Title
The Other Arbitrary Cutoff

Journal
The American Statistician, 73(4)

ISSN
0003-1305 (Print) 1537-2731 (Online)

Author
Bacchetti, Peter

Publication Date
2019-08-29

Peer reviewed

Below is the author's accepted version of
Bacchetti P (2019). The Other Arbitrary Cutoff. *The American Stat*istician, **73**:4, 413-414.
The published version is at: https://doi.org/10.1080/00031305.2019.1654920.

The Other Arbitrary Cutoff

While I welcome the call to abandon statistical significance (Wasserstein, Schirm, and Lazar 2019), there is another arbitrary cutoff, power≥80%, that is also quite harmful (Bacchetti 2010). Although power requirements may be less frequently enforced than p-value requirements in some fields, I believe that efforts to improve the use of statistics in research will have more success if they are consistent and comprehensive, which will require us to follow the logic and spirit of "beyond $p<0.05$" into the realm of sample size justification and grant proposals. In medical research, this realm has entrenched conventions that are incompatible with moving beyond $p<0.05$, so substantial cultural change will be needed. Although Wasserstein, et al. (2019), along with some other concurrent calls for abandoning the concept of statistical significance (Amrhein, Greenland, and McShane 2019; Hurlbert, Levine, and Utts 2019; McShane, et al. 2019), did not emphasize the need for changes at the grant proposal stage, I believe that addressing the challenges in this area could substantially enhance the prospects for meaningful reform.

Abandoning the concept of statistical significance implies also abandoning the concept of statistical power, because power is only meaningful if what matters most about a planned study is whether or not it reaches statistical significance. Furthermore, the conventional requirement for at least 80% power is an additional arbitrary cutoff (Bacchetti 2010). Unfortunately, statisticians and researchers have amassed much experience with power-based sample size justification, and many might feel adrift if this came to be recognized as a poor approach. In the special issue that Wasserstein, et al. (2019) introduced, a few articles presented alternatives to power-based sample size planning (Manski 2019; Manski and Tetenov 2019; Trafimow 2019), but they appeared to retain an undesirable feature of the conventional approach: calculations designed to achieve a fixed goal for just one aspect of the many things that change as sample size increases. In particular, the calculations do not involve factors that worsen as sample size increases, such as cost, feasibility, and the number of study participants put at risk. Ignoring such factors while calculating sample size would only make sense if there were a meaningful

dichotomization of sample size into “adequate” and “inadequate,” so that the study should either be done right or not at all. The reality of how sample size influences the projected scientific or practical value of a study, however, precludes any valid justification for such dichotomization (Bacchetti, McCulloch, and Segal 2008), even in the unusual case where study results are to be dichotomized into a binary decision. In actual practice, costs are usually given some consideration, at least informally, but this is rarely spelled out in the formal sample size justifications that appear in grant proposals and publications. I recommend formal methods that better align with practical considerations, specifically methods that weigh the information gains from larger sample sizes against the increased costs, thereby identifying a reasonable tradeoff (Bacchetti, et al. 2008; Wilson 2015).

The recommendations from Wasserstein, et al. (2019) to “accept uncertainty” and “be modest” may be particularly difficult to follow in grant proposals. My impression over a 35 year career in medical research is that nearly every study is expected to “answer the question,” and the concept of an “adequate” sample size is a key underpinning of the widespread expectation that every study should reach a final, definitive conclusion about the issue it investigates. Although this is usually quite unrealistic, grant proposals generally seek to comply, phrasing their aims as “to determine whether” and almost never as “to estimate how much.” They also follow power-based conventions to claim “adequate” sample size. McShane, et al. (2019) acknowledged that failing to “blithely promise 80% power” may cause a proposal to miss out on funding, but I would go further and argue that the modesty needed to go beyond $p<0.05$ would usually amount to peer review suicide for those seeking funding in the current culture, at least in the medical area. Individual researchers are therefore not in a good position to lead in this regard, and institutional change may be needed to blaze the trail. An ASA statement on power calculation might be a worthwhile contribution. A clear and forceful statement of the problems with current practices could allow forward-thinking funders or consortiums of funders to discourage power calculations in applications. That would empower applicants to safely venture beyond power$\geq$80% and beyond $p<0.05$.

Peter Bacchetti
Professor of Biostatistics
Department of Epidemiology and Biostatistics

University of California
San Francisco, California, USA
peter.bacchetti@ucsf.edu

Amrhein, V., Greenland, S., and McShane, B. (2019), "Retire Statistical Significance," *Nature*, 567, 305-307.

Bacchetti, P. (2010), "Current Sample Size Conventions: Flaws, Harms, and Alternatives," *BMC Medicine*, 8, 17.

Bacchetti, P., McCulloch, C. E., Segal, M.R. (2008), "Simple, Defensible Sample Sizes Based on Cost Efficiency" (with discussion), *Biometrics*, 64, 577–594.

Hurlbert, S., Levine, R., and Utts, J. (2019), "Coup de Grâce for a Tough Old Bull: 'Statistically Significant' Expires," *The American Statistician*, 73, 352-357.

Manski, C. (2019), "Treatment Choice With Trial Data: Statistical Decision Theory Should Supplant Hypothesis Testing," *The American Statistician*, 73, 296–304.

Manski, C., and Tetenov, A. (2019), "Trial Size for Near Optimal Choice between Surveillance and Aggressive Treatment: Reconsidering MSLTII," *The American Statistician*, 73, 305–311.

McShane, B., Gal, D., Gelman, A., Robert, C., and Tackett, J. (2019), "Abandon Statistical Significance," *The American Statistician*, 73, 235–245.

Trafimow, D. (2019), "Five Nonobvious Changes in Editorial Practice for Editors and Reviewers to Consider When Evaluating Submissions in a Post $p < .05$ Universe," *The American Statistician*, 73, 340–345.

Wasserstein, R. L., Schirm A .L., and Lazar, N. A. (2019), "Moving to a World Beyond '$p < 0.05$'", *The American Statistician*, 73, 1-19.

Wilson, E. C. (2015), "A practical guide to value of information analysis," *Pharmacoeconomics*, 33, 105–21.