Check for updates

Received: 15 August 2019 | Accepted: 15 August 2019

DOI: 10.1111/eci.13165

EDITORIAL

WILEY

When we make recommendations for scientific practice, we are (at best) acting as social scientists

The foundations and the practice of statistics are in turmoil, with corresponding threads of argument in biology, economics, political science, psychology, public health, and other fields that rely on quantitative research in the presence of variation and uncertainty. Lots of people (myself included) have strong opinions on what should not be done, without always being clear on the best remedy. This makes sense: good work can be difficult and context dependent, while bad work can be easy to identify.

Given all the open questions in applied statistics, it is no surprise that statistical reformers (including not just statisticians but also users and developers of statistics in medicine and other fields) are proceeding on several tracks, including changes in the publication process (preregistration, results-blind reviewing, open post-publication review, etc), expansion of scientific culture (open data and code, more active inclusion of women, ethnic minorities, and researchers outside of traditional academic power centers), recognition of uncertainty (going beyond binary statistical significance rules, greater acknowledgment of researcher degrees of freedom), identification of particular poor statistical practices (most notably regarding misunderstandings of *P*-values, power analyses, and causal identification), and research into improved methods.

Like many people, I think the current system of scientific research and communication has big problems, especially when it comes to handling uncertainty,[1] and, even if I cannot offer a once-size-fits all replacement, I think there are useful directions for improvement. This sort of attitude motivates people, my colleagues and myself included, to not just criticize but to suggest incremental improvements for general practice. One such suggestion that received much attention was a paper[2] with 71 authors, including John Ioannidis, that proposed "to change the default *P*-value threshold for statistical significance from .05 to .005 for claims of new discoveries," which was followed up by an 88-authored paper proposing that "researchers should transparently report and justify" their significance levels. Some colleagues and I disagreed with both these recommendations and wrote our own paper[3] recommending that statistical significance be abandoned entirely. A related paper[4] was accompanied by an endorsement letter with over 800 signatures.

The challenge is that the debates we are having are both scientific and sociological. The merits of various proposed reforms can be argued on theoretical grounds or on the basis of individual cases, it would be impossible from a purely empirical basis to judge the potential effects of proposed policies.

On a mathematical or scientific basis we can correct misconceptions about *P*-values and causal inference, we can argue about the reasonableness of particular assumptions (for example, in various attempts to connect Bayesian and classical hypothesis testing), and we can conduct empirical work, for example finding[5] from "surveys of researchers across a wide variety of fields" that "a substantial majority" make systematic errors regarding the interpretation of data and statistical significance. We can also make scientific arguments for how a particular alternative procedure in some particular example or class of problems.

But when it comes to speculating about the potential efficacy of solutions, we are, at best, doing social science--and speculative social science at that. I agree with Hardwicke and Ioannidis[6] that it is not clear how the political act of signing a petition is supposed to advance a scientific argument. It can, however, be relevant given that there are policy questions at hand. I signed the form because I feel that this would do more good than harm, but I fully respect the position of not signing any petitions. I don't think that my signing of the form is an act of campaigning or politics. I just think it's a shorthand way of saying that I agree with the general points of the published article and that I agree with most of its recommendations. I'm more comfortable signing on and endorsing the general message of a three-authored article,[4] than I would be to add my name to an article with 70 or 80 authors.[2,7]

One issue that came up in this discussion is, what's the point of those articles with all those authors, or the letter with all those signatories? Is it mob rule, the idea that scientific positions should be determined by those people who are loudest and most willing to express strong opinions? Or is it "the silent majority" representing sensible opinion? Or does it represent an attempt by well-connected elites to tell people what to think? Are these mass efforts attempting to serve a gatekeeping function by restricting how researchers can analyze their data? Or can this all be seen as a crude attempt to establish a consensus of the scientific community?

© 2019 Stichting European Society for Clinical Investigation Journal Foundation

None of these seem so great! Science should be determined my truth, accuracy, reproducibility, strength of theory, real-world applicability, moral values, etc All sorts of things, but these should not be the property of the mob, or the elites, or gatekeepers, or a consensus.

That said, the mob, the elites, gatekeepers, and the consensus aren't going anywhere, and I see massively-authored papers and long lists of signatories as being part of the necessary process of influencing the influencers. Also a valuable step in this process are critical assessments such as that of Hardwicke and Ioannidis,[6] along with criticisms of such criticisms, and so forth.

Andrew Gelman
Department of Statistics and Department of Political Science, Columbia University, New York

Correspondence
Andrew Gelman, Department of Statistics and Department of Political Science, Columbia University, New York.
Email: gelman@stat.columbia.edu

REFERENCES

1. Gelman A, Carlin J, Nallamothu B. Objective Randomised Blinded Investigation With Optimal Medical Therapy of Angioplasty in Stable Angina (ORBITA) and coronary stents: A case study in the analysis and reporting of clinical trials. *Am Heart J*. 2019;214:54-59.
2. Benjamin DJ, et al. Redefine statistical significance. *Nature Human Behaviour*. 2018;2:6-10.
3. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat*. 2019;73(S1):235-245.
4. Amrhein V, Greenland S, McShane B. Retire statistical significance. *Nature*. 2019;567:305-307.
5. McShane BB, Gal D. Blinding us to the obvious? The effect of statistical training on the evaluation of evidence. *Manage Sci*. 2016;62:1707-1718.
6. Hardwicke TA, Ioannidis J. Petitions in scientific argumentation: Dissecting the request to retire statistical significance. *Eur J Clin Invest*. 2019;2019(49):e13162.
7. Lakens D, et al. Justify your alpha. Nature Human. *Behaviour*. 2018;2:168-171.