Editorial Commentary: The Power of Interpretation: Utilizing the *P* Value as a Spectrum, in Addition to Effect Size, Will Lead to Accurate Presentation of Results

Payam W. Sabetian, M.D., and Benjamin G. Domb, M.D., Editorial Board

Abstract: Statistics have helped develop evidence-based medicine. Comparing groups and rejecting (or not) a null hypothesis is a main principle of the scientific method. Many studies have demonstrated that drawing conclusions based on the statistical result of a dichotomic *P* value instead of a spectrum can mislead us to conclude that there is "no difference" between two groups, or two treatments. In addition to the *P* value, the utilization of effect size (magnitude of difference between studied groups), may help us obtain a better global understanding of the statement "no effect". Although statistical significance does not mean clinical significance, by learning to adequately interpret data, we can disclose transparent results and conclusions, as we ward off our own bias. After all, without appropriate interpretation, we may be blinded from the truth.

See related article on page 1315

The ancient Hindu parable of The Blind Men and an Elephant,[1] describes a group of blind men who have never come across an elephant before, and who learn and imagine what an elephant is by touching it. However, since each blind man feels a different part of the elephant's body, they only learn to describe it as the part of the elephant they have felt. Since their descriptions widely vary from each individual's experience, they come to suspect that the other person is dishonest, and they come to a heated argument, only to later realize they were biased to their partial understanding of the truth. The parable may be applied to medical research, as our individual findings of a statistically significant difference in the treatment to a medical problem may seem an absolute truth. Nevertheless, to fully understand the "elephant" requires us to truly learn to interpret our findings, as well as those of others.

Statistics have helped medical research from experience-based opinions to evidence-based medicine. Comparing groups and rejecting—or not—a null hypothesis is a main principle of the scientific method

Chicago, Illinois

The authors report the following potential conflicts of interest or sources of funding: B.G.D. reports grants and other from American Orthopedic Foundation, during the conduct of the study; personal fees from Adventist Hinsdale Hospital, personal fees and nonfinancial support from Amplitude, grants, personal fees and non-financial support from Arthrex, personal fees and non-financial support from DJO Global, grants from Kaufman Foundation, grants, personal fees and non-financial support from Medacta, grants, personal fees, nonfinancial support and other from Pacira Pharmaceuticals, grants, personal fees, nonfinancial support and other from Stryker, grants from Breg, personal fees from Orthomerica, grants, personal fees, nonfinancial support and other from Mako Surgical Corp, grants and nonfinancial support from Medwest Associates, grants from ATI Physical Therapy, grants, personal fees and nonfinancial support from St. Alexius Medical Center, grants from Ossur, outside the submitted work. In addition, Dr. Domb has a patent 8920497 - Method and instrumentation for acetabular labrum reconstruction with royalties paid to Arthrex, a patent 8708941 - Adjustable multi-component hip orthosis with royalties paid to Orthomerica and DJO Global, and a patent 9737292 - Knotless suture anchors and methods of tissue repair with royalties paid to Arthrex. Also, Dr. Domb is the Medical Director of Hip Preservation at St. Alexius Medical Center, a board member for the American Hip Institute Research Foundation, AANA Learning Center Committee, the Journal of Hip Preservation Surgery, the Journal of Arthroscopy; has HAD ownership interests in the American Hip Institute, Hinsdale Orthopedic Associates, Hinsdale Orthopedic Imaging, SCD#3, North Shore Surgical Suites, and Munster Specialty Surgery Center. Full ICMJE author disclosure forms are available for this article online, as supplementary material.

0749-8063/211466/$36.00
https://doi.org/10.1016/j.arthro.2021.10.009

(frequentist approach). Despite its rigors, the scientific method presents the risk of rejecting a true hypothesis (type 1 error) or failing to reject a false one (type 2 error), based on the subjective threshold given by the researcher of the probability (*P*) and power.[2,3] Interpretation based on the statistical result of a dichotomic *P* value instead of a spectrum may, therefore, mislead us to conclude that there is "no difference" between two groups, or two treatments.

In the article 'No Effect' Conclusions in Studies Reporting Nonsignificant Results Are Potentially Incorrect",[4] authors Uimonen, Ponkilainen, Raittio, and Reito evaluate the sizes of observed effects in 'no effect' statements on high-impact, orthopedic journals. A total of 255 articles were reviewed, from which 18% were randomized controlled trials. Cohen's d value, Phi and odds ratios, and hazard ratios were used to calculate the effect sizes. The asymmetry ratio was assessed and averaged 1.9 in all studies; however, in 22% of them, it exceeded 5. This suggests that although these studies did not show statistical significance, there may have.

This well-designed study shows us how the interpretation of our results is certainly the most valuable part of our research. If we consider that we give our studies designated limits for type I and type II errors, making conclusions based solely on these values may cause us to write off treatments that do not meet the threshold, but they could, in fact, have an effect.[5,6] With their research, the authors highlight the importance of the use of effect size, which is the magnitude of the difference between our studied groups.[7,8] In reality, effect size is the main statistical finding of a quantitative study and should be reported in addition to *P*, as clearly demonstrated.

We commend the authors on a very analytical study that calls us to reflect on the way we interpret and present our results by filtering them through the *P* value. We agree that it should be accompanied by confidence interval and effect size, in addition to clinically reported outcomes. Lest we forget, statistical significance is not equal to clinical effect, for which a Bayesian approach may provide the answer.[9] By improving our knowledge of statistics, we gain a powerful tool to interpret data, we disclose transparent results and conclusions, and we ward off our own bias. After all, describing only a part of the "elephant" may blind us to the whole truth.

References

1. The Blind Men and the Elephant, https://americanliterature.com/author/james-baldwin/short-story/the-blind-men-and-the-elephant. Accessed October 8, 2021.
2. Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research pearls: The significance of statistics and perils of pooling. Part 1: Clinical versus statistical significance. *Arthroscopy* 2017;33:1102-1112.
3. Cote MP, Lubowitz JH, Brand JC, Rossi MJ. Misinterpretation of *P* values and statistical power creates a false sense of certainty: Statistical significance, lack of significance, and the uncertainty challenge. *Arthroscopy* 2021;37:1057-1063.
4. Uimonen M, Ponkilainen V, Raittio L, Reito A. 'No effect' conclusions in studies reporting nonsignificant results are potentially incorrect. *Arthroscopy* 2022;38:1315-1323.
5. Szucs D, Ioannidis JPA. When null hypothesis significance testing is unsuitable for research: A reassessment. *Front Hum Neurosci* 2017;11:390.
6. Domb BG, Sabetian PW. The blight of the type II error: When no difference does not mean no difference. *Arthroscopy* 2021;37:1353-1356.
7. Sullivan GM, Feinn R. Using effect size—or why the *P* value is not enough. *J Grad Med Educ* 2012;4:279-282.
8. Rosenthal R, Rubin DB. The counternull value of an effect size: A new statistic. *Psychol Sci* 1994;5:329-334.
9. Hohmann E, Wetzler MJ, D'Agostino RB. Research pearls: The significance of statistics and perils of pooling. Part 2: Predictive modeling. *Arthroscopy* 2017;33:1423-1432.