Abandoning statistical significance is both sensible and practical

Valentin Amrhein[1], Andrew Gelman[2], Sander Greenland[3],
Blakeley B. McShane[4]

[1] Department of Environmental Sciences, Zoology, University of Basel, Basel, Switzerland;
v.amrhein@unibas.ch

[2] Department of Political Science and Department of Statistics, Columbia University,
New York City, NY, USA; gelman@stat.columbia.edu

[3] Department of Epidemiology and Department of Statistics, University of California,
Los Angeles, CA, USA; lesdomes@g.ucla.edu

[4] Kellogg School of Management, Northwestern University, Evanston, IL, USA;
b-mcshane@kellogg.northwestern.edu

To the Editor of *JAMA*

Dr Ioannidis[1] writes against our proposals[2-4] to abandon statistical significance in scientific reasoning and publication, as endorsed in the editorial[5] of a recent special issue of an American Statistical Association journal devoted to moving to a "post $p<0.05$ world." We appreciate that he echoes our calls for "embracing uncertainty, avoiding hyped claims…and recognizing 'statistical significance' is often poorly understood." We also welcome his agreement that the "interpretation of any result is far more complicated than just significance testing" and that "clinical, monetary, and other considerations may often have more importance than statistical findings."

Nonetheless, we disagree that a statistical significance-based "filtering process is useful to avoid drowning in noise" in science and instead view such filtering as harmful. First, the implicit rule to not publish nonsignificant results biases the literature with overestimated effect sizes and encourages "hacking" to get significance. Second, nonsignificant results are often wrongly treated as zero. Third, significant results are often wrongly treated as truth rather than as the

noisy estimates they are, thereby creating unrealistic expectations of replicability. Fourth, filtering on statistical significance provides no guarantee against noise. Instead, it amplifies noise because the quantity on which the filtering is based (the p-value) is itself extremely noisy and is made more so by dichotomizing it.

We also disagree that abandoning statistical significance will reduce science to "a state of statistical anarchy." Indeed, the journal *Epidemiology* banned statistical significance in 1990 and is today recognized as a leader in the field.

Valid synthesis requires accounting for all relevant evidence—not just the subset that attained statistical significance. Thus, researchers should report more, not less, providing estimates and uncertainty statements for *all* quantities, justifying any exceptions, and considering ways the results are wrong. Publication criteria should be based on evaluating study design, data quality, and scientific content—not statistical significance.

Decisions are seldom necessary in scientific reporting. However, when they are required (as in clinical practice), they should be made based on the costs, benefits, and likelihoods of all possible outcomes, not via arbitrary cutoffs applied to statistical summaries such as p-values which capture little of this picture.

The replication crisis in science is not the product of the publication of unreliable findings. The publication of unreliable findings is unavoidable: as the saying goes, if we knew what we were doing, it would not be called research. Rather, the replication crisis has arisen because unreliable findings are presented as reliable.

References

[1] Ioannidis JPA. The importance of predefined rules and prespecified statistical analyses: Do not abandon significance. *JAMA.* Published online April 4, 2019. doi:10.1001/jama.2019.4582

[2] Amrhein V, Greenland S, McShane B. Retire statistical significance. *Nature*. 2019, 567, 305-307. doi:10.1038/d41586-019-00857-9

[3] McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am. Stat.* 2019, 73:sup1, 235-245. doi:10.1080/00031305.2018.1527253

[4] Amrhein V, Trafimow D, Greenland S. Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *Am. Stat.* 2019, 73:sup1, 262-270. doi:10.1080/00031305.2018.1543137

[5] Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "*p*<0.05". *Am Stat*. 2019, 73:sup1, 1-19. doi:10.1080/00031305.2019.1583913