

Anchored Minimal Clinically Important Difference Metrics: Considerations for Bias and Regression to the Mean

Matthew S. Tenan, PhD, ATC*†; Janet E. Simon, PhD, ATC‡§;
Richard J. Robins, MD||¶; Ian Lee, PT, DSc*; Andrew J. Sheean, MD#;
Jonathan F. Dickens, MD¶**††

*Defense Health Management Systems, Rosslyn, VA; †Optimum Performance Analytics Associates, Apex, NC; ‡College of Health Sciences and Professions and §Ohio Musculoskeletal and Neurological Institute, Ohio University, Athens; ||United States Air Force Academy Ambulatory Surgical Center, Colorado Springs; ¶Uniformed Services University of the Health Sciences, Bethesda, MD; #San Antonio Military Medical Center, Fort Sam Houston, TX; **Walter Reed National Military Medical Center, Bethesda, MD; ††John A. Feagin, Jr Orthopaedic Sports Medicine Fellowship, West Point, NY

Minimal clinically important differences (MCIDs) are used to understand clinical relevance. However, repeated observations produce biased analyses unless one accounts for baseline observation, known as *regression to the mean* (RTM). Using an International Knee Documentation Committee (IKDC) survey dataset, we can demonstrate the effect of RTM on MCID values by (1) MCID-estimate dependence on baseline observation and (2) MCID-estimate bias being higher when the posttest-pretest data correlation is lower. We created 10 IKDC datasets with 5000 patients and a specific correlation under both equal and unequal variances. For each 10-point increase in baseline IKDC, MCID decreased by 3.5, 2.7, 1.9, 1.2, and 0.7 points

when posttest-pretest correlations were 0.10, 0.25, 0.50, 0.75, and 0.90, respectively, under equal variances. Not accounting for RTM resulted in a static 20-point MCID. Minimal clinically important difference estimates may be unreliable. Minimal clinically important difference calculations should include the correlation and variances between posttest and pretest data, and researchers should consider using a baseline covariate-adjusted receiver operating characteristic curve analysis to calculate MCID.

Key Words: validity, patient-reported outcomes, statistics

In recent years, the difference between statistically different and clinically important changes in patient-reported outcomes (PROs) has become widely accepted, and it has become increasingly common for authors to report the minimal clinically important difference (MCID) and substantial clinical benefit (SCB) for a number of measures intended to quantify and characterize patients' responses to different interventions. Two types of MCID metrics are commonly used: distributional and anchored. Distribution-based MCIDs typically rely on the standard error of measurement, straight SD calculations from sample data, or the minimal detectable change¹; however, no information related to clinical relevance is contained in these distributional metrics, so their clinical interpretability may be limited. Anchor-based MCIDs are typically produced by using change over time to predict a binary outcome in a receiver operating characteristic (ROC) curve analysis. This binary outcome is the anchor, which conveys the clinical-outcome information of interest. The purpose of MCID and SCB metrics has been well intended, but the actual execution and application of these metrics in the literature may be problematic given the methods commonly

used to calculate these values. For example, many MCID and SCB values have been reported for the same instrument. Although the discrepancies among studies are commonly attributed to population differences, a number of other potential sources of error exist: (1) MCID and SCB calculations are often anchored to different outcomes, which will naturally result in different MCID and SCB estimates; (2) MCID and SCB metrics are commonly presented as a single estimate when they should be reported with the corresponding CIs (ie, actual error variance around estimates is seldom reported); and (3) MCID and SCB metrics are calculated with Δ scores, making them susceptible to regression to the mean (RTM). In this technical note, we detail the potential concern about how RTM may bias existing MCID and SCB metrics in the sports medicine literature.

Repeated observations of a patient produce biased analyses unless the baseline observation is accounted for statistically, a phenomenon known as RTM. Regression to the mean was first documented by Galton² in 1886 and termed *regression to mediocrity*. Whereas RTM is often taught in undergraduate health science curriculums, Gal-

ton's example of parents with above-average height tending to have shorter offspring can seem esoteric and not applicable to a sports medicine setting. A statistically identical but more easily understood example of RTM commonly observed in sports medicine occurs when multiple range-of-motion measurements are obtained for a single group of patients. Patients who present with extreme range of motion at their first visit will naturally tend toward more average ranges of motion at later visits, and the same is true for those who initially present with restricted ranges of motion. This statistical phenomenon is due to both measurement error and inherent variation in the phenomenon being measured.^{3,4} The effects of RTM can be accounted for or mitigated in a number of ways, but the most common and effective statistical method is to use analysis of covariance in which the baseline score is a covariate in the model.³ How this baseline covariate is able to account for RTM can be seen in Figures 1 and 2: the change across time (observed Δ of posttest pretest [post-pre] data) is highly related to the baseline score (pretest data), and the level of this relatedness depends on the correlation between the posttest and pretest data (ie, the slope of the regression line is greater when the correlation between posttest and pretest data is lower). Whereas it is widely accepted that analysis of covariance can be used to account for the effects of RTM on model estimates in regression analyses, it is less well recognized that not accounting for RTM can substantially bias nearly all models with repeated measures, including the ROC curve-based MCID and SCB metrics.

The purpose of this technical note is to demonstrate how existing anchored MCID and SCB estimates are biased by RTM. The methods for calculating MCID and SCB are identical and differentiated only by whether one judges the anchor to be minimal or substantial. For simplicity, we will demonstrate the RTM concern with MCIDs, but this work can be extended to SCB. Using a simulated International Knee Documentation Committee (IKDC) survey dataset, we can demonstrate the effect of RTM on MCID values by evaluating 2 classic characteristics of RTM: (1) MCID estimates are highly dependent on the magnitude of the baseline observation⁵ and (2) the effect of the baseline observation on MCID estimates is higher when the correlation between the posttest and pretest measurements is lower.^{3,5}

METHODS

Data Simulation

Ten simulated datasets containing post-pre data for the IKDC survey (score range = 0–100) were created via statistical simulation. Each dataset consisted of 5000 patients and had a designated correlation between the posttest and pretest data. The 10 datasets were divided into 2 types: 1 in which the variances between posttest and pretest data were equal and 1 in which they were unequal. In the equal-variances datasets, the pretest (baseline) values were normally distributed ($SD = 20$) and centered at 45, and the posttest values were normally distributed ($SD = 20$) and centered at 65. In the unequal-variances datasets, the pretest (baseline) values were normally distributed ($SD = 20$) and centered at 45, and the posttest values were normally distributed but with a dispersion 50% greater than that of

the pretest data ($SD = 30$) and centered at 65. Each dataset was simulated so that the posttest-pretest data had correlation coefficients of 0.10, 0.25, 0.50, 0.75, and 0.90. The binary outcome, on which MCID calculations were based, was simulated using a Bernoulli distribution, in which a positive outcome was more likely when the difference between posttest and pretest values was 20. The code for all data simulations, analyses, and figures was written in R (version 3.6.2; The R Project for Statistical Computing) programming language and can be found at <https://osf.io/56u37/>.

The MCID Calculations

The widely used method for determining an anchored MCID or SCB in the orthopaedic and sports medicine literature leverages the ROC curve analysis.^{6–9} The difference (Δ) between the posttest and pretest values is calculated (ie, $\text{posttest} - \text{pretest} = \Delta$), and then this Δ score is used to predict the binary outcome of the ROC curve analysis. In practice, the binary outcome is often a dichotomized continuous or ordinal variable (eg, Did an athlete return to play in <10 days? *yes* or *no*), but it can also be anchored to a true binary outcome (eg, Did an athlete return to sport after surgery? *yes* or *no*). The researcher then determines the MCID from the ROC curve by attempting to balance specificity and sensitivity using either the top-left corner method or Youden index (J), although the Youden J has been shown to be less biased.¹⁰

It is well known that RTM can be accounted for in a regression analysis by using the baseline measure as a covariate.³ Covariate-adjusted ROC curve analyses are a relatively recent statistical development¹¹ and exist in several forms, incorporating both frequentist and Bayesian approaches and nonlinear covariates.^{11–13} Given the appropriate background of the investigator, any of these options are viable for controlling for RTM in an MCID calculation. For our analysis, we used a 1-dimensional continuous covariate (baseline IKDC score) based on the induced nonparametric ROC curve to calculate a baseline-adjusted MCID for each of the 10 simulated datasets.¹² Similar to the standard MCID calculation method, the MCID was extracted from the covariate-adjusted ROC curve analysis via the Youden J index. To examine the statistical implications and summarize the magnitude of change, we fit a linear regression between the estimated MCID and the observed baseline.

RESULTS

For each of the 5 datasets examining equal variances, the baseline observed IKDC score influenced the MCID estimates (P values < .001). As evidenced by the regression coefficients for the observed baseline IKDC score, the magnitude of effect on the MCID estimate was larger when the post-pre scores had a lower correlation. For every 10-point increase in baseline IKDC score, MCID decreased by 3.5, 2.7, 1.9, 1.2, and 0.7 points in the dataset in which post-pre correlations were 0.10, 0.25, 0.50, 0.75, and 0.90, respectively (Figure 3). When the baseline observations range from 35 to 60 points and account for RTM, the calculated IKDC MCID can vary substantially, ranging from 14 to 24 points based on the magnitude of correlation between posttest and pretest values (Figure 3A).

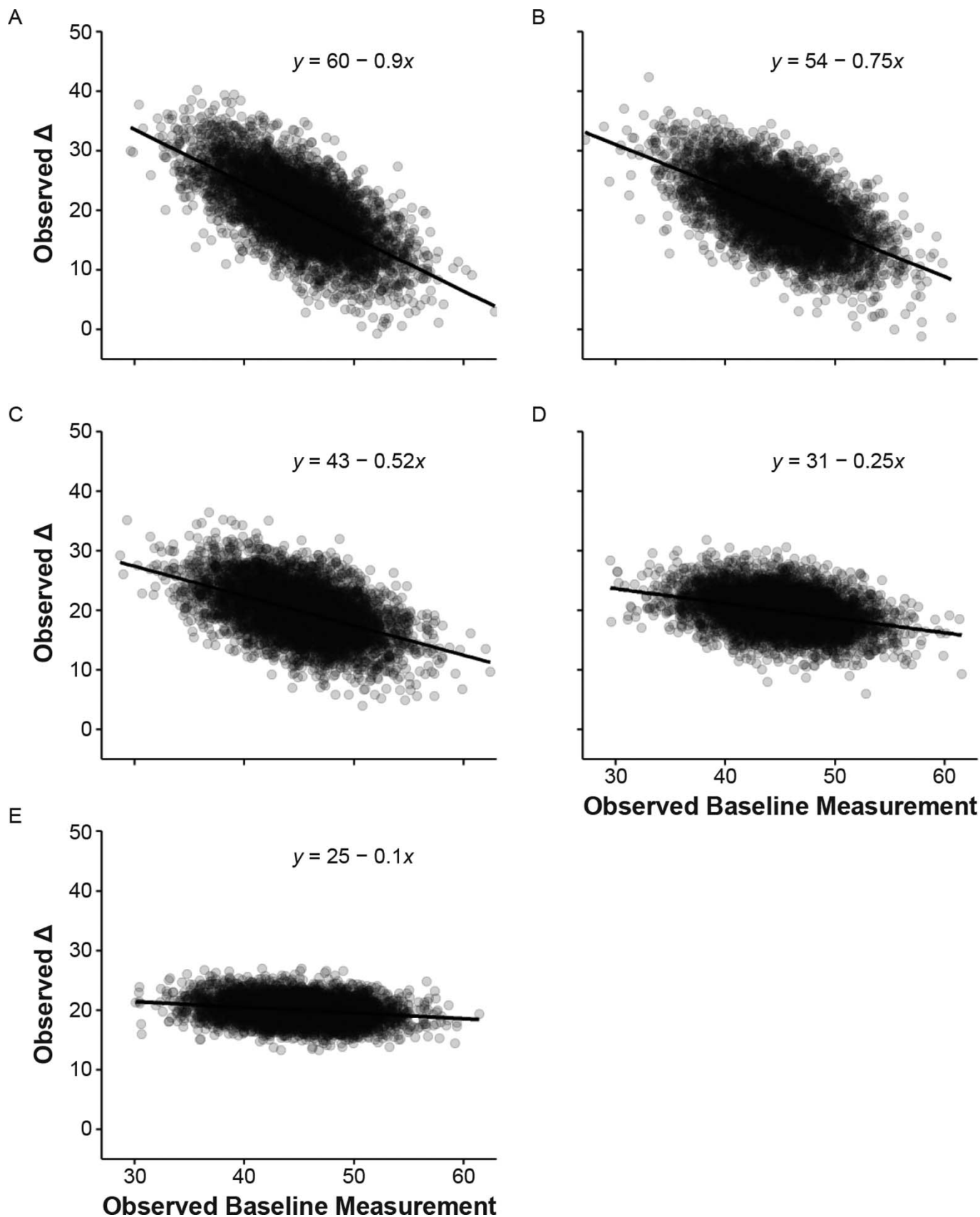


Figure 1. Plots of the Δ (posttest–pretest = Δ) against the observed baseline data (ie, pretest data) for various levels of posttest–pretest data correlation under the scenario of equal variances. A, 0.10 correlation. B, 0.25 correlation. C, 0.50 correlation. D, 0.75 correlation. E, 0.90 correlation. The linear regression lines and equations are plotted to depict that the slope of the correlation is higher when the posttest–pretest data correlation is lower.

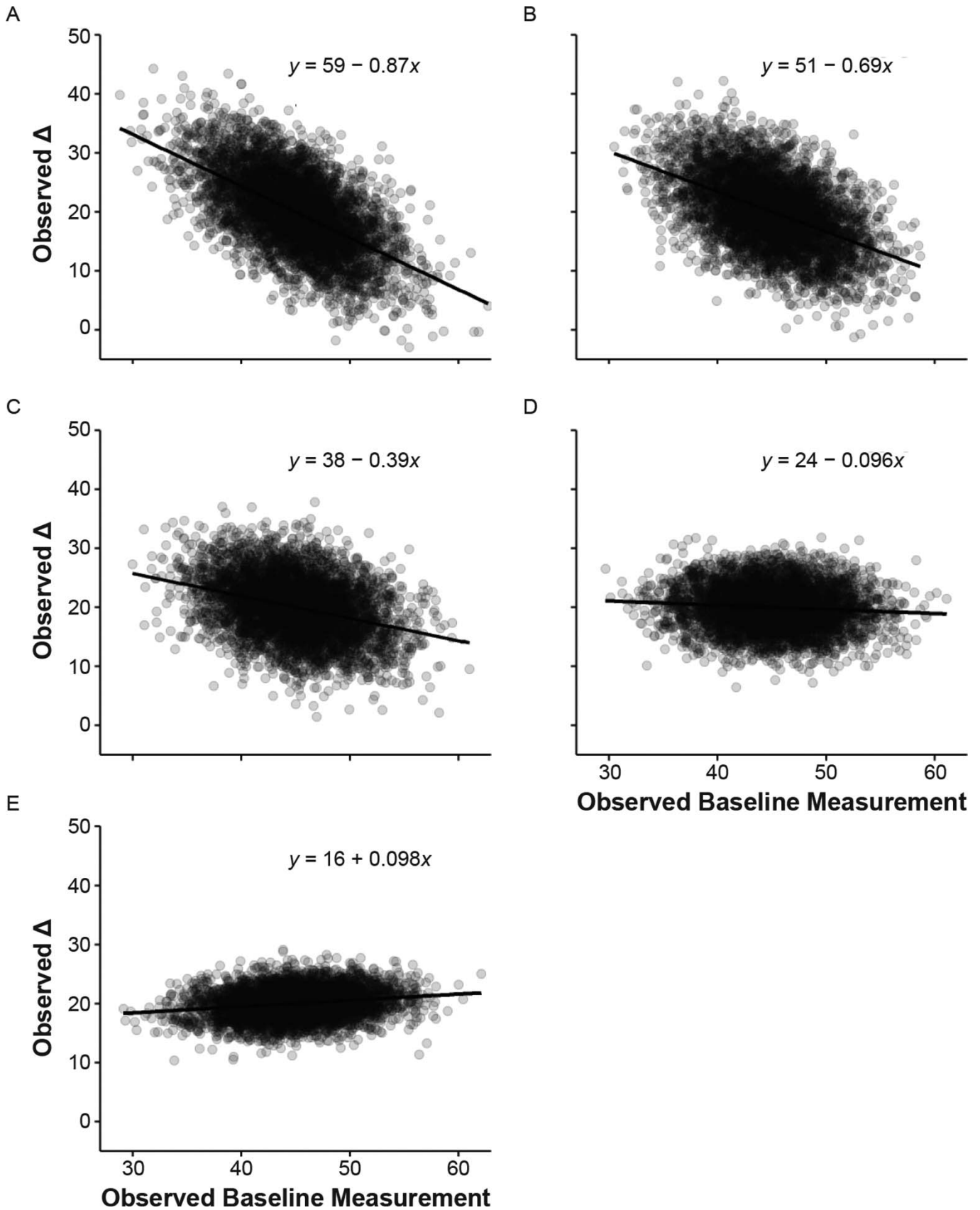


Figure 2. Plots of the Δ (posttest–pretest = Δ) against the observed baseline data (ie, pretest data) for various levels of posttest–pretest data correlation under unequal variances. A, 0.10 correlation. B, 0.25 correlation. C, 0.50 correlation. D, 0.75 correlation. E, 0.90 correlation. The linear regression line and equation are plotted to depict that the slope of the correlation is generally higher when the posttest–pretest data correlation is lower, although both D and E approximate a slope of zero.

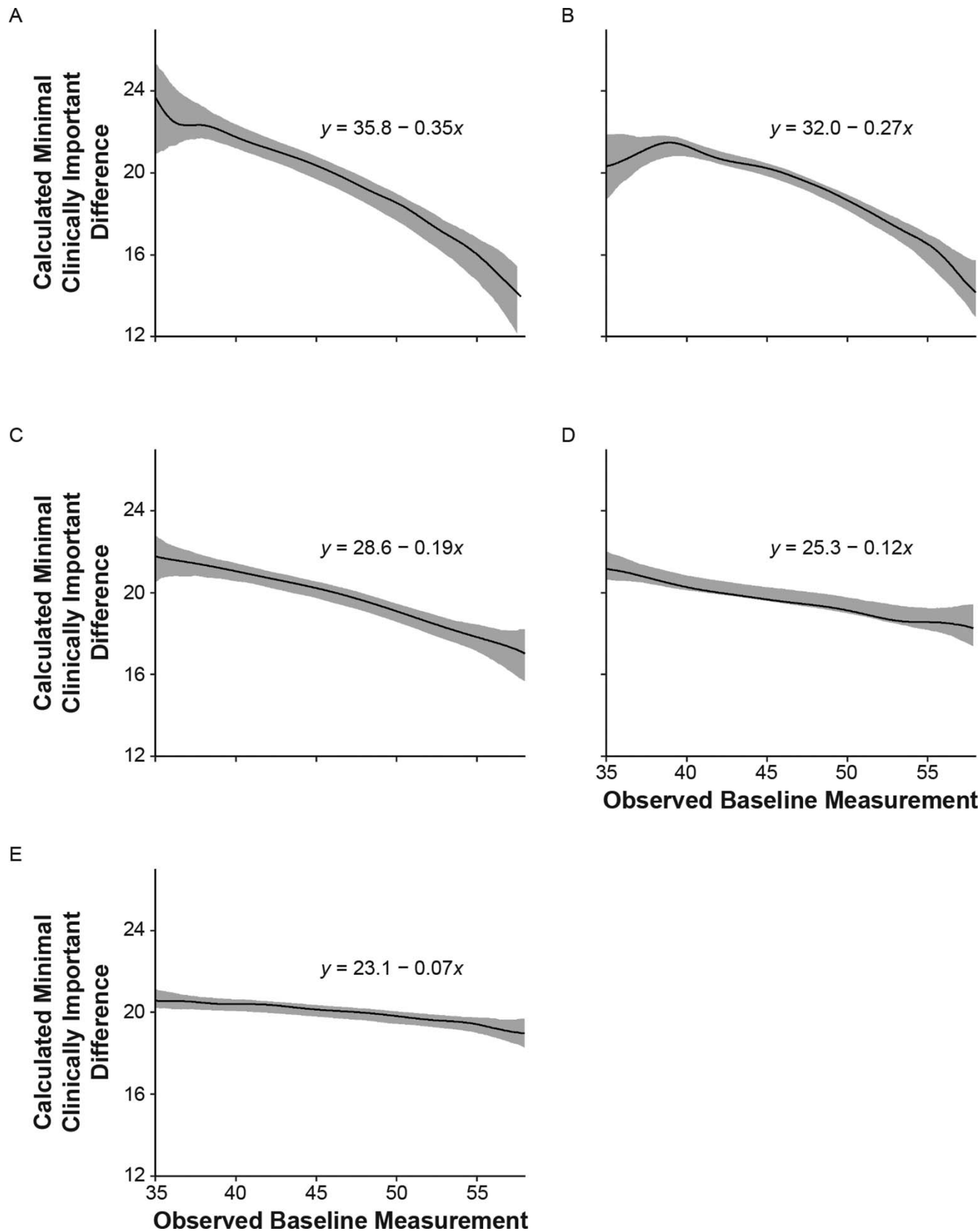


Figure 3. Simulated minimal clinically important differences (MCIDs) are demonstrated with increasing levels of correlation under the scenario of equal variances. A, 0.10 correlation. B, 0.25 correlation. C, 0.50 correlation. D, 0.75 correlation. E, 0.90 correlation. Lower levels of data correlation result in a greater rate of change in the MCID estimate as well as a wider range of MCIDs as a function of the observed baseline data. The regression equations are for a linear regression fit to the data (MCID and observed baseline measurement) and not for the plotted data and bootstrapped 95% CIs.

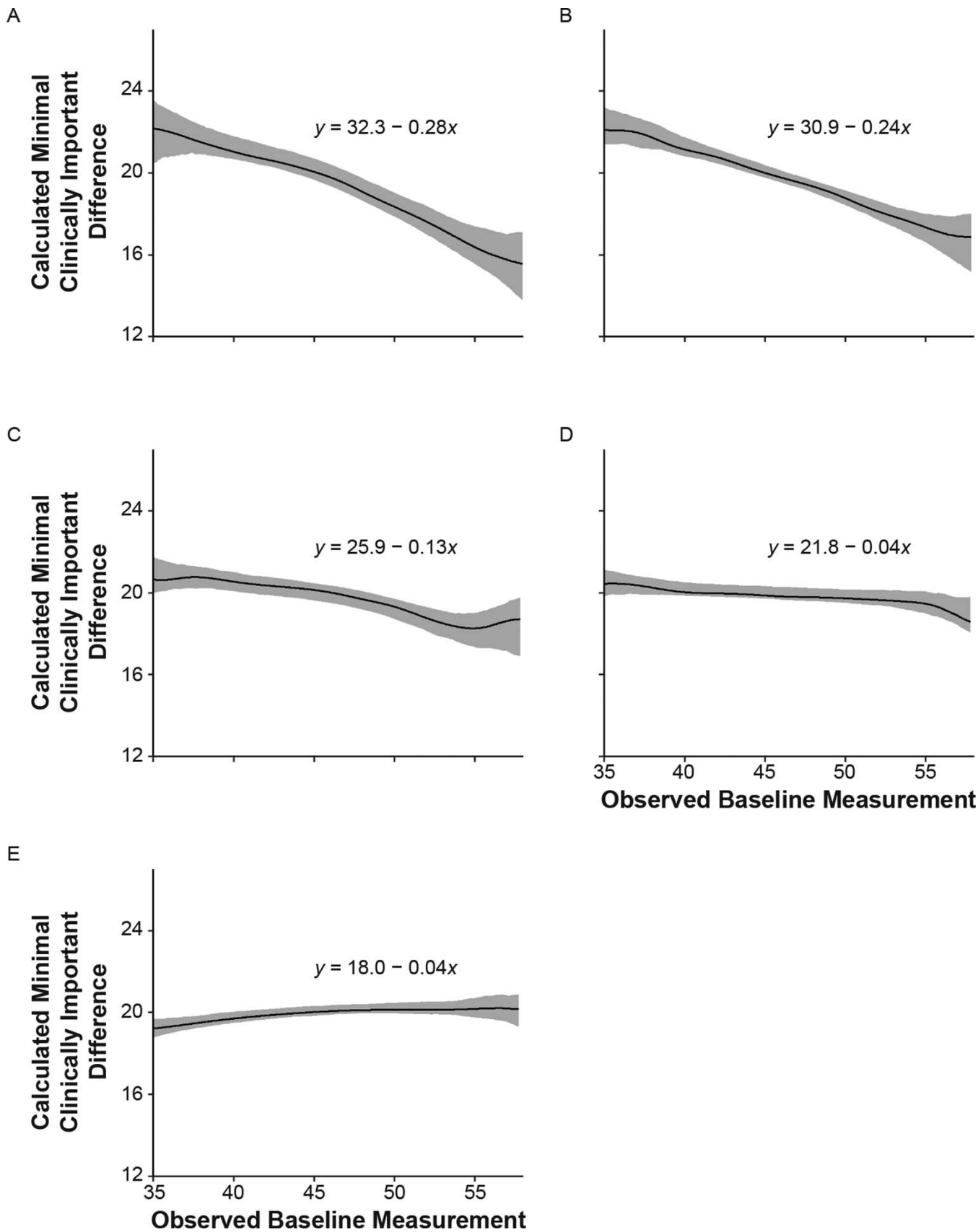


Figure 4. Simulated minimal clinically important differences (MCIDs) are demonstrated with increasing levels of correlation under unequal variances. A, 0.10 correlation. B, 0.25 correlation. C, 0.50 correlation. D, 0.75 correlation. E, 0.90 correlation. Lower levels of data correlation result in a greater rate of change in the MCID estimate as well as a wider range of MCIDs as a function of the observed baseline data, although both D and E approximate a slope of zero. The regression equations are for a linear regression fit to the data (MCID and observed baseline measurement) and not the plotted data and bootstrapped 95% CIs.

By contrast, not accounting for RTM via a baseline covariate results in an MCID that is static at 20 points.

For the 5 datasets with unequal variances, the baseline observed IKDC score had an effect ($P < .001$); however, the regression coefficients for the baseline IKDC scores were not as profound as those observed in the equal-variances datasets. The magnitude of the baseline IKDC effects was still highest when the post-pre correlations were low but was close to zero when the correlations were high (Figure 4). For every 10-point increase in the baseline IKDC score, the MCID decreased by 2.8, 2.4, 1.3, and 0.4 when the posttest-pretest correlations were 0.10, 0.25, 0.50, and 0.75, respectively, and increased by 0.4 when the post-pre correlation was 0.90. As prescribed in the simulation, the static standard MCID calculation always approximates 20 points, but the RTM-controlled MCID can vary from approximately 15 to 23 points, depending on the properties of the underlying data.

DISCUSSION

The results of this study demonstrated that existing anchored MCID and SCB values, both of which are often derived from ROC curve analyses, can be biased by RTM. Indeed, any method using repeated observations, even as Δ scores, can be biased by RTM if the correlation between repeated observations is low and the baseline value is not statistically controlled.

The MCID and SCB are commonly based on a single study, using a researcher-chosen anchor, of a subsample of a population; therefore, the MCID and SCB values derived from a single study should be expected to have some level of subjectivity and error variance around the reported estimates. Our results and the impetus behind this broader effort are not to argue against the use of the MCID and SCB. Rather, we suggest that authors reporting on MCID and SCB should not report these metrics as a singular value. Based on our findings, we propose that future researchers using and reporting anchored MCID and SCB metrics should consider several guidelines:

- (1) Always report the corresponding CI around the MCID or SCB estimate. The interpretative difference between an MCID of 20 and an MCID of 20 with CIs ranging from 10 to 65 is enormous. The former suggests an absolute level of clinical difference, whereas the latter shows that the derived MCID is inconclusive at best.
- (2) Be transparent about the anchors used, especially if the research dichotomized a continuous or ordinal variable.
- (3) Report the correlation coefficient between the posttest and pretest data and the variance of the data at each time point to give informed readers a basic understanding of whether RTM potentially biases the reported MCID or SCB metric.
- (4) Consider using a baseline covariate-controlled ROC curve analysis to calculate the MCID or SCB metric and report this metric with the associated CI.

In our view, the fourth recommendation in particular presents the best picture of what an MCID or SCB value should be for an individual. However, in this conceptualization, the MCID or SCB is not a singular value but rather a function of the baseline. For example, if Figure 3A were a real analysis, one could refer to a figure to say that the

patient who came in with a baseline IKDC of 45 had an MCID of 20, whereas the patient who had a baseline IKDC of 50 had an MCID of 18. The same information could be automatically extracted from an arithmetic equation using an online calculator or application. For example, the MCID calculation for Figure 3A is

$$\text{MCID} = 35.8 - 0.35 * \text{Baseline}$$

It would be easy to automate these calculations in an electronic health record format or web application. A secondary option would be to simply say that the MCID for IKDC is a range from roughly 13 to 24 points. As with many aspects of analytics, there is a trade-off between complexity and accuracy.

These conclusions should be interpreted not as an effort to dictate to researchers or clinicians how they should derive or use MCID and SCB metrics but rather as a demonstration that existing conceptualizations are potentially biased and do not reflect the real-world variability inherent in any analysis. The simulations provided here are useful reminders that fundamental statistical concepts, such as RTM, need to be considered across a wide array of analyses, including MCID and SCB calculations. These proposed guidelines for conceptualizing MCID and SCB calculations should aid clinicians and researchers in recognizing the limitations of current metrics and contemplating future efforts to refine the analysis of clinical information.

ACKNOWLEDGMENTS

The views expressed in this article are ours as authors and do not reflect the official policy of the Department of the Army, Department of the Navy, Department of the Air Force, Department of Defense, or US government.

REFERENCES

1. Smith-Forbes EV, Howell DM, Willoughby J, Pitts DG, Uhl TL. Specificity of the minimal clinically important difference of the quick Disabilities of the Arm Shoulder and Hand (QDASH) for distal upper extremity conditions. *J Hand Ther.* 2016;29(1):81–88. doi:10.1016/j.jht.2015.09.003
2. Galton F. Regression towards mediocrity in hereditary stature. *J Anthropol Inst G B Irel.* 1886;15:246–263. doi:10.2307/2841583
3. 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
4. Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol.* 1976;104(5):493–498. doi:10.1093/oxfordjournals.aje.a112321
5. Morton V, Torgerson DJ. Effect of regression to the mean on decision making in health care. *BMJ.* 2003;326(7398):1083–1084. doi:10.1136/bmj.326.7398.1083
6. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res.* 2015;10(1):24. doi:10.1186/s13018-014-0144-x
7. Hung M, Baumhauer JF, Licari FW, Voss MW, Bounsanga J, Saltzman CL. PROMIS and FAAM minimal clinically important differences in foot and ankle orthopedics. *Foot Ankle Int.* 2019;40(1):65–73. doi:10.1177/1071100718800304
8. Chahal J, Thiel GSV, Mather RC, Lee S, Salata MJ, Nho SJ. The minimal clinically important difference (mcid) and patient acceptable symptomatic state (PASS) for the modified Harris Hip Score and

- Hip Outcome Score among patients undergoing surgical treatment for femoroacetabular impingement. *Orthop J Sports Med.* 2014;2(7 suppl2):2325967114S00105. doi:10.1177/2325967114S00105
9. Lu Y, Beletsky A, Chahla J, et al. How can we define clinically important improvement in pain scores after biceps tenodesis? *J Shoulder Elbow Surg.* Published online June 25, 2020. doi:10.1016/j.jse.2020.05.038
 10. Perkins NJ, Schisterman EF. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol.* 2006;163(7):670–675. doi:10.1093/aje/kwj063
 11. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol.* 2008;168(1):89–97. doi:10.1093/aje/kwn099
 12. Rodríguez-Álvarez MX, Roca-Pardiñas J, Cadarso-Suárez C. ROC curve and covariates: extending induced methodology to the non-parametric framework. *Stat Comput.* 2011;21(4):483–499. doi:10.1007/s11222-010-9184-1
 13. de Carvalho VI, Rodriguez-Alvarez MX. Bayesian nonparametric inference for the covariate-adjusted ROC curve Web site. *ArXiv.* Published May 30, 2018. Accessed July 6, 2020. <http://arxiv.org/abs/1806.00473>

Address correspondence to Matthew S. Tenan, PhD, ATC, 1700 North Moore Street, Suite 2300, Rosslyn, VA 22209. Address email to matt@optimumanalytics.org.