Commentary

Invited commentary on ‘different minimal clinically important difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index when using the same sample of patients’

Steven Z. George, PT, PhD

Associate Professor and Assistant Department Chair,
University of Florida, Gainesville FL (USA)

Minimal clinically important difference (MCID) scores are frequently used by clinicians and researchers to describe outcomes following treatment episodes. There is debate regarding appropriate interpretation and application of MCID's because multiple values exist for commonly used outcome measures.^{1,2} Schwind et al³ add to the MCID literature with an interesting secondary analysis from a randomized clinical trial of patients with low back pain receiving thrust or non thrust manipulation in combination with other physical therapist directed approaches. In this analysis, Schwind et al³ identify baseline predictors for 6 previously established MCID's for a commonly used outcome measure, the Oswestry Disability Index (ODI). Previous studies have mostly focused on development of a particular ODI MCID, so this analysis is a change in direction that may have relevance for clinicians and researchers. This commentary will highlight key issues that affect interpretation of this paper with both of those audiences in mind.

There are some general methodological factors to consider before specific issues are discussed in more detail. This study had notable strengths that included a uniform and well-described treatment approach and a sufficient (though not ideal, as the authors acknowledge) sample size for these analyses. In addition, one of the regression models showed overall good fit (prediction of final ODI score $\leq 20\%$; Nagelkerke $R^2=0.49$) which gives readers some confidence in that particular model in comparison to others reported in the paper (Nagelkerke $R^2=0.14$ to 0.27). An issue that may affect interpretation is related to how the authors implemented prognostic factors in the analysis. These factors were included in the study based on prior evidence and clinical experience, which is a sound approach. However, some prognostic factors represented individual variables (e.g. age and BMI) while other variables represented clusters of variables (e.g. clinical prediction rule status and irritability). The authors did include

individual variables that were also used for cluster variables in the analysis (e.g. FABQW scores and symptom duration are part of the clinical prediction rule), but not all were considered in this manner (e.g. spinal mobility and location of pain were not considered independently). The difference in what each variable represented is an important issue because not all variables included in the logistic regression models accounted for the same amount of information. This difference in amount of information may lead to cluster variables being more likely to be included in regression models simply because they account for more information in comparison to an individual variable.

Schwind et al³ highlighted the ‘extreme variability’ from their regression models and such results are not surprising because the authors essentially had different, but presumably correlated, dependent variables for each model (i.e., differing estimates based on ODI scores). The authors correctly point out that their findings reinforce the ‘inherent fragility of predictive modeling’ which is a key point for readers to understand. Robust predictors are expected to replicate across different study settings when the same outcome measures are used; however, it is not clear if robust predictors are also expected to replicate across different outcome measures used in the same study setting. This analysis addressed this issue and clearly demonstrated that specificity of prediction from regression-based techniques is inevitable when the same outcome measure is expressed in different forms for a single patient cohort. It is at this point where interpretation of these findings hinges on how different the 6 MCID variables are from each other. These data were not directly reported in the study, but from the variability in the baseline predictor variables that remained in the regression models and the data reported in Figure 1, it appears that the different MCID's for the ODI are unlikely to be highly correlated with each other.

It is expected that readers accepting the suggestion that the variability in predictors reported in this paper can then identify consistent themes to guide interpretation of these data. For the sake of simplicity, these observations will focus on the models that included all study participants (Table 2), as results were very similar for the subset with higher ODI scores. First, the results from Schwind et al³ reinforce the importance of baseline scores in predicting clinical outcomes of the same measure. In 4/6 of the regression models, the baseline ODI (in differing forms but all related to the initial score level as higher or lower) was a predictor

Correspondence to: Steven Z. George, University of Florida, 925 NW 56th Terrace, Gainesville, FL, USA. Email: szgeorge@phhp.ufl.edu

of the MCID. In fact, the baseline ODI was only NOT involved in prediction when the MCID included the baseline value in the calculation of a percentage change. This is an expected finding and is consistent with studies from our own group using outcomes in the original continuous scale,^{4,5} and from other clinical outcome studies in primary care settings.^{6,7} To paraphrase Everlast, these analyses are further demonstration that, even for categorical outcomes, 'where it ends up usually depends on where you start'.

Second, these results indicate that variables representing clusters of examination findings may be more powerful in predicting MCID outcomes when compared to individual variables. For example, the previously validated clinical prediction rule followed baseline ODI as a predictor in 3/6 of the regression models, and 2 individual components of the clinical prediction rule (FABQW score and symptom duration) were predictors in a fourth model. The identification of the clinical prediction rule is a surprising finding since previous clinical trials have suggested that this rule is specific to thrust manipulation responses.^{8,9} However, the current study suggests that the clinical prediction rule has broader implications for clinical decision making on non-thrust manipulation techniques. An important methodological distinction to make is that in some studies the clinical prediction rule was included as a treatment modifier, and in this study it was investigated as a prognostic factor.¹⁰ Therefore, these findings do give further credence to the clinical prediction rule as an indicator of favorable clinical outcomes when manual therapy is involved. However, the current study only adds to the debate on whether this clinical prediction rule is a favorable prognostic factor (general indicator of high chance of successful outcome), as the methodology did not directly test for treatment moderation based on clinical prediction rule status (specific indicator that receiving spinal manipulation increases chance of successful outcome).¹⁰ As previously mentioned in this commentary, the consistent identification of the clinical prediction rule could be related to the fact that this cluster variable represented more than one examination finding in the analysis.

In summary, the overall interpretation of this article from Schwind et al³ depends on whether prediction of MCID levels becomes of widespread interest. If that is the case, then the differences in predictors reported in this analysis could be an important issue to resolve. In the current *JMMT* paper and in a companion paper published by Cook et al¹¹ in *PTJ*, the authors bring up the idea of identifying robust prognostic factors to develop 'universal predictors' for various low back pain clinical outcomes. This is an intriguing idea, but its pursuit begs the question of whether outcome measures that are correlated, but not redundant (e.g. pain and disability), should be expected to have a universal set of

predictors. Certainly the current study demonstrates that there are inherent differences in predictors even when the same measure was dichotomized into 6 different categories. The pursuit of a 'universal MCID' would be the easiest solution to this problem, as that would allow for relatively quick development of universal predictors for that particular MCID value. This is unlikely to occur because the magnitude of a 'clinically important difference' will vary based on practice settings and sample characteristics, making consensus on a MCID with universal application to individual patients next to impossible to establish.¹² Therefore, this interesting article from Schwind et al³ reinforces the need for individual clinics and health care systems to establish a standard MCID based on characteristics of the specific environment. Once established, a standard MCID would allow for outcome comparisons across patients and practitioners within the same environment. In contrast, use of multiple MCID's in the same environment or comparisons of the same MCID across environments should be discouraged because doing so may complicate decision making for outcomes rather than simplify it.

Acknowledgments

Jason Beneciuk and Mark Bishop provided review and comments on an earlier version of this commentary.

References

- 1 Ostelo RW, Deyo RA, Stratford P *et al.* Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008;33:90–94.
- 2 Terwee CB, Roorda LD, Dekker J *et al.* Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524–534.
- 3 Schwind J, Learman K, O'Halloran B *et al.* Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry Disability Index when using the same sample of patients. *J Man Manip Ther* 2013;21:in press.
- 4 Beneciuk JM, Bishop MD, Fritz JM *et al.* The STarT Back Screening Tool and Individual Psychological Measures: Evaluation of Prognostic Capabilities for Low Back Pain Clinical Outcomes in Outpatient Physical Therapy Settings. *Phys Ther* 2012.
- 5 George SZ, Coronado RA, Beneciuk JM *et al.* Depressive symptoms, anatomical region, and clinical outcomes for patients seeking outpatient physical therapy for musculoskeletal pain. *Phys Ther* 2011;91:358–372.
- 6 Foster NE, Thomas E, Bishop A *et al.* Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain* 2010;148:398–406.
- 7 Sieben JM, Vlaeyen JW, Portegijs PJ *et al.* A longitudinal study on the predictive validity of the fear-avoidance model in low back pain. *Pain* 2005;117:162–170.
- 8 Childs JD, Fritz JM, Flynn TW *et al.* A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med* 2004;141:920–928.
- 9 Cleland JA, Fritz JM, Kulig K *et al.* Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: a randomized clinical trial. *Spine (Phila Pa 1976)* 2009;34:2720–2729.
- 10 Hill JC, Fritz JM. Psychosocial influences on low back pain, disability, and response to treatment. *Phys Ther* 2011;91:712–721.
- 11 Cook CE, Learman KE, O'Halloran BJ *et al.* Which prognostic factors for low back pain are generic predictors of outcome across a range of recovery domains? *Phys Ther* 2013;93:32–40.
- 12 de Vet HC, Terluin B, Knol DL *et al.* Three ways to quantify uncertainty in individually applied 'minimally important change' values. *J Clin Epidemiol* 2010;63:37–45.