorates by universities in Stockholm, Paris, and Aix-en-Provence. He has held faculty positions in economics, sociology, or management at Purdue University, Johns Hopkins University, Cornell University, University of Wisconsin-Milwaukee, and New York University, as well as visiting positions in England, France, New Zealand, Norway, Sweden, and the United States. He was also senior research fellow at the International Institute of Management in Berlin. He edited *Administrative Science Quarterly*, chaired the screening committee for Fulbright awards in business management, directed the doctoral program in business administration at New York University, and was President of the Academy of Management. He is a fellow of the Academy of Management, the American Psychological Association, the American Psychological Society, the British Academy of Management, and the Society for Industrial and Organizational Psychology.

He has published over 160 articles on accounting, bargaining, business strategy, computer programming, computer simulation, forecasting, decision making, human-computer interaction, learning, organizational design, organizational growth and development, perception, scientific methods, and social revolutions. He has also authored two books and edited seventeen books. In his latest book *The Production of Knowledge*, he reflects on his own academic journey and on the challenges associated with management and social science research.

Andreas Schwab

Andreas Schwab is an Associate Professor and Dean's Fellow in the College of Business at Iowa State University. He received his Ph.D. in management from the University of Wisconsin-Madison. He studies organizational learning processes in dynamic environments, such as project ventures and temporary organizations. He is also engaged in several initiatives to advance methodological practices in the management field by promoting alternatives to statistical significance tests and introducing scholars to Bayesian statistics.

His research has been published in the *Academy of Management Journal*, *Organization Science*, *Academy of Management Learning & Education*, *Strategic Organization*, *Entrepreneurship Theory & Practice*, *Industrial and Corporate Change*, *Industrial Relations*, *Family Business Review*, *Group and Organization Management*, *Management and Organization Review* and other outlets. He was a Contributing Editor for *Research Methodology and Statistics at Entrepreneurship, Theory & Practice* and serves on the editorial boards of *Organization Science*, *Group and Organization Management*, and the *Strategic Entrepreneurship Journal*. He was a representative-at-large for the Research Method Division of the Academy of Management and is the current Chair of the Research Community of the Strategic Management Society.