Ibrahim Card Med J 2025; 15 (1): 05-06

EDITORIAL

The "So What" Factor: Navigating the Divide Between Statistical Significance and Clinical Reality

Md. Nurul Amin[1]

In the modern landscape of evidence-based medicine, a long-standing "alphabet soup" has blurred the lines between mathematical probability and medical relevance.[1] As Einstein is often credited with saying, "Not everything that can be counted counts, and not everything that counts can be counted." This sentiment is the core of the debate surrounding statistical versus clinical significance-a tension recently highlighted by many clinical researchers in their studies.[1-4]

The P-Value: A Reliable but Narrow Metric

Statistical significance, typically defined by a p-value < 0.05, is merely a statement about the likelihood that a finding occurred by chance.[1,3] In Amin's study, "Impact of Cellphone on Human Health," local temperature increased significantly (97.4 ± 0.7°F to 98.0 ± 0.8°F, p < 0.001) with a large effect size (d = 0.80) following a median talk-time of 20 minutes over a mobile-phone. Pulse rate also showed a statistically significant increase (89 ± 17 to 92 ± 17 bpm, p = 0.009), but with a negligible effect size (d = 0.18). However, no linear correlation was observed between talk-time and either local temperature or pulse rate.[2] While these pulse "count" as statistically significant, they represent a minute physiological shift. As Fethney[1] and Ranganathan et al.[3] argue, p-values are sample-specific estimates that offer no information regarding the magnitude of the effect or its future impact on the wider population.

A p-value of 0.009 tells us the result is unlikely to occur by chance, but it cannot answer the "so what?" question. Is a 3-beat increase in pulse clinically meaningful? To answer this, we must look beyond the p-value to the effect-size and the minimum important difference (MID).

The "Sample Size Fallacy" and Type I Errors

A common pitfall in research is the "sample size fallacy." As noted in the medical literature, too large a sample may lead researchers to report a difference that exists mathematically but not biologically-a variation of the Type I error where "significance" is an artifact of the study's power rather than its practical implication for clinical practice.[5]

Amin's study utilized a within-sample pre-test/post-test design with 118 participants, which virtually enehnced the statistical power.[2] While this rigour is commendable, it can lead to "false positive" clinical impressions regarding trivial variations. This mirrors the "REVEL" trial discussed by Dahlberg et al., where an enormous sample of 1,253 patients resulted in a statistically significant survival benefit of only 1.4 months-a result that many oncologists deemed a "clinically irrelevant improvement" despite its p-value of 0.023.[3,6]

Effect Size: The Bridge to Clinical Relevance

Clinical significance reflects the practical value of a treatment or exposure-specifically, whether it makes a real difference to subject lives.[3,4] One robust way to quantify this is through Cohen's d. Amin's research found a "negligible" effect size for pulse rate (d = 0.18) despite its statistical significance.[2] In contrast, local temperature over parotid gland showed both statistical significance (p < 0.001) and a "large" effect size (d = 0.80).[2]

Author's information:

[1]***Dr. Md. Nurul Amin,*** *PhD. Senior Fellow International Diabetes Federation (IDF), Associate Professor (Research) & Executive Editor (Ibrahim Cardiac Medical Journal), Ibrahim Cardiac Hospital & Research Institute, Shahbag, Dhaka, Bangladesh.*

Correspondence: ***Dr. Md. Nurul Amin,*** *Mobile: 01753178452, E-mail: mdamin01@yahoo.com*

As Heavey notes, a statistically significant result that is not clinically significant is often not clinically meaningful.[5] In medical science and general practice, experts must determine if a finding is substantial enough to direct patient care. For instance, a trivial blood pressure drop of 3 mmHg might be statistically significant due to a large sample size but clinically "inconsequential" if it does not equate to improved outcomes for the patient.[1]

The Role of Confidence Intervals

A primary limitation of the p-value is its dichotomous (yes or no) nature.[1,6] To bridge this gap, experts advocate for the reporting of Confidence Intervals (CIs).[1,3,5,6] CIs provide a plausible range of values for the "true" population parameter. If a 95% CI includes zero (in a difference of means), the result is statistically non-significant. However, even if the CI excludes zero, we must check if the lower limit meets the MID-the smallest difference that would actually be of interest to a clinician.[1] In Amin's study, the lack of linear correlation between talk-time and hemodynamic variables (r = 0.007) suggests that while the "average" effect may be caught by a p-value, the "true" clinical impact varies wildly between individuals possibly due to differences in radiofrequency emitting rate, measured in terms specific absorption rate (W/kg), from different brands of mobile.[2]

CONCLUSION

While statistical significance follows rigid mathematical rules, clinical significance remains a subjective judgment made by clinicians and patients.[3,5] It involves weighing the extent of change against patient acceptability, cost-effectiveness, and safety.[1,4]

The studies referenced and discussed hitherto serves as a vital reminder that researchers must move away from reflexive reliance on p-values Clearly, we must stop treating the p-value as a binary "pass/fail" grade. By defining MIDs and reporting effect sizes, we ensure that our research does not just "count" in a statistical table, but truly "counts" for human health. Identifying clinical importance is what clinicians are ultimately aiming for; to halt interpretation at the p-value is to afford it more meaning than it deserves.

Researchers should move toward the standard use of confidence intervals (CIs) and effect sizes (Cohen's d) to provide a range of certainty and a measure of impact. In clinical trials a result can be "fantastic" in an analytical software but "disappointing" when applied to a patient. If we are to truly improve human health, our definition of "significance" must return to its literal meaning: something that is truly important to the patient

REFERENCES

1. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both. *Aust Crit Care* 2010;23(2):93-7. doi: 10.1016/j.aucc.2010.03.001.
2. Amin MN, Islam MSU, Haque MJ Association between Mobile Phone Use and Physical Wellbeing. *Ibrahim Card Med J* 2021;112:8-15 DOI:https://doiorg/103329/icmjv11i266572
3. Ranganathan P, Pramesh CS, Buyse M Common pitfalls in statistical analysis: Clinical versus statistical significance. *Perspect Clin Res* 2015;63:169-170.
4. LeFort SM. The statistical versus clinical significance debate. *Image J Nurs Sch* 1993;25(1):57-62. doi: 10.1111/j.1547-5069.1993.tb00754.x
5. Heavey E. Differentiating statistical significance and clinical significance. *Nursing* 2015;455:26-28.
6. Dahlberg SE, Korn EL, Le-Rademacher J, Mandrekar SJ. Clinical Versus Statistical Significance in Studies of Thoracic Malignancies. *J Thorac Oncol* 2020;159: 1406-1408.