

P value, it is just not enough

Boohwi Hong^{1,2}

¹Department of Anesthesiology and Pain Medicine, Chungnam National University Hospital, Daejeon, Korea

²Department of Anesthesiology and Pain Medicine, College of Medicine, Chungnam National University, Daejeon, Korea

Received June 7, 2023; Revised June 12, 2023; Accepted June 12, 2023

Handling Editor: Francis S. Nahm

Correspondence: Boohwi Hong

Department of Anaesthesiology and Pain Medicine, Chungnam National University Hospital, 282 Munhwa-ro, Jung-gu, Daejeon 35015, Korea

Tel: +82-42-280-7840, Fax: +82-42-280-7968, E-mail: koho0127@gmail.com

According to the instructions for authors in The Korean Journal of Pain (KJP), it is recommended to provide an effect size and its corresponding estimates. Additionally, it suggests being cautious about reporting *P* values alone.

“Confidence intervals or effect sizes should be presented with *P* values. *P* values should not be presented alone and should be presented with confidence intervals.”

Furthermore, in the consolidated standards of reporting trials (CONSORT) statement, which serves as the reporting guidelines for randomized controlled trials (RCTs), the following reporting items are recommended to be included [1].

“For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval). For binary outcomes, presentation of both absolute and relative effect sizes is recommended.”

From 2020 onwards, a search in the KJP archive with the terms "randomized" or "randomised" in the title yielded a total of 23 RCTs (**Supplementary Table 1**). Among these studies, only three reported the effect size and confidence interval (CI) for the primary outcome [2-4]. In some studies, only the effect size according to the measurement time points within the groups was reported without the effect size and CI between the groups.

A *P* value is a statistical measure employed in null hypothesis significance testing to indicate the likelihood of obtaining the observed results assuming the null hypothesis is true. It also serves as a criterion for accepting or rejecting the alternative hypothesis, based on a predetermined threshold. However, this binary decision-making process has created an obsession and misuse of *P* values, often treating a *P* value < 0.05 as a guarantee that the hypothesis being tested is true [5]. Paradoxically, this practice has led to an increase in false positive studies and raised concerns about the reproducibility of research findings. As a result, there is a growing consensus advocating for a reduction in the threshold for statistical significance [6].

However, it is important to note that a *P* value < 0.05 (in general) does not represent the probability of the alternative hypothesis being true. Moreover, the *P* value itself does not provide information about the extent of the observed difference, regardless of the chosen threshold for significance. Therefore, it is not appropriate to determine the significance of the research results solely based on the magnitude of the *P* value. In 2016, the American Statistical Association released principles that addressed the appropriate use of *P* values [7]. Among them, I believe the following sentence carries significant implications.

“A *P*-value, or statistical significance, does not mea-



sure the size of an effect or the importance of a result.”

The limitations of the *P* value can be overcome by providing the effect size and its estimated CI, which can more effectively express research findings [8,9].

Furthermore, clinical significance in research findings can be evaluated by establishing a “minimally clinically important difference (MCID)” prior to the study commencement and comparing the CI of the effect size to that predetermined difference [10]. The term MCID refers to the smallest change in a clinical outcome that patients perceive as meaningful and significant [11]. The emphasis is on recognizing clinically important or meaningful changes from the patient's perspective, rather than simply focusing on statistically significant differences. For example, if the MCID for a pain score is determined to be ‘1’ based on prior evidence [12], and the lower limit of the CI in the study is greater than ‘1’, it indicates that the study findings are not only statistically significant but also hold clinical significance. In well-designed RCTs, readers should be able to assess the clinical significance of the research findings independently of statistical significance, and the effect size and CI can play a role in facilitating this assessment [13].

Although readers can calculate the effect size directly using the provided data (mean, standard deviation, and sample size for continuous outcome, or a contingency table for binary outcome), providing the effect size can alleviate this burden for readers. Moreover, it enhances the clarity of research findings, promotes transparency, and facilitates a better understanding of the study's impact.

Given these considerations, I strongly advocate for a more robust and explicit recommendation regarding reporting effect size measures and their corresponding estimates.

DATA AVAILABILITY

Data sharing is not applicable to this article as no datasets were generated or analyzed for this paper.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

FUNDING

No funding to declare.

AUTHOR CONTRIBUTIONS

Boohwi Hong: Writing/manuscript preparation.

ORCID

Boohwi Hong, <https://orcid.org/0000-0003-2468-9271>

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via <https://doi.org/10.3344/kjp.23173>.

REFERENCES

1. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.
2. Govil N, Parag K, Arora P, Khandelwal H, Singh A; Ruchi. Perioperative duloxetine as part of a multimodal analgesia regime reduces postoperative pain in lumbar canal stenosis surgery: a randomized, triple blind, and placebo-controlled trial. *Korean J Pain* 2020; 33: 40-7.
3. Jeong H, Choi JW, Sim WS, Kim DK, Bang YJ, Park S, et al. Ultrasound-guided erector spinae plane block for pain management after gastrectomy: a randomized, single-blinded, controlled trial. *Korean J Pain* 2022; 35: 303-10.
4. Lee GY, Lee JW, Lee E, Yeom JS, Kim KJ, Shin HI, et al. Evaluation of the efficacy and safety of epidural steroid injection using a nonparticulate steroid, dexamethasone or betamethasone: a double-blind, randomized, crossover, clinical trial. *Korean J Pain* 2022; 35: 336-44.
5. Hadjipavlou G, Siviter R, Feix B. What is the true worth of a P-value? Time for a change. *Br J Anaesth* 2021; 126: 564-7.
6. Chuang Z, Martin J, Shapiro J, Nguyen D, Neocleous P, Jones PM. Minimum false-positive risk of primary outcomes and impact of reducing nominal P-value

- threshold from 0.05 to 0.005 in anaesthesiology randomised clinical trials: a cross-sectional study. *Br J Anaesth* 2023; 130: 412-20.
7. Wasserstein RL, Lazar NA. The ASA statement on *p*-values: context, process, and purpose. *Am Stat* 2016; 70: 129-33.
 8. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol* 2016; 69: 555-62.
 9. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. *J Grad Med Educ* 2012; 4: 279-82.
 10. Lee S. Avoiding negative reviewer comments: common statistical errors in anesthesia journals. *Korean J Anesthesiol* 2016; 69: 219-26.
 11. Muñoz-Leyva F, El-Boghdadly K, Chan V. Is the minimal clinically important difference (MCID) in acute pain a good measure of analgesic efficacy in regional anesthesia? *Reg Anesth Pain Med* 2020; 45: 1000-5.
 12. Myles PS, Myles DB, Gallagher W, Boyd D, Chew C, MacDonald N, et al. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. *Br J Anaesth* 2017; 118: 424-9.
 13. Park S. Significant results: statistical or clinical? *Korean J Anesthesiol* 2016; 69: 121-5.