

Commentary



The difference between statistical and clinical significance is subtle but important. A result that we identify as statistically significant (usually defined at the 5% level, with $P \leq 0.05$) may not be clinically significant (usually defined as an effect size, or difference between treatments, of $\geq 10\%$). The P-value, though readily attainable through any standard statistical software program, may not be very meaningful in real life. What we really want to know is the probability that an observed outcome is of clinical or practical significance. To this end, we have to use our experience and understanding of a clinical situation to define a threshold of clinical significance, which is often defined as a difference $\geq 10\%$. An alternative to the P-value is the confidence interval, which defines a range, in the same terms by which the data were measured, within which the true value probably (95% probability) resides. The confidence interval also provides information about statistical significance, the strength and direction of the effect, and enables us to consider the clinical relevance of the outcome. To make these points clear, authors and editors should make an effort to report confidence intervals about their point estimates. In the following editorial, Dr. Joseph explains the importance of reporting confidence intervals in the context of health care reform.

D. Scot Malay, DPM, MSCE, FCFAS
Editor
The Journal of Foot & Ankle Surgery

Health Care Reform, Statistical Significance, Effect Size, and the Future of the Profession

Health care reform is transforming the medical landscape for clinicians, patients, and administrators alike, with initiatives that strive to maximize health care access and quality while containing expense. The theme of “reform culture” is that of evolving processes of measurement, analysis, and change that are not unique to medicine but rather are a component of a new global economy. Research and analytics have never been so pervasive throughout industries ranging from communications and manufacturing to finance. Outside the health care sector, the vernacular of Six Sigma, Fair Market Valuation, and even Google Analytics are all hallmarks of a culture geared toward analytics. In health care, we have come to recognize this culture in the form of Pay for Performance, Physician Quality Reporting Initiative (PQRI), and Evidence-Based Medicine. No matter what the industry, society now expects greater clarity, transparency, justification, and accountability through analytics.

Reform culture has had an impact on all aspects of the profession of foot and ankle surgery, from the way we evaluate patients

and render care, to the finer details of practice management. Although the means of health care reform are controversial, the unifying message is that reform must be justified by *significant* discrepancies in our health system that can be changed to cause the *greatest impact* on improving the balance between health care quality, access, and cost. As a college of surgeons, the concept of *significance* is well known to us as we use statistical calculations every day to report the *P* values described in our research. We are comfortable with this statistical measurement. At its core, the *P* value merely tells us that a measured difference between 2 treatments was probably attributable to the intervention rather than chance. The new demands of reform, however, call for accountability beyond a simple claim of statistical significance. To meet this higher degree of accountability, a more meaningful and robust measurement is needed.

Outcomes research is one of the fundamental ways by which we demonstrate the importance of our interventions. *Large effects* are a critical component in establishing the practical significance and



potential impact of our clinical care, yet P values alone fall short in demonstrating the impact of a particular treatment. The problem with P values alone is twofold: they may obscure clinically important findings that are nonetheless statistically insignificant, and they may overemphasize the importance of effects that are statistically significant but of little clinical value. This is because a P value does not convey whether statistical significance is a result of a large

sample size, small data variance, or (most importantly) a large effect. An example would be a randomized controlled trial that reports that pain medication A displayed a duration of pain control that was statistically significantly ($P = .002$) longer than that observed with medication B . Although statistically significant, the P value in this example does not reveal whether the average duration of pain control achieved with medication A was 12 hours and that with medication B was 13 hours, a difference that many clinicians would not consider clinically significant or economically justified if medication A was exorbitantly more expensive than medication B . Scenarios like this exemplify the limitations of the P value in regard to describing the research in practical terms that can be useful to clinicians and, as such, potentially dilute the value of our outcomes research.

In an effort to overcome the shortcomings associated with reporting only the P value, we must transition toward using research measures that more clearly explain not just statistical significance, but the magnitude and direction of an effect, as well. Confidence intervals do this, yet they are rarely used to validate orthopedic and podiatric research. A recent study estimated that only 22% of orthopedic studies reported confidence intervals in research validation (Vavken P, Heinrich KM, Koppelhuber C, Rois S, Dorotka R. The use of confidence intervals in reporting orthopaedic research findings. *Clin Orthop Relat Res* 467:3334–9, 2009). The same study also reported that the probability that a study reporting statistically significant findings would predict a 10% or greater difference between comparison groups was only 69%. These estimates suggest that there may be substantial challenges in the way that we apply research findings to real life clinical situations. These challenges may be partly overcome by more diligent use of confidence intervals in the analyses that we undertake to validate the treatments that we recommend and provide. Confidence intervals provide a graphical and numerical representation of data that show statistical significance and the magnitude of effects not portrayed by P values alone. To meet the increasing demands of an analytic economy, we must adopt instruments that more clearly demonstrate the impact that foot and ankle surgery has on our nation's health. As foot and ankle surgery evolves, so will our need to use more powerful analytical methods such as confidence intervals in an effort to make our research results more relevant to clinical care.

Robert M. Joseph, DPM, PhD