Commentary: Confidence intervals are only one piece of the puzzle

David Zurakowski, PhD, MS, and Steven J. Staffa, MS

In the Expert Opinion article published in the April 2021 issue of The Journal of Thoracic and Cardiovascular Surgery entitled “Moving Beyond Significance Testing: Confidence Intervals in Clinical Research,” Visintainer⁰ recommends increasing the use of point estimates and confidence intervals to help avoid misinterpretation of $P$ values in isolation. Along with avoiding misguided inferences, Visintainer astutely states that providing point estimates and confidence intervals can lead to more “critically and clinically relevant discussions of study results.” As a commentary written in light of recent articles regarding $P$ values published in the New England Journal of Medicine¹ and Nature², among other journals, Visintainer’s Expert Opinion article³ expands the discussion by making a call for increased use of confidence intervals.

Relevant to surgical research, Visintainer⁰ discusses how the width of confidence intervals, in other words the variability around the effect estimate or point estimate (eg, odds ratio, difference in proportions, hazard ratio) needs to be considered for proper interpretation of study results. The $P$ value is unable to give this information by itself, and the $P < .05$ significance threshold is arbitrary.⁴ Furthermore, regardless of statistical significance, the width of confidence intervals can vary substantially. Of course, the width of a confidence interval depends heavily on sample size, with a larger $N$ leading to a narrower confidence interval. Visintainer⁰ discusses the advantages of narrower confidence intervals in light of clinical interpretation of study results.

Confidence intervals, however, are only one piece of the puzzle required for getting thoracic and cardiovascular surgeons beyond complete reliance on $P$ values. There are many pieces of the puzzle that need to be incorporated to achieve more useful and impactful publications (Figure 1). The research question and study plan should be clearly and deliberately specified. As part of the design phase, surgeons must state the relevant primary and secondary outcomes, as well as the corresponding statistics that will be calculated. A power analysis and sample size justification should be performed before the study to ensure sufficient power for detecting clinically significant effects. Along with reporting confidence intervals or other estimates of variability, it is important to ensure that study results are thoughtfully interpreted in the context of both statistical and clinical significance. Statistical point estimates and associated confidence intervals need to be translated into language that has clinical meaning in cardiovascular surgery.

Reporting confidence intervals is a definite step in the right direction for improving the quality of surgical research; however, readers of the Journal are accustomed to seeing and often focusing only on $P$ values. All the pieces of the puzzle are essential and highly relevant for successful research. Therefore, following a strategy that incorporates these elements will lead to more insightful interpretations of the data and thus more impactful articles. When reporting $P$ values, we urge researchers to do so judiciously and never in isolation, because, as Visintainer⁰ has stated, the $P$ value by itself may lead to faulty interpretations and misguided inferences.

From the Departments of Surgery and Anesthesiology, Critical Care and Pain Medicine, Boston Children’s Hospital, Harvard Medical School, Boston, Mass. Disclosures: Authors have nothing to disclose with regard to commercial support. Received for publication Feb 5, 2020; revisions received Feb 6, 2020; accepted for publication Feb 7, 2020; available ahead of print Feb 14, 2020. Address for reprints: David Zurakowski, PhD, MS, Department of Surgery, Boston Children’s Hospital, 300 Longwood Ave, Boston, MA 02115 (E-mail: David.Zurakowski@childrens.harvard.edu). J Thorac Cardiovasc Surg 2022;164:e37-8 0022-5223/36.00 Copyright © 2020 by The American Association for Thoracic Surgery https://doi.org/10.1016/j.jtcvs.2020.02.029

CENTRAL MESSAGE
Confidence intervals are important for quantifying the precision of effect estimates, but they are only one piece of the puzzle in carrying out and achieving impactful surgical research.
FIGURE 1. The surgical research puzzle involves multiple pieces pertaining to a priori study design (blue), comprehensive and thorough data analysis (red), and clear and insightful presentation to highlight and communicate the most important findings (green). The puzzle represents the cornerstone of good surgical research and requires a collaborative effort between surgeons and statisticians. All the pieces of the puzzle are important for successful research, and incorporating these elements will lead to more insightful interpretations of the data and more impactful articles in the Journal.

References