Journal of Athletic Training 2022;57(11/12):1124–1125
doi: 10.4085/1062-6050-1007.22

www.natajournals.org

Minimal Clinically Important Differences Require Baseline Adjustments to Maintain Theoretical Viability

Dear Editor:

We thank Terluin et al for their letter as well as their extensive publication record about clinical thresholds. We were unaware of some of their publications, largely due to the disparate nomenclature that plagues this area of work (eg, minimal clinically important difference [MCID], clinically important difference [CID], minimal clinically important change [MCIC], clinically important change [CIC], minimal clinically important improvement [MCII], and minimal important difference [MID]). In our response, we will first detail an area in which our groups largely have mutual agreement. We will then show that the main thesis of Terluin et al's letter is incorrect based on simulation and empirical data. Finally, we will conclude by discussing some valuable points they raise.

The letter concludes with the statement, "ROC [receiver operating characteristic] analysis is not a good method for estimating the MCID," and we wholeheartedly agree. In fact, MCIDs and all their related clinical thresholds are highly flawed from any number of conceptual angles. Humans are multidimensional, and it does not make sense to use univariate thresholds to determine when someone is clinically better. A far better approach would be a multivariable model incorporating various dimensions and environmental factors.[1] Additionally, not all clinical thresholds are thresholds per se; they are point estimates abstracted from a population-based sample. It is unfortunate that the point estimate is often used with no regard for the confidence intervals around those estimates. For these reasons, we believe that MCIDs and related metrics do not effectively capture real-world dynamics in their current form.

The main thesis of the letter is that, although we have shown MCIDs to vary according to the baseline score (which we attributed to regression to the mean), they believe this variance in baseline score is an artifact due to the prevalence of the outcome or anchor (ie, proportion of improved patients). Here, we demonstrate that this thesis is incorrect based on simulation and empirical data. As shown here, using the same simulated model the authors used in their letter, the MCID varies according to the baseline score, both with a natural dataset (ie, unbalanced outcomes) and when that dataset is resampled to contain a 50/50 distribution of outcomes. We simulated the same sample from our Figure 1A as in the letter, only far larger to allow for resampling (n = 1 000 000), and then subdivided the sample into 4 subgroups with baseline scores of (1) $\geq$32 and $\leq$38, (2) $\geq$37 and $\leq$43, (3) $\geq$47 and $\leq$53, and (4) $\geq$52 and $\leq$58. We then calculated MCIDs for these subsamples on the natural data and after resampling the data (without replacement) such that there was a 50/50 split in the anchor (500 samples in each group, n = 1000). As seen in the Table, the MCID varies according to the baseline score; however, the prevalence of the outcome (ie, proportion improved) has no effect.

Furthermore, we have shown in a recent empirical data paper[1] that MCID point estimates vary according to baseline scores, and these data can be similarly resampled to a 50/50 outcome proportion to show that outcome proportion has no effect on MCID estimate (data available: https://osf.io/9ktsb/). Therefore, although we agree with Terluin et al that we, in our paper published in *JAT*, do not conclusively demonstrate that MCID estimates varying according to baseline score is attributable to regression to the mean, it does have the hallmarks of the phenomenon (varying by baseline and affected by data correlation), and it can be clearly shown that this is not simply an artifact of outcome prevalence.

One criticism of our initial *JAT* paper that Terluin et al highlight is that we modeled a ground-truth static MCID of 20, so the illustration that this MCID could vary according to the baseline score and was affected by correlation in the repeated-measures data was somehow flawed. This is a reasonable criticism. However, we would like to make 2 points:

1. The goal of the *JAT* manuscript was to didactically demonstrate a statistical principle to a clinical audience, not model a full system of data in its complete form. We appreciate the work that Terluin et al have done in the past, but the goal of their work was different from ours. Had we submitted the full simulation, it would not have been appropriate for the readership of *JAT*.
2. If we modeled the underlying data such that ground-truth MCID varied according to the baseline score, a valid criticism would have been that, of course, a baseline-adjusted MCID better represents the data than a static MCID. It would have been a self-fulfilling prophecy. The fact remains that we do not fully know the real data-generating process underlying patient-reported outcomes, so any simulation is based on a flawed, assumed reality.

In their letter, Terluin et al also make the peculiar decision to take a continuous variable, baseline score, and make it polychotomous. This step is entirely unnecessary, and the practice of unnecessarily dichotomizing continuous variables has been widely criticized in the statistical community.[2,3] We are unclear how this decision may have influenced their findings or interpretations in earlier work or in their current letter, but we do not advocate for this practice. A baseline-adjusted ROC used in our *JAT* manuscript maintains the continuous nature of the data

Table. The MCID Calculation is Affected By the Baseline Score But Not By the Proportion of Improved Patients

Baseline Score Subgroup	Mean Baseline Score	Proportion Improved	Receiver Operating Characteristic–Based MCID
≥32 and ≤38	36.3	0.94	22.9
≥32 and ≤38	36.4	0.50	22.2
≥37 and ≤43	40.7	0.78	21.2
≥37 and ≤43	40.9	0.50	21.6
≥47 and ≤53	49.3	0.22	18.7
≥47 and ≤53	49.1	0.50	19.5
≥52 and ≤58	53.7	0.06	17.0
≥52 and ≤58	53.5	0.50	17.0

Abbreviation: MCID, minimal clinically important difference.

and would be unbiased by this unnecessary dichotomization.

From a clinical standpoint, theoretical concerns exist with static MCIDs that do not account for baseline scores. For example, take a simple scale that ranges from 0 to 10 points for which 0 = *feeling terrible* and 10 = *feeling great*. If we assume that a static MCID for this scale is 3, how would a clinician interpret a patient entering a clinic with a baseline score of 8? Is a clinician supposed to believe that an 11 out of 10 on the scale is needed to reach MCID? Surely not. A baseline-adjusted MCID can account for this concern. When statisticians or epidemiologists develop clinimetrics, how they will be used or abused in practice must be considered. If clinicians decide they need to have a univariate MCID for clinical practice, baseline-adjusted MCIDs strike a reasonable balance between statistical and theoretical validity and ease of use.

Matthew S. Tenan, PhD, ATC
Rockefeller Neuroscience Institute
West Virginia University, Morgantown

Janet E. Simon, PhD, ATC
School of Applied Health Sciences and Wellness
Ohio University, Athens

REFERENCES

1. Boyer CW, Lee IE, Tenan MS. All MCIDs are wrong, but some may be useful. *J Orthop Sports Phys Ther*. 2022;52(6):401–407. doi:10.2519/jospt.2022.11193
2. Senn S. Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials. In: *Proceedings of the International Statistical Institute*. International Statistical Institute; 2005.
3. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat*. 2009;8(1):50–61. doi:10.1002/pst.331