Journal of Athletic Training 2022;57(11/12):1122–1123
doi: 10.4085/1062-6050-1006.22

www.natajournals.org

Minimal Clinically Important Difference Estimates Are Biased by Adjusting for Baseline Severity, Not by Regression to the Mean

Dear Editor:

Minimal clinically important difference (MCID) estimates are often used to interpret change scores from measurement instruments. Researchers debate how MCID values should be estimated. In a recent paper, Tenan et al[1] recommended adjusting for baseline severity in the analysis to avoid biased MCID estimates due to regression to the mean (RTM). They stated that anchored MCID estimation can be biased by RTM due to repeated measurements. They also stated that including baseline severity as a covariate in the analysis (the authors used baseline covariate-adjusted receiver operating characteristic [ROC] analysis) averts this bias. No proof or justification was offered to support these statements. In this letter, we argue that adjusting for baseline severity is bound to introduce bias, instead of warding it off.

Regression to the mean refers to change that occurs due to random fluctuations.[2] Following a relatively high (or low) observation of a randomly fluctuating construct (eg, physical fitness), a repeated measurement will likely demonstrate a more moderate observation. Extreme values tend to regress to the mean. Therefore, RTM expresses itself as a negative correlation between baseline and change scores.

However, such a negative correlation is not necessarily a sign of RTM. Real reasons may explain why more severely affected patients improve more than less severely affected patients, eg, a treatment might be more effective in more severely affected patients.

The MCID is the change score deemed a minimal improvement that is considered important (we limit the discussion to improvement). The assessment of the improvement is based on an external criterion, namely, the anchor, which is often a single question that asks patients to rate their perceived change. It is assumed that patients have their own minimally important change thresholds, and it seems reasonable to consider the mean of the individual thresholds as the MCID to be estimated.[3,4]

Tenan et al simulated several datasets (n = 5000 patients) consisting of a baseline score (T1; mean = 45 ± 20) and a follow-up score (T2; mean = 65 ± 20), with variable correlations between T1 and T2. A binary anchor variable was added based on a Bernoulli distribution such that a positive outcome was more likely when the T1 − T2 difference was ≥20. It should be noted that the authors thus simulated the *true MCID* (defined as the average individual minimally important change threshold) as 20, independent of the baseline score. Indeed, a standard ROC analysis, using the anchor as the state variable and the change score as the test variable, yielded 20 as the MCID estimate.

Next, because the authors believed that MCID estimates can be biased by RTM and that accounting for the baseline score avoids this bias, they performed baseline covariate-adjusted ROC analysis. This analysis resulted in MCID estimates that were correlated with the baseline score, dependent on the correlation between T1 and T2. At this point, the authors believed their adjusted results (showing baseline-dependent MCID estimates) to be more true than their unadjusted results (which reflected the baseline-independent MCID values they had actually simulated). Why the authors came to this conclusion is a mystery to us. However, we do understand why baseline adjustment may lead to baseline-dependent MCID estimates (and this has nothing to do with RTM).

To clarify what happens when adjusting for the baseline score, we repeated the simulation in a different way. We adjusted for the baseline score by performing standard ROC analyses on baseline-stratified subgroups. We simulated a sample, similar to the first sample of the authors, but 5 times larger (n = 25 000). Then we split the sample into 5 subgroups based on quintiles of the baseline score. The correlation between the baseline score and the follow-up score was 0.11, and the correlation between the baseline score and the change score was –0.66. The results of the subgroup analyses are shown in the Table.

Due to the stratification, subgroups 1 through 5 showed increasing mean baseline scores. The MCID estimates mirrored the results of the authors' analysis in their Figure 3A; lower baseline scores were associated with higher MCID estimates and vice versa. Because of the negative correlation between the baseline and change scores, the mean change score was higher in subgroups with lower baseline scores and lower in subgroups with higher baseline scores. Given the simulated minimal important change threshold of 20, this resulted in greater proportions of patients who improved in the subgroups with lower baseline scores and smaller proportions who improved in the subgroups with higher baseline scores.

The cause of this baseline dependency of the MCID lies in the fact that the optimal ROC cutoff point (which defines the MCID value) depends on the prevalence of the condition (presently, the proportion of improved patients).[5,6] The optimal ROC cutoff point (Youden criterion) is the cutoff that classifies improved and not-improved patients with the least (weighted) misclassification. In large samples with normally distributed scores, this cutoff is characterized by equality of sensitivity and specificity. However, if the prevalence increases, the sensitivity of a given cutoff increases while its specificity decreases, and the opposite occurs if the prevalence decreases.[5] Therefore,

Table. Minimal Clinically Important Difference Estimates in Subgroups Stratified by the Baseline Score

Subgroup	Mean Baseline Score	Mean Change Score	Proportion Improved	Receiver Operating Characteristic-Based Minimal Clinically Important Difference[a]
1	38.7	25.6	0.86	22.0
2	42.7	22.1	0.66	20.9
3	45.0	20.0	0.50	20.3
4	47.4	17.9	0.33	19.3
5	51.3	14.4	0.14	18.1
Total group	45.0	20.0	0.50	20.0

[a] Based on the Youden criterion.

in (sub)samples with greater proportions improved, the optimal ROC cutoff point will be higher, whereas in (sub)samples with smaller proportions improved, the optimal ROC cutoff point will be lower. Only if the proportion improved is 0.5 does the ROC-based MCID estimate reflect the true MCID (as shown in subgroup 3 in the Table).[7] In other words, the simulation study that Tenan et al performed actually showed that the ROC-based MCID depends on the proportion improved, even though the authors did not recognize it as such. In a previous paper,[7] we demonstrated this phenomenon extensively.

The bottom line is that, generally, ROC analysis is not a good method for estimating the MCID. Better methods include the adjusted predictive modeling method[7] and a method based on item response theory.[8] If one is concerned about the MCID being baseline dependent, simply stratifying on the baseline score or, for that matter, baseline covariate adjusting is not a good idea, but solutions do exist.[9]

Berend Terluin, MD, PhD
Caroline Terwee, PhD
Amsterdam University Medical Centers, the Netherlands

Iris Eekhout, PhD
TNO, Amsterdam, the Netherlands

REFERENCES

1. Tenan MS, Simon JE, Robins RJ, Lee I, Sheean AJ, Dickens JF. Anchored minimal clinically important difference metrics: considerations for bias and regression to the mean. *J Athl Train*. 2021;56(9):1042–1059. doi:10.4085/1062-6050-0368.20
2. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. 2005;34(1):215–220. doi:10.1093/ije/dyh299
3. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171–184. doi:10.1586/erp.11.9
4. Vanier A, Sebille V, Blanchin M, Hardouin JB. The minimal perceived change: a formal model of the responder definition according to the patient's meaning of change for patient-reported outcome data analysis and interpretation. *BMC Med Res Methodol*. 2021;21(1):128. doi:10.1186/s12874-021-01307-9
5. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16(9):981–991. doi:10.1002/(sici)1097-0258(19970515)16:9<981::aid-sim510>3.0.co;2-n
6. Terluin B, Griffiths P, van der Wouden JC, Ingelsrud LH, Terwee CB. Unlike ROC analysis, a new IRT method identified clinical thresholds unbiased by disease prevalence. *J Clin Epidemiol*. 2020;124:118–125. doi:10.1016/j.jclinepi.2020.05.008
7. Terluin B, Eekhout I, Terwee CB. The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. *J Clin Epidemiol*. 2017;83:90–100. doi:10.1016/j.jclinepi.2016.12.015
8. Bjorner J. Analysis of minimal important change through item response theory methods [abstract]. *Value Health*. 2019;22(3):S818. doi:10.1016/j.jval.2019.09.2220
9. Terluin B, Roos EM, Terwee CB, Thorlund JB, Ingelsrud LH. Assessing baseline dependency of anchor-based minimal important change (MIC): don't stratify on the baseline score! *Qual Life Res*. 2021;30(10):2773–2782. doi:10.1007/s11136-021-02886-2