Journal of Patient-Centered Research and Reviews

Volume 10 | Issue 3 Article 1

7-18-2023

Interpreting *P* Values in 2023

Jennifer K. Homa-Bonell

Recommended Citation

Homa-Bonell JK. Interpreting P values in 2023. J Patient Cent Res Rev. 2023;10:102-3. doi: 10.17294/2330-0698.2064

Published quarterly by Midwest-based health system Advocate Aurora Health and indexed in PubMed Central, the Journal of Patient-Centered Research and Reviews (JPCRR) is an open access, peer-reviewed medical journal focused on disseminating scholarly works devoted to improving patient-centered care practices, health outcomes, and the patient experience.

Interpreting *P* Values in 2023

Jennifer K. Homa-Bonell, MS

Advocate Aurora Research Institute, Aurora BayCare Medical Center, Green Bay, WI

Biostatisticians are often asked, much to our chagrin, "What's the *P* value?" We have mixed feelings about this question. It's not that we dislike *P* values, we appreciate what they tell us (ie, the probability of obtaining particular results assuming the null hypothesis is true[1]). However, we are also well aware of what *P* values *don't* tell us (eg, the magnitude of the association and its clinical importance), which is why the American Statistical Association (ASA) encourages broader discussion and interpretation of results beyond looking at a *P* value by itself.[2]

In truth, this can be hard to implement in practice. Pressure toward binary reporting, publication bias, and fear of manuscript rejection all conspire to reinforce the conventional misuse of *P* values that has developed over decades.[1,3] And yet, the biostatistician community has recently noticed a most-welcome culture shift in that more researchers, editors, and reviewers are engaging in dialogue around this topic and see the value in more comprehensive reporting of results.[4] Herein, we offer guidance for researchers who would like to incorporate more comprehensive reporting in their research papers.

In 2016, the ASA published a statement on statistical significance and *P* values in hopes of drawing renewed attention to this important topic.[2] In nontechnical terms, the article offered a few principles underlying the proper use and interpretation of the *P* value, intending to reach writers who are not primarily statisticians.[2] The ASA statement succeeded in outlining appropriate statistical practice for the broader research community, and this editorial aims to suggest how to incorporate and operationalize its principles.

Confidence Intervals

A confidence interval (CI) is an interval estimate calculated from sample data that indicates the precision of a point estimate.[1] It provides a range of plausible values you expect your point estimate (eg, sample mean) to fall between if you repeat your experiment with a given level of confidence.[1] Whether your discipline uses the *AMA Manual of Style*[3] or the *Publication Manual of the American Psychological Association*[5] (APA Style), both strongly recommend including both the *P* value and CI wherever possible.

P values only provide information about whether the null hypothesis is rejected.[3] CIs convey the degree of precision of point estimates as well as statistical significance.[3] Precision is indicated by the width of a CI, in that wide CIs mean low precision and narrow CIs mean high precision (eg, an odds ratio of 0.5 with a 95% CI of 0.05 to 4.5 indicates low precision of the estimate). A standalone *P* value does not convey precision of the estimate.[1,3] The *P* value and CI together provide more information — *important* information — than either estimate alone.[3,5]

Effect Sizes

Effect size is a numeric value measuring the strength or magnitude of the relationship between two variables.[1,3] It indicates *how much* association there is and the importance of an observed effect.[1,3] For all relevant results, including statistically nonsignificant findings, *P* values should be accompanied by effect size estimates.[1,2] The relative size or strength of a research outcome indicates the practical significance of the results, something which *P* values do not measure.[2]

Since *P* values depend not only on the effect size but also the sample size, it is not unusual for studies with large sample sizes to produce small *P* values even when the actual treatment effect is so small as to be clinically irrelevant.[2,4] Similarly, large effects may yield unimpressive *P* values if the sample size is too small to detect a meaningful difference.[2,4] Discussing effect size estimates in your interpretation will provide information on the magnitude of the association or phenomenon of interest.[1,3,5]

Clinical Implications

Deriving a *P* value, confidence interval, and effect size estimate yields strong statistical support, but the researcher's job should not end there. The next step is to

Corresponding author: Jennifer K. Homa-Bonell, Aurora BayCare Medical Center, 2845 Greenbrier Rd., Green Bay, WI 54311 (jennifer.homa@aah.org)

discuss the clinical significance and practical context of the results, in addition to the *P* value, when interpreting results.[1,2] Attempt to interpret the relevance of a given result by explaining how it might impact your practice and/or study population. Translating findings into real-life scenarios helps to convey the importance (or lack thereof) of the results.

Results Near Statistical Significance

Statistical significance is established for each quantitative research study as the point at which rejection of the null hypothesis is demonstrated by the statistical analysis; the concept is generally (but not always) interpreted as a *P* value of ≤0.05.[1-3] When reporting results that were close to, but did not reach, statistical significance as prespecified in the study design, a critical question is whether the sample size was large enough to detect meaningful differences (ie, what was the power of the study?). Statistical power is the ability to detect an effect if it truly does exist.[1,3]

Many studies have insufficient power, and it's important to recognize and discuss this limitation.[1] Make sure not to overreach or overaggrandize findings. For example, we advise staying away from the language "approached the level of statistical significance" or "trending toward statistical significance," as it is suggestive of a direction.[3] The direction in which the *P* value could move given more data (ie, increase or decrease) is unknown. You may keep it simple by referring to these results as "near statistical significance." Or, if the sample size was too small, you can state that there was a difference that did not reach statistical significance and hence should be examined further in larger, more robust studies.[1,5]

Your interpretation of the results should take into account the limitations and weaknesses of the study design.[3,5]

Effect Sizes & Confidence Intervals & Significance, Oh My!

In summary, all-encompassing solutions do not come easy. If the connection between these statistics seems a little overwhelming or vague, we assure you that you are not alone in that feeling. Ideally, you will have, at a minimum, consulted with a biostatistician early on to determine which analyses are proper for the data you plan to collect as a part of your study design. Include a biostatistician if there are any major study modifications, and reach out to a biostatistician when drafting your manuscript to ensure not only accurate interpreting but also reporting of results. Their expertise can be of service throughout the duration of a study, and we are here to help your research succeed!

Acknowledgments

Many thanks to Advocate Health's Midwest region biostatistician workgroup, which meets monthly, for their insightful dialogue and input on this topic.

Conflicts of Interest

None.

References

1. Cumming G. *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis.* Routledge; 2012.
2. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat.* 2016;70:129-33. CrossRef
3. Christiansen S, Iverson C, Flanagin A, et al. *AMA Manual of Style: A Guide for Authors and Editors, 11th Edition.* Oxford University Press; 2020. CrossRef
4. Madjarova SJ, Williams RJ 3rd, Nwachukwu BU, et al. Picking apart p values: common problems and points of confusion. *Knee Surg Sports Traumatol Arthrosc.* 2022;30:3245-8. CrossRef
5. American Psychological Association. *Publication Manual of the American Psychological Association, Seventh Edition.* American Psychological Association; 2020.