Redefining treatment success: Shifting from p-value to clinical meaningfulness

"The greatest obstacle to discovery is not ignorance – it is the illusion of knowledge."

Daniel J Boorstin, 1984

For the longest time, researchers and clinicians have relied on statistical significance to assess treatment efficacy and guide clinical decisions. Conventionally, a p-value <0.05 is taken as 'proof' of treatment efficacy. But this perception, while widespread, is deeply flawed. Our obsession with p-values has contributed to much of the published scientific research being not reproducible in real life. After all, statistical significance is not the same as clinical significance.

Imagine a trial (n=100) evaluating a new treatment for vitiligo. The treatment reduces the baseline Vitiligo area scoring index (VASI) from 35% to 25% (p<0.001). While this change may be statistically significant, a 10% repigmentation may be too little to be seen as clinically meaningful by the patients or physicians. Nonetheless, this often gets interpreted as the treatment being effective (when it clearly is not!). The p-value does not comment on the magnitude of the treatment effect (how good is the treatment; its effect size).[1,2] Moreover, p-values can only be applied to a group of patients, and not to an individual patient in the clinic, limiting their utility in the real-life setting.

So, how should physicians analyse research to inform clinical practice? For starters, they should consider not just the p-values (is the effect due to treatment?), but also the effect size (how much is the treatment effect?) and its confidence intervals (how precise are the results?).[2] A statistically significant result with a small effect size might not be clinically meaningful. Take a large enough sample size, and even tiny changes may become statistically significant. Conversely, a large effect size even in the absence of statistical significance (as may happen in an underpowered study) could still be meaningful, and should not be dismissed outright.

But how large should the effect size be for it to be clinically relevant? Cohen classified effect size as small (d=0.2), medium (d=0.5), and large (d ≥0.8),[3] but these cut-offs are arbitrary and not without limitations.[4]

A better way to assess clinical relevance may be to view it through the lens of a patient-centered metric. That is where the concept of minimal important difference (MID) comes in. MID refers to the smallest change in the disease severity that is considered meaningful by the patients (or the clinicians). For instance, a 4-point decrease in the Dermatology Life Quality Index (DLQI, a skin disease-related quality-of-life instrument) represents a clinically meaningful improvement for patients with inflammatory skin diseases, a signal that the treatment is working and is worth continuing.[5] However, treatments are not aimed at producing only the smallest perceptible change, but rather a large one. In this context, MID may be considered too low a bar for determining treatment success, and alternative thresholds such as 'substantial clinical benefit' (SCB) may be more appropriate.[6] For DLQI, an 8-point change (i.e., two times the MID) cut-off has been proposed as its SCB cut-off, a level at which patients are likely to report a significant improvement.[7] These thresholds are not just useful at the individual level; they are well-suited for comparing treatments in clinical trials as well.

Some such thresholds are already being used in dermatology. PASI75 (and now PASI90) in psoriasis has been traditionally used as a clinical benchmark, while others, such as SALT50 (alopecia areata), EASI75 (atopic dermatitis), and HiSCR50 (hidradenitis suppurativa), are now being increasingly used. But instead of relying on what are perhaps intuitively chosen cut-offs (like 50% or 75% reduction), we should strive to define meaningful thresholds based on patient-perceived improvement. This can be achieved by anchoring outcome measures to patient-reported changes, through Likert scales during clinical trials or focused patient interviews. Recently, a multiple anchors-approach estimated that improvements of 30% in T-VASI and 50% in F-VASI scores reflected clinically meaningful repigmentation in patients with vitiligo. Interestingly, these cut-offs are lower than those historically used (T-VASI50, F-VASI75) in clinical trials.[8] Looking back at our earlier example of the vitiligo trial, would the treatment still be considered effective if only 8% of patients achieve F-VASI75?

Editors, reviewers, and researchers should prioritise reporting outcomes in terms of clinically meaningful benefit (effect size, MID, SCB) in addition to, and not just, statistical significance (p-value). As we move toward a more personalised and patient-centered era in dermatology, it is essential that the way we interpret our data evolves as well. Clinical trials must tell us not only whether a treatment works, but whether it works enough to matter.

Vishal Gupta[1]

[1]Department of Dermatology and Venereology,
All India Institute of Medical Sciences,
New Delhi, India

Corresponding author:
Dr. Vishal Gupta,
Associate Editor, IJDVL
Department of Dermatology and Venereology,
All India Institute of Medical Sciences, New Delhi, India.
doctor.vishalgupta@gmail.com

References

1. Nahm FS. What the P values really tell us. Korean J Pain 2017;30:241-2.
2. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. J Grad Med Educ 2012;4:279-82.
3. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Lawrence Erlbaum Associates; 1988.
4. Lee DK. Alternatives to P value: Confidence interval and effect size. Korean J Anesthesiol 2016;69:555-62.
5. Basra MK, Salek MS, Camilleri L, Sturkey R, Finlay AY. Determining the minimal clinically important difference and responsiveness of the dermatology life quality index (DLQI): Further data. Dermatology 2015;230:27-33.
6. Rossi MJ, Brand JC, Lubowitz JH. Minimally clinically important difference (MCID) is a low bar. Arthroscopy 2023;39:139-41.
7. Ali FM, Salek MS, Finlay AY. Two minimal clinically important difference (2MCID): A new twist on an old concept. Acta Derm Venereol 2018;98:715-7.
8. Ezzedine K, Soliman AM, Camp HS, Ladd MK, Pokrzywinski R, Coyne KS, *et al*. Psychometric properties and meaningful change thresholds of the vitiligo area scoring index. JAMA Dermatol 2025;161:39-46.