The Annals of Applied Statistics
2021, Vol. 15, No. 3, 1084–1085
https://doi.org/10.1214/21-AOAS1501

THE ASA PRESIDENT'S TASK FORCE STATEMENT ON STATISTICAL SIGNIFICANCE AND REPLICABILITY

BY YOAV BENJAMINI[1], RICHARD D. DE VEAUX[2], BRADLEY EFRON[3], SCOTT EVANS[4], MARK GLICKMAN[5,*], BARRY I. GRAUBARD[6], XUMING HE[7], XIAO-LI MENG[5,†], NANCY REID[8], STEPHEN M. STIGLER[9], STEPHEN B. VARDEMAN[10], CHRISTOPHER K. WIKLE[11], TOMMY WRIGHT[12], LINDA J. YOUNG[13] AND KAREN KAFADAR[14]

[1]*Department of Statistics and Operations Research, Tel Aviv University, ybenja@tauex.tau.ac.il*
[2]*Department of Mathematics and Statistics, Williams College, deveaux@williams.edu*
[3]*Department of Statistics and Department of Biomedical Data Sciences, Stanford University, brad@stat.stanford.edu*
[4]*Department of Biostatistics & Bioinformatics, George Washington University, sevans@bsc.gwu.edu*
[5]*Department of Statistics, Harvard University, *glickman@fas.harvard.edu; †meng@stat.harvard.edu*
[6]*Biostatistics Branch, National Cancer Institite, barry.graubard@nih.gov*
[7]*(Co-chair), Department of Statistics, University of Michigan, xmhe@umich.edu*
[8]*Department of Statistics, University of Toronto, reid@utstat.utoronto.ca*
[9]*Department of Statistics, University of Chicago, stigler@uchicago.edu*
[10]*Department of Statistics and Department of Industrial & Manufacturing Systems Engineering, Iowa State University, vardeman@iastate.edu*
[11]*Department of Statistics, University of Missouri, wiklec@missouri.edu*
[12]*Center for Statistical Research and Methodology, United States Bureau of the Census, tommy.wright@census.gov*
[13]*(Co-chair), Research & Development, National Agricultural Statistics Service, linda.j.young@usda.gov*
[14]*(Ex-officio), Department of Statistics, University of Virginia, kkafadar@virginia.edu*

Over the past decade, the sciences have experienced elevated concerns about replicability of study results. An important aspect of replicability is the use of statistical methods for framing conclusions. In 2019 the President of the American Statistical Association (ASA) established a task force to address concerns that a 2019 editorial in *The American Statistician* (an ASA journal) might be mistakenly interpreted as official ASA policy. (The 2019 editorial recommended eliminating the use of "$p < 0.05$" and "statistically significant" in statistical analysis.) This document is the statement of the task force, and the ASA invited us to publicize it. Its purpose is two-fold: to clarify that the use of P-values and significance testing, properly applied and interpreted, are important tools that should not be abandoned, and to briefly set out some principles of sound statistical inference that may be useful to the scientific community.

P-values are valid statistical measures that provide convenient conventions for communicating the uncertainty inherent in quantitative results. Indeed, P-values and significance tests are among the most studied and best understood statistical procedures in the statistics literature. They are important tools that have advanced science through their proper application.

Much of the controversy surrounding statistical significance can be dispelled through a better appreciation of uncertainty, variability, multiplicity, and replicability. The following general principles underlie the appropriate use of P-values and the reporting of statistical significance and apply more broadly to good statistical practice.

Capturing the uncertainty associated with statistical summaries is critical. Different measures of uncertainty can complement one another; no single measure serves all purposes. The sources of variation that the summaries address should be described in scientific articles and reports. Where possible, those sources of variation that have not been addressed should also be identified.

Received June 2021.

Dealing with replicability and uncertainty lies at the heart of statistical science. Study results are replicable if they can be verified in further studies with new data. Setting aside the possibility of fraud, important sources of replicability problems include poor study design and conduct, insufficient data, lack of attention to model choice without a full appreciation of the implications of that choice, inadequate description of the analytical and computational procedures, and selection of results to report. Selective reporting, even the highlighting of a few persuasive results among those reported, may lead to a distorted view of the evidence. In some settings this problem may be mitigated by adjusting for multiplicity. Controlling and accounting for uncertainty begins with the design of the study and measurement process and continues through each phase of the analysis to the reporting of results. Even in well-designed, carefully executed studies, inherent uncertainty remains, and the statistical analysis should account properly for this uncertainty.

The theoretical basis of statistical science offers several general strategies for dealing with uncertainty. P-values, confidence intervals and prediction intervals are typically associated with the frequentist approach. Bayes factors, posterior probability distributions and credible intervals are commonly used in the Bayesian approach. These are some among many statistical methods useful for reflecting uncertainty.

Thresholds are helpful when actions are required. Comparing P-values to a significance level can be useful, though P-values themselves provide valuable information. P-values and statistical significance should be understood as assessments of observations or effects relative to sampling variation, and not necessarily as measures of practical significance. If thresholds are deemed necessary as a part of decision-making, they should be explicitly defined based on study goals, considering the consequences of incorrect decisions. Conventions vary by discipline and purpose of analyses.

In summary, P-values and significance tests, when properly applied and interpreted, increase the rigor of the conclusions drawn from data. Analyzing data and summarizing results are often more complex than is sometimes popularly conveyed. Although all scientific methods have limitations, the proper application of statistical methods is essential for interpreting the results of data analyses and enhancing the replicability of scientific results.

> "*The most reckless and treacherous of all theorists is he who professes to let facts and figures speak for themselves, who keeps in the background the part he has played, perhaps unconsciously, in selecting and grouping them.*" (Alfred Marshall, 1885)

SUPPLEMENTARY MATERIAL

Supplement A to "The ASA president's task force statement on statistical significance and replicability" (DOI: 10.1214/21-AOAS1501SUPPA; .pdf). Brief biographies of Task Force members.

Supplement B to "The ASA president's task force statement on statistical significance and replicability" (DOI: 10.1214/21-AOAS1501SUPPB; .pdf). Chinese translation of the statement.