Clinical Practice Study

Establishing minimum clinically important difference values for the Patient-Reported Outcomes Measurement Information System Physical Function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics

Man Hung, Jerry Bounsanga, Maren W Voss, Charles L Saltzman

Man Hung, Jerry Bounsanga, Maren W Voss, Charles L Saltzman, Department of Orthopaedic Surgery Operations, University of Utah, Salt Lake City, UT 84108, United States

Man Hung, Division of Public Health, University of Utah, Salt Lake City, UT 84108, United States

Man Hung, Population Health Research Foundation, University of Utah, Salt Lake City, UT 84112, United States

ORCID number: Man Hung (0000-0003-2827-3740); Jerry Bounsanga (0000-0001-6852-4650); Maren W Voss (0000-0002-9054-7389); Charles L Saltzman (0000-0003-2877-2339).

Author contributions: Hung M was instrumental in study design, data collection, data processing, data analysis, interpretation, study oversight, literature review, manuscript writing and editing; Bounsanga J was instrumental in data processing, data analysis, manuscript writing and editing; Voss MW was instrumental in literature review, data analysis, manuscript writing and editing; Saltzman CL was instrumental in study design, data collection, interpretation and manuscript editing.

Supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health, No. U01AR067138. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Institutional review board statement: Institutional review board approval/waiver was obtained through the University of Utah, IRB #00053404.

Informed consent statement: Study participants completed PRO measures as part of the standard of care in treatment.

Conflict-of-interest statement: The authors have no conflicts of interest to disclose.

Data sharing statement: The dataset analyzed during the current study is not publicly available due to HIPAA regulations but may be available from the University of Utah if the requestor has obtained proper IRB approval from the University of Utah and its Enterprise Data Warehouse.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Manuscript source: Invited manuscript

Correspondence to: Man Hung, PhD, Associate Professor, Department of Orthopaedic Surgery Operations, University of Utah, 590 Wakara Way, Salt Lake City, UT 84108, United States. man.hung@hsc.utah.edu

Telephone: +1-801-5875372

Fax: +1-801-5875411

Received: January 4, 2018

Peer-review started: January 4, 2018

First decision: January 23, 2018

Revised: January 30, 2018

Accepted: February 6, 2018

Article in press: February 6, 2018
Abstract

AIM
To establish minimum clinically important difference (MCID) for measurements in an orthopaedic patient population with joint disorders.

METHODS
Adult patients aged 18 years and older seeking care for joint conditions at an orthopaedic clinic took the Patient-Reported Outcomes Measurement Information System Physical Function (PROMIS® PF) computerized adaptive test (CAT), hip disability and osteoarthritis outcome score for joint reconstruction (HOOS JR), and the knee injury and osteoarthritis outcome score for joint reconstruction (KOOS JR) from February 2014 to April 2017. MCIDs were calculated using anchor-based and distribution-based methods. Patient reports of meaningful change in function since their first clinic encounter were used as an anchor.

RESULTS
There were 2226 patients who participated with a mean age of 61.16 (SD = 12.84) years, 41.6% male, and 89.7% Caucasian. Mean change ranged from 7.29 to 8.41 for the PROMIS® PF CAT, from 14.81 to 19.68 for the HOOS JR, and from 14.51 to 18.85 for the KOOS JR. ROC cut-offs ranged from 1.97-8.18 for the PF CAT, 6.33-43.36 for the HOOS JR, and 2.21-8.16 for the KOOS JR. Distribution-based methods estimated MCID values ranging from 2.45 to 21.55 for the PROMIS® PF CAT; from 3.90 to 43.61 for the HOOS JR, and from 3.98 to 40.67 for the KOOS JR. The median MCID value in the range was similar to the mean change score for each measure and was 7.9 for the PF CAT, 18.0 for the HOOS JR, and 15.1 for the KOOS JR.

CONCLUSION
This is the first comprehensive study providing a wide range of MCIDs for the PROMIS® PF, HOOS JR, and KOOS JR in orthopaedic patients with joint ailments.

Key words: Hip disability and osteoarthritis outcome score for joint reconstruction; Patient-Reported Outcomes Measurement Information System Physical Function; Knee injury and osteoarthritis outcome score for joint reconstruction; Minimum clinically important difference; Joint; Physical function; Minimum detectable change; Arthroplasty; Orthopaedics; Clinical outcomes

© The Author(s) 2018. Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: Personal value judgments should be used to apply these minimum clinically important difference (MCID) values to treatment planning and in guiding patient expectations of change. We recommend applying low values of MCIDs for screening purposes and median values as a more conservative cut-off for evaluating longitudinal change.

INTRODUCTION
The National Institutes of Health sponsored an initiative to develop a Patient-Reported Outcomes Measurement Information System (PROMIS®) using item response theory (IRT) methods. These methods have been recommended for use in the evaluation of arthroplasty outcomes because of their unique measurement properties. Computerized adaptive testing (CAT) using IRT minimizes respondent burden without sacrificing instrument precision. The PROMIS® Physical Function (PF) CAT assesses physical function in five domains but is not directly targeted at joint function. Yet the useful measurement and administration qualities of the PROMIS® PF CAT make it a valuable addition to patient-reported outcomes (PRO) assessments in joint reconstruction. Aside from minimizing burden, top quality PRO instruments also offer reliable and valid scores that are easy to interpret.

Recent attention in PRO development has focused on the interpretability of scores, particularly in terms of how meaningful the outcomes are to patients. Change in function is an important clinical outcome, thus PROs should be able to detect change in patient function. In interpreting change, the minimum clinically important difference (MCID) reflects the smallest amount of meaningful change. “MCIDs are patient derived scores that reflect changes in a clinical intervention that are meaningful for the patient.” Meaningful change is important, as it can serve as a benchmark of treatment effect, and it is critical in decision making. Whether or not a treatment produces a statistically significant outcome is less informative than whether it produces meaningful change, as an inflated sample size can yield statistical significance without clinical relevance. Multiple methods have been developed for determining MCID values and there is little agreement on the best standard to apply. Distribution-based approaches rely on statistical methods and probability sampling. They describe how much change falls beyond random levels of variation. They rely on distributions of scores and how much the scores vary between patients in reaching a magnitude of change that is beyond chance fluctuation. But distribution methods cannot
tell us whether the amount of change is meaningful from the clinician’s or the patient’s perspective\textsuperscript{[8]}. An alternative approach is to use anchor-based methods which relates the change in patient scores to some other measure of health outcomes\textsuperscript{[13]}.

Determining the MCID of the PROMIS\textsuperscript{®} PF is an important step to understand the meaning of the scores. Collecting the longitudinal data necessary to analyze meaningful change takes time. Because the PROMIS\textsuperscript{®} development began quite recently in 2004\textsuperscript{[1]}, there have been very few studies estimating MCIDs for PROMIS\textsuperscript{®} measures and little is known about MCID values in the orthopaedic adult reconstruction population\textsuperscript{[14]}. Initial MCID development for PROMIS\textsuperscript{®} instruments has been in specific patient populations such as pediatrics\textsuperscript{[15]} or cancer patients\textsuperscript{[16]} but studies are lacking in orthopaedic patients.

Other PROs have been developed that are specific to the domain of joint function and it is helpful to consider the measurement properties of these newer instruments side by side. The knee injury and osteoarthritis outcome score for joint reconstruction (KOOS JR) and the hip disability and osteoarthritis outcome score for joint reconstruction (HOOS JR) are two joint specific instruments recently introduced in the arthroplasty arena\textsuperscript{[17-20]}. Both instruments have been approved by the Centers for Medicaid and Medicare Services for use in joint replacement registries\textsuperscript{[21-24]}. The HOOS and KOOS were selected given their use in joint replacement registries in the United States, though other valid measures such as the Oxford hip or knee score are more common in European registries and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) has a longer-term use for more general assessments of osteoarthritis\textsuperscript{[21]}. Prior research has demonstrated that the HOOS and KOOS measures are sensitive to change\textsuperscript{[25]}; however, we are unaware of any studies that develop MCIDs for these instruments\textsuperscript{[26]}. The purpose of this study was to determine the MCIDs of the PROMIS\textsuperscript{®} PF CAT, the HOOS JR, and the KOOS JR in an orthopaedic population with lower-extremity joint conditions.

**MATERIALS AND METHODS**

**Patient sample**
After receiving approval from the University of Utah Institutional Review Board, we analyzed data from an adult reconstruction population at a single academic medical center. All patients, aged 18 and older with pathology of the hip and knee were eligible for inclusion. As part of the standard of care, patients were provided with internet enabled hand-held tablets to answer demographic and PRO instruments which are linked to the electronic health record system via mEval. Questionnaires were either completed at clinic check-in or within 7-d prior to their visit via email. Follow-up data were obtained in the same manner at future clinic visits. No informed consent was obtained as all PRO measurement was conducted as a part of customary standard of care. For this study, these visits were organized into four follow-up periods based upon the first recorded score within the database (baseline initial assessment): (1) 3-mo follow-up (80 to 100 d after initial assessment); (2) > 3-mo follow-up (90 d or more after initial assessment); 6-mo follow-up (170 to 190 d after initial assessment); and (3) > 6-mo follow-up (180 d or more after initial assessment). These follow-up periods were selected based upon recommendations within the literature\textsuperscript{[27-37]}. It should be noted that the baseline score may not necessarily correlate with a specific intervention. Nonetheless, this method still allows for the monitoring of change over time.

A major goal of MCID determination is to allow for meaningful interpretation of scores for clinical decision-making, in addition to sample size calculation for investigating treatment effectiveness. As there is essentially no evidence that MCID values are dependent on the severity of disease conditions, length of follow-ups, or specific patient groups\textsuperscript{[38-40]}, it is appropriate for this study to establish MCIDs among orthopaedic patients with a full range of joint impairments and varying follow-up time points. The present evaluation of MCID values was conducted in a general joint clinic population, among surgical and non-surgical patient samples regardless of specific treatment or intervention. Since MCID development were not meant to be treatment specific\textsuperscript{[40]}, MCID values derived from this study can be applied to adult reconstruction patients with all types of surgical and non-surgical interventions.

**Patient-reported outcomes**
The PROMIS® PF CAT, v1.2, draws from a 121-item test bank that contains both upper extremity and lower extremity functional items. The PROMIS® PF CAT algorithms were established by PROMIS® developers\textsuperscript{[41]}; and the instrument was scored using T-scores, a standardized metric that has a mean of 50 and a standard deviation of 10\textsuperscript{[5]}. Higher scores on the PROMIS® PF CAT indicate higher physical function.

**HOOS JR:** The HOOS JR is a 6-item measure assessing function and pain\textsuperscript{[22]} with psychometric properties similar to the full HOOS\textsuperscript{[42]}. The HOOS JR was scored on a 0 - 100 scale with larger numbers indicating higher hip function.

**KOOS JR:** The KOOS JR is a 7-item measure assessing function and pain\textsuperscript{[24]}. The KOOS JR is scored on a 0-100 scale with larger numbers indicating higher levels of knee function.

**Statistical analysis**
Descriptive statistics regarding patient characteristics and demographics were calculated. Mean change scores for the patients were evaluated for each time-period. Change scores were calculated as the follow-up score minus the baseline score on each measure, and
Table 1  Demographics of patients (n = 2226)

| Patient characteristics | n    | Percent | Mean (SD) | Range |
|-------------------------|------|---------|-----------|-------|
| Age (yr)                | 61.16 (12.84) | 18-93   |
| Gender                  |      |         |           |       |
| Male                    | 927  | 41.6    |           |       |
| Female                  | 1299 | 58.4    |           |       |
| Race                    |      |         |           |       |
| White or Caucasian      | 1997 | 89.7    |           |       |
| Black or African        | 1.2  |         |           |       |
| American Indian and     | 32   | 1.4     |           |       |
| Alaska Native           |      |         |           |       |
| Native Hawaiian or Pacific Islander | 11 | 0.5 |       |
| Other                   | 113  | 5.1     |           |       |
| Unknown/missing         | 26   | 1.2     |           |       |
| Ethnicity               |      |         |           |       |
| Hispanic                | 114  | 5.1     |           |       |
| Non-Hispanic            | 2075 | 93.2    |           |       |
| Missing                 | 37   | 1.7     |           |       |

recorded as the absolute value difference between the scores. They were calculated for each of the follow-up periods described above including 3-mo, > 3 mo, 6-mo, and > 6 mo time-points.

The anchor-based methods applied patients’ perspective to the question: “Compared to your FIRST EVALUATION at the University Orthopaedic Center: how would you describe your physical function now?” (much worse, worse, slightly worse, no change, slightly improved, improved, much improved) as a determinant of meaningful change. No change equates to a 0 value; the negative ratings are from -3 to -1 and positive ratings are from 1 to 3. When change is anchored to the patient perception or report of deterioration or improvement, it can be interpreted as a meaningful (or noticeable) level of change. Patients with a ± 2 or ± 3 point change (much worse, worse, improved, much improved) were included in each analysis of change, a method used to distinguish noticeable change from no-change. Patients reporting no change or only slight change were considered together as the no-change group. We combined the improved and deteriorated conditions using absolute values of the change scores to distinguish change from stable symptomology.

The distribution-based methods included calculations based on the standard deviation (SD) and on the minimum detectable change (MDC). The SD approach used the 1/2 SD and 1/3 SD as variation-based estimates of MCID. The MDC is based on the standard error of measurement (SEM) of the follow-up scores and is the smallest score change that likely reflects a true change in condition. We calculated MDC at three confidence levels: 90%, 95% and 99%. The formulas for calculating the MDC are: MDC_{90%} = 1.65 \times SD; MDC_{95%} = 1.96 \times SD; MDC_{99%} = 2.56 \times SD. The SEM equals to SD\times(1-r)^{1/2}, where r is the reliability represented by Cronbach alpha and SD is the standard deviation of the follow-up scores.

We fitted the receiver operating curve (ROC) to measure the best cut-off points to maximize sensitivity and specificity of the instruments. The cut-off was calculated as (sensitivity + specificity) -1, based on Youden’s J value. Sensitivity is the proportion of correct identification of patients who showed changes, and specificity is the proportion of correct identification of patients showing no meaningful change. All statistical analyses were performed using SPSS 24.0 (IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp.) and R 3.30 (R Development Core Team, Vienna, AT: R Foundation for Statistical Computing).

RESULTS

Demographics
A total of 2226 patients were included in the study. Mean age of participants was 61.16 years (SD = 12.84), with a range of 18-93 years. Most participants in the study were White or Caucasian (89.7%) and Non-Hispanic (93.2%). Detailed demographics can be found in Table 1.

Anchor-based methods
Mean change: Mean change scores varied at each follow-up period. In terms of patients experiencing change, the highest mean change scores were observed at the 6-mo follow-up for the PF CAT (8.41), HOOS JR (19.68), and KOOS JR (18.85). The lowest mean change scores for patients experiencing change was at the 3-mo follow-up for all three measures. For the PF CAT mean change at 3-mo was 7.60 points, for the HOOS JR it was 14.81 points, and for the KOOS JR. It was 14.51 points. The median MCID value from the complete range of all two anchor-based methods and five distribution-based methods analyzed at all four time-points was similar to the mean change score for each measure and was 7.9 for the PF CAT, 18.0 for the HOOS JR, and 15.1 for the KOOS JR (Figure 1).

ROC curve: The ROC area under the curve was used to identify the optimal cut-off point between meaningful and non-meaningful change, by calculating the point at which the sum of false positive and false negative identifications are the fewest. The highest ROC MCID value for the PF CAT was 8.18 at > 6-mo follow-up. For the HOOS JR the highest value was 43.36 observed at the 3-mo follow-up with a small sample size (n = 24). For the KOOS JR, the highest MCID value was 8.16 at > 3-mo follow-up. Detailed mean change and ROC cut-off values for the PF CAT, HOOS JR. and KOOS JR. can be found in Table 2.

Distribution-based methods
1/2 standard deviation of each of the function and pain scores: The PF CAT (4.35) had the highest ½ SD value at the > 6-mo follow-up whereas the HOOS
JR (10.96) had the highest 1/2 SD value at the 6-mo follow-up. The highest 1/2 SD value for the KOOS JR was 8.02 at the > 3-mo follow-up. At the 3-mo follow-up, PF CAT and HOOS JR had the lowest 1/2 SD values of 3.68 and 5.86. The KOOS JR had the lowest 1/2 SD value of 5.96 at the 6-mo follow-up.

1/3 standard deviation of each of the function and pain scores: The highest 1/3 SD value for the PF CAT (2.90) was observed at the > 6-mo follow-up whereas the HOOS JR (7.30) showed the highest 1/3 SD value at the 6-mo follow-up. The KOOS JR had the highest 1/3 SD value of 5.34 at the > 3-mo follow-up but the lowest 1/3 SD value of 3.98 at the 6-mo follow-up. The lowest 1/3 SD values for both the PF CAT (2.45) and HOOS JR (3.90) were observed at the 3-mo follow-up.

Minimum detectable change (MDC) @90%, @95%, @99%: The MDC was highest for the PF CAT (MDC90% = 13.89, MDC95% = 16.50, MDC99% = 21.55) at the 6-mo follow-up. At the > 6-mo follow-up, the MDC was highest for the HOOS JR (MDC90% = 28.86, MDC99% = 41.67) and highest at > 3-mo HOOS JR (MDC99% = 31.90). The highest MDC for the KOOS JR (MDC90% = 26.21, MDC95% = 31.14, MDC99% = 40.67) was observed at the > 3-mo follow-up (Table 3).

**DISCUSSION**

Determining the values of MCID for PRO measures enhances the interpretability and utility of the instruments. Determining score changes which reflect meaningful change requires understanding the patient perception of improvement or worsening. Anchor-based methods of determining an MCID tie the change scores to the patient reports of change and supplement the distribution-based methods which focus on precision in scores from a statistical standpoint. The current
The findings demonstrate that the method used

comprehensive assessment of MCID values utilized multiple methods of anchor-based and distribution-based estimation.

The PROMIS® PF CAT mean change scores in this population of orthopaedic joint patients ranged from 7.29 to 8.41 depending on follow-up points. The ROC area under the curve ranged from 1.97 to 8.18. Distribution-based methods ranged using the SD method from a low of 2.45 (1/3 SD) to 4.35 (½ SD). The values using the MDC approach produced much larger values from 11.64 (MDC90) to 21.55 (MDC95). MCID values vary depending on the type of patient population (e.g., orthopaedic, cancer, pediatric) being evaluated, but overall the MCID values identified with the MDC approach in this study are much higher than the 2-3 point MCID of the PROMIS® mobility subscale identified in a pediatric population[15] or the 4-6 point PROMIS® PF CAT MCID values identified in a cancer population[16].

The larger MCID values in this study could have been obtained for a number of reasons. The values could be population specific, reflecting an underlying difference in functional gain between the orthopaedic adult patient population and the pediatric and cancer patient populations previously studied. However, the most likely explanation is that the MCID value obtained is a reflection of the method used for calculation. Unlike other studies that commonly use just one approach to estimate MCID values, the current study used multiple approaches and multiple time points to arrive at twenty-eight MCID values per instrument, allowing better triangulation of results. This study found a great deal of consistency across the follow-up time-points for any one particular method of calculating MCIDs (Tables 2 and 3), suggesting that the methods were responsible for the variation of the MCID values and that the length of follow-ups was mostly irrelevant. It is important to recognize that this study incorporated a wide array of cut-off standards, ranging from lenient to extremely strict. Since MDC90 and MDC99 are extremely strict cut-off standards, it is not surprising that these MCID values are high. None of the research reviewed as background for this study calculated MCIDs using such strict criteria[15,16]. The lowest MCID values derived from this study may be appropriate for screening purposes, but the median value (Figure 1) may be a better estimate of true and meaningful change when applying a conservative standard for evaluating treatment effects or for respondent analyses[50]. The more stringent cut-off standard derived MCID values reported in this study provide more definite assurance that important change has occurred.

The mean change for the HOOS JR ranged from 14.81 to 19.68. The ROC area under the curve ranged from 6.33 to 43.36. The 43.36 value appeared to be an outlier and is not considered a reliable estimate. The SD values ranged from 3.90 (1/3 SD) to 10.96 (½ SD). The MDC method of detecting change yielded a range of MCIDs of 17.54 (MDC90) to 43.61 (MDC99). These SD values are consistent with previous research identifying an MCID of 9.1 (1/2 SD method) for the HOOS in an arthroplasty population[51]. The median of the HOOS JR MCIDs was 18.0 and may be an appropriate conservative estimate for evaluating treatment effects.

For the KOOS JR, the mean change ranged from 14.51 to 18.85. The ROC maximized MCID values ranged from 2.21 to 8.16. The SD values ranged from 3.98 (1/3 SD) to 8.02 (½ SD). The MDC method of detecting change yielded a range of MCIDs from 20.84 (MDC90) to 40.67 (MDC99). The MDC99 values reported here are actually lower than previous research on the KOOS JR which produced a range for improved individuals from 28.3 to 35.5[52]. The median KOOS JR MCID was 15.1 in this patient population.

The findings demonstrate that the method used

| Instrument | n | SD 1/2 | SD 1/3 | MDC 90% | MDC 95% | MDC 99% |
|------------|---|--------|--------|---------|---------|---------|
| 3-mo from baseline follow-up | | | | | | |
| PF CAT | 663 | 3.68 | 2.45 | 11.64 | 14.14 | 18.47 |
| HOOS JR | 99 | 5.86 | 3.9 | 17.54 | 20.84 | 27.22 |
| KOOS JR | 319 | 6.52 | 4.35 | 24.49 | 29.09 | 37.99 |
| > 3-mo from baseline follow-up | | | | | | |
| PF CAT | 2133 | 4.02 | 2.68 | 12.47 | 14.82 | 19.35 |
| HOOS JR | 245 | 9.33 | 6.22 | 28.11 | 33.39 | 43.61 |
| KOOS JR | 365 | 8.02 | 5.34 | 26.21 | 31.14 | 40.67 |
| 6-mo from baseline follow-up | | | | | | |
| PF CAT | 264 | 3.93 | 2.62 | 13.89 | 16.5 | 21.55 |
| HOOS JR | 31 | 10.96 | 7.3 | 26.36 | 31.31 | 40.89 |
| KOOS JR | 55 | 5.96 | 3.98 | 20.84 | 24.75 | 32.33 |
| > 6-mo from baseline follow-up | | | | | | |
| PF CAT | 1520 | 4.35 | 2.9 | 13.46 | 15.98 | 20.88 |
| HOOS JR | 112 | 10.73 | 7.15 | 28.86 | 31.9 | 41.67 |
| KOOS JR | 154 | 7.12 | 4.75 | 22.91 | 27.22 | 35.55 |

MDC: Minimum detectable change.
to estimate MCIDs has a large impact on MCID value determination. The KOOS JR, for example, had a 2 point to 40 point difference in MCID depending on the method used. There is not yet consensus on a standardized approach for establishing MCID\(^{[10]}\). The lack of agreement between MCID values reported in this study, depending on the method used, is consistent with findings in the literature\(^{[33]}\). There have been recommendations that MCID determination should be standardized, but the best methods have not been agreed upon\(^{[38]}\). Until a more standardized method is established, the comprehensive range of MCIDs presented in this study provided much deeper insights than many existing studies. This comprehensive presentation enables patients, clinicians, care-givers and decision makers to be well-positioned in making their judgment call as to which MCID value(s) they should select based on how lenient or strict a standard they would like to set for their patients. It is tempting for clinicians to oversimplify and search for a single fixed MCID value, yet a single MCID value is often unstable\(^{[39]}\). A range of MCID values such as those presented in this study should indeed be considered by clinicians and health care practitioners.

The study may be limited by the type of patients who self-selected to return for follow-up visits, which is a common phenomenon across all orthopaedic and other clinics. Patients returning for longer-term follow-up were generally those experiencing more severe conditions and thus may not reflect the full range of improvement in condition. However, this should not be of much concern, as these patients returning for follow-ups were representative of the ones treated regularly in clinics, rendering the results of this study even more practical for standard orthopaedic practice. MCID value determination is generally not dependent on the severity of condition, thus the shorter-term or longer-term follow-up periods would not have impacted the MCID values, as evidenced from the empirical results of this study. In addition, not every patient in this study completed the outcome measures pre-treatment, thus some change scores may reflect a baseline time-point that was post-intervention. Yet these evaluations still produced meaningful change since longitudinal change over time was the main focus, not necessarily change from a specific time point to another specific time point.

One additional limitation comes from the use of anchor-based questions to determine clinically relevant change. An anchor-based question with a global rating of change measure is subject to recall bias, with some research indicating that reports of change may be more related to the current health status than real change from baseline\(^{[10]}\). Different anchor questions may produce different results as well. Distribution methods also have limitations as they derive MCIDs based on the variance of the data, which might be difficult to interpret. Since all methods have strengths and limitations, it was thus our intention for this study to be comprehensive in nature to cover a variety of approaches, allowing readers to make informed decisions in selecting MCID values based on their personal value judgments as to which target MCID is worthwhile to pursue. Lastly, the demographic characteristics of this sample may not be representative of those in the United States and may affect MCID scores. Future research should investigate MCIDs in a more diverse demographic sample. It should also establish MCIDs linking baseline scores using Rasch methodology to provide deeper insights.

Overall, this study utilized rigorous methodologies to develop a wide range of MCID values for the PROMIS\(^{[53]}\) PF, HOOS JR, and KOOS JR in an orthopaedic sample of patients with joint disorders. As there is no such concept as a correct or incorrect MCID, individual value judgments are necessary to apply MCIDs to treatment planning and in guiding patient expectations of treatment change.

**ARTICLE HIGHLIGHTS**

**Research background**

Newly developed patient-reported outcomes have many advantages, but require further studies, including establishing minimum clinically important difference (MCID) values. Determining the MCID for the Patient-Reported Outcomes Measurement Information System Physical Function (PROMIS\(^{P}PF\)) will be useful for orthopaedic clinical practice and is helpful to understand the MCID in the context of previously used measures like the hip disability and osteoarthritis outcome score for joint reconstruction (HOOS JR), and the knee injury and osteoarthritis outcome score for joint reconstruction (KOOS JR). Anchor-based and distribution-based methods can both be used to determine MCID.

**Research motivation**

New instruments require studies to inform their score interpretation. Because of the lack of consensus on MCID methods, a comprehensive approach was taken, using both anchor- and distribution-based methods at multiple levels of precision and multiple follow-up time points. Cross verification of MCID values using powerful triangulation methods allow researchers and clinicians to understand the complexity of MCID evaluation and conscientiously select the most appropriate one for themselves.

**Research objectives**

To determine MCIDs for the PROMIS PF, HOOS JR and KOOS JR in a general joint orthopaedic patient population applying comprehensive methods.

**Research methods**

Consecutively enrolled patients aged 18 and older from a large academic orthopaedic joint clinic completed PROs at their first clinic visit and at follow-up points from 3-mo to 6-mo and beyond. These patients also completed an anchor question that queried how much their physical function had improved up points from 3-mo to 6-mo and beyond. They were grouped into change and no-change categories. Anchor-based analyses looked at mean change scores and the receiver operating curve to maximize the best cut-off based on sensitivity and specificity. Distribution-based analyses looked at the standard deviation, and minimum detectable change.

**Research results**

There were 2226 patients who participated with a mean age of 61.16 (SD = 12.84) years, 41.6% male, and 89.7% Caucasian. Mean change ranged from 7.29 to 8.41 for the PROMIS\(^{P}PF\) CAT; from 14.81 to 19.68 for the HOOS JR; and from 14.51 to 18.85 for the KOOS JR. ROC cut-offs ranged from 1.97-8.18 for the PF CAT, 6.33-43.36 for the HOOS JR, and 2.21-8.16 for the KOOS JR. Distribution-based methods estimated MCID values ranging from 2.45 to 21.55 for the PROMIS\(^{P}PF\) CAT; from 3.90 to 43.61 for the HOOS JR; and from 3.98 to 40.67 for the KOOS JR. The median MCID value in the range was similar to...
the mean change score for each measure and was 7.9 for the PF CAT, 18.0 for the HOOS JR, and 15.1 for the KOOS JR.

Research conclusions
Overall this study identified a large range of MCIDs for the PROMIS® PF, HOOS JR, and KOOS JR in an orthopaedic sample of patients with joint ailments. This range reflects the comprehensive strategies applied to determine MCIDs at varying levels of precision and cut off standards. The range of MCIDs presented in this study can be incorporated into decision making to guide treatment recommendations, compute sample size for research studies and clinical trials, and conduct respondent analyses.

Research perspectives
Decisions on which MCID value to select or which MCID value is useful should be based on an individual’s personal value and belief. Future research direction should focus on investigation of MCIDs with a more diverse demographic sample and to link MCIDs with baseline scores using Rasch-based methods.

REFERENCES
1 Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007; 45: S3-511 [PMID: 17443116 DOI: 10.1097/01.mmr.0000285615.4278.55]
2 Hassain FS, Konan S, Patel S, Rodriguez-Merchan EC, Haddad FS. The assessment of outcome after total knee arthroplasty: are we there yet? Bone Joint J 2015; 97-B: 3-9 [PMID: 25568406 DOI: 10.1302/0301-620x.97b1.34434]
3 Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Antmann D, Bode R, Buyse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pitkonis P, Revicki D, Rose M, Weinert K, Hays R; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010; 63: 1179-1194 [PMID: 20685078 DOI: 10.1016/j.jclinepi.2010.04.011]
4 Hung M, Nickisch F, Beals TC, Greene T, Clegg DO, Saltzman CL. New paradigm for patient-reported outcomes assessment in foot & ankle research: computerized adaptive testing. Foot Ankle Int 2012; 33: 621-626 [PMID: 22955227 DOI: 10.1113/ fai.2012.0620]
5 Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function Item Bank Was Calibrated to a Standardized Metric and Shown to Improve Measurement Efficiency. J Clin Epidemiol 2014; 67: 516-526 [PMID: 24698295 DOI: 10.1016/j.jclinepi.2013.10.024]
6 Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014; 41: 153-158 [PMID: 24241485 DOI: 10.3899/jrheum.130813]
7 Deutsch L, Smith L, Gage B, Kelleher C, Garfinkel D. Patient-reported outcomes in performance measurement: commissioned paper on PRO-based performance measures for healthcare accountable entities. Proceedings of the Washington, DC: National Quality Forum; 2012
8 Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: The Minimal Clinically Important Difference (MCID). J Man Manip Ther 2012; 20: 160-166 [PMID: 23904756 DOI: 10.1179/2042618612y.0000000001]
9 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989; 10: 407-415 [PMID: 2691207]
10 Cook CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. J Man Manip Ther 2008; 16: E82-E83 [PMID: 19771185 DOI: 10.1179/ jmt.2008.16.4.82E2]
11 Turner D, Schülemann JH, Griffith LE, Beaton DE, Griffiths AM, Critch JN, Guyatt GH. The minimal detectable change cannot reliably replace the minimal important difference. J Clin Epidemiol 2010; 63: 28-36 [PMID: 19800198 DOI: 10.1016/j.jclinepi.2009.01.024]
12 Hung M, Bousanga J, Voss MW. Interpretation of correlations in clinical research. Postgrad Med 2017; 129: 902-906 [PMID: 28936887 DOI: 10.1097/01PGM.0000000000000322]
13 MeC Glothin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014; 312: 1342-1343 [PMID: 25268441 DOI: 10.1001/jama.2014.13128]
14 White DK, Master H. Patient Reported Measures of Physical Function in Knee Osteoarthrit. Rheum Dis Clin North Am 2016; 42: 239-252 [PMID: 27133487 DOI: 10.1016/j.rdc.2016.01.005]
15 Thissen D, Liu Y, Magnus B, Quinn H, Gipson DS, Dampier C, Huang IC, Hinds PS, Selewski DT, Reeve BB, Gross HE, DeWalt DA. Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. Qual Life Res 2016; 25: 13-23 [PMID: 26187688 DOI: 10.1007/s11136-015-1058-8]
16 Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. J Clin Epidemiol 2011; 64: 507-516 [PMID: 21444727 DOI: 10.1016/j.jclinepi.2010.11.018]
17 Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)–development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28: 88-96 [PMID: 9699158 DOI: 10.2529/jospt.1998.28.2.88]
18 Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003; 1: 17 [PMID: 12801417 DOI: 10.1186/1477-7525-1-17]
19 Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review. Br J Sports Med 2010; 44: 1186-1196 [PMID: 19666629 DOI: 10.1136/bjsm.2009.068899]
20 Nilsson AE, Lohmander LS, Källbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)–validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003; 4: 10 [PMID: 12777182 DOI: 10.1186/1477-7242-4-10]
21 Ayers DC, Franklin PD. Joint replacement registries in the United States: a new paradigm. J Bone Joint Surg Am 2014; 96: 1576-1569 [PMID: 25232081 DOI: 10.2106/jbjs.06.00641]
22 Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: A Short-form Hip Replacement Survey. Clin Orthop Relat Res 2016; 474: 1472-1482 [PMID: 26296772 DOI: 10.1177/001126941664718-2]
23 Tait MA, Dredge C, Barnes CL. Preoperative Patient Education for Hip and Knee Arthroplasty: Financial Benefit? J Surg Orthop Adv 2015; 24: 246-251 [PMID: 26731389]
24 Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. Clin Orthop Relat Res 2016; 474: 1461-1471 [PMID: 26296773 DOI: 10.1007/s11999-016-4719-1]
25 Hung M, Saltzman CL, Greene T, Voss MW, Bousanga J, Gu Y, Andersson MB, Peters CL, Gilliland J, Pelt CE. Evaluating instrument responsiveness in joint function: The HOOS JR, the KOOS JR, and the PROMIS PF CAT. J Orthop Res 2017; Epub ahead of print [PMID: 28921658 DOI: 10.1002/jor.23739]
26 Paradowski PT, Kęska R, Witowski D. Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in patients with osteoarthritis undergoing total knee replacement. BMJ Open 2015; 5: e006047 [PMID: 26141360 DOI: 10.1136/bmjopen-2014-006047]
27 Paetelma M, Kilkikoski S, Simonen R, Heinonen A, Alen M, Videman T. Orthopaedic manual therapy, McKenzie method or advice only for low back pain in working adults: a randomized
controlled trial with one year follow-up. J Rehabil Med 2008; 40: 858-863 [PMID: 19242624 DOI: 10.2340/16501977-0262]

28 Uchiyama S, Imaeda T, Toh S, Kasunose K, Suzukiumi T, Wada T, Okinaga S, Nishida J, Onosaka S; Clinical Outcomes Committee of the Japanese Orthopaedic Association; Impairment Evaluation Committee of the Japanese Society for Surgery of the Hand. Comparison of responsiveness of the Japanese Society for Surgery of the Hand of the varus and the cupul tunnel syndrome instrument to surgical treatment with DASH, SF-36, and physical findings. J Orthop Sci 2007; 12: 249-253 [PMID: 17530376 DOI: 10.1007/s00776-007-1128-1]

29 Cournos MR, Silbernaegel KG, Nilsson-Helannder K, Mei-Dan O, Karlsson J, Maffulli N. Cross cultural adaptation of the Achilles tendon Total Rupture Score with reliability, validity and responsiveness evaluation. Knee Surg Sports Traumatol Arthrosc 2013; 21: 1356-1360 [PMID: 22847247 DOI: 10.1007/s00167-012-1216-8]

30 Landauer F, Wimmer C, Behensky H. Estimating the final outcome of brace treatment for idiopathic thoracic scoliosis at 6-month follow-up. Pediatr Rehabil 2003; 6: 201-207 [PMID: 14713586 DOI: 10.1080/13638490310001636817]

31 Little DG, MacDonald D. The use of the percentage change in Oswestry Disability Index score as an outcome measure in lumbar spinal surgery. Spine (Phila Pa 1976) 1994; 19: 2139-2143 [PMID: 7809744]

32 Cornell CN, Levine D, O’Doherty J, Lyden J. Unipolar versus bipolar hemiarthroplasty for the treatment of femoral neck fractures in the elderly. Clin Orthop Relat Res 1998; (348): 67-71 [PMID: 9553535]

33 Kotis SV, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand questionnaire in carpal tunnel surgery. J Hand Surg Am 2005; 30: 81-86 [PMID: 15680560 DOI: 10.1016/j.jhsa.2004.10.006]

34 MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Surg Am 2000; 25: 330-340 [PMID: 10722826 DOI: 10.1016/j.jsha.2000.03.030]

35 Ibrahim T, Beiri A, Azzabi M, Best AJ, Taylor GJ, Menon DK. Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. J Foot Ankle Surg 2007; 46: 65-74 [PMID: 17331864 DOI: 10.1053/j.jfas.2006.12.002]

36 Segal NA, Glass NA, Teran-Yengle P, Singh B, Wallace RB, Yack HJ. Intensive Gait Training for Older Adults with Symptomatic Knee Osteoarthritis. J Rheumatol 2014; 41: 509-515 [PMID: 24429183 DOI: 10.3899/jrheum.130609]

37 Gregory J, Haward D, Gochanour E, Sherman S, Romeo A. Clinical outcomes of revision biceps tenodesis. Int J Shoulder Surg 2012; 6: 45 [PMID: 2278333 DOI: 10.4103/0973-6042.96993]

38 Terwee CB, Roorda LD, Dekker J, Bijera-Neira SM, Peat G, Jordan KP, Croft P, de Vet HC. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010; 63: 524-534 [PMID: 19926446 DOI: 10.1016/j.cej.2009.08.010]

39 Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, Bouter LM, de Vet HC. 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 [PMID: 18165753 DOI: 10.1097/BR5.0b013e31815ea3a10]

40 Jones PW, Beech KM, Chapman KR, Decramer M, Mahler DA, Wedzicha JA. Minimal clinically important differences in pharmacological trials. Am J Respir Crit Care Med 2014; 189: 250-255 [PMID: 24383418 DOI: 10.1164/ajrccm.201310-1863PP]

41 Gershon RC, Rothrock N, Hanrahan R, Bass M, Cell D. The Use of PROMIS and Assessment Center to Deliver Patient-Reported Outcome Measures in Clinical Research. J Appl Meas 2010; 11: 304-314 [PMID: 20847477]

42 Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefer JF, Lohmander LS. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis Cartilage 2009; 17: 843-847 [PMID: 19215728 DOI: 10.1016/j.joca.2009.01.005]

43 Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. Man Ther 2010; 15: 404-407 [PMID: 20434942 DOI: 10.1016/j.math.2010.03.008]

44 Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E, Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). J Orthop Sports Phys Ther 2014; 44: 30-39 [PMID: 24175606 DOI: 10.2519/jospt.2014.4893]

45 Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. Qual Life Res 1996; 5: 35-46 [PMID: 8901365]

46 Youden WJ. Index for rating diagnostic tests. Cancer 1950; 3: 32-35 [PMID: 15405679]

47 IBM Corp. SPSS Statistics for Windows. Armonk, New York: IBM Corp., 2015

48 R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2010

49 de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, Boers M, Bouter LM. Minimally important change defined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007; 16: 131-142 [PMID: 17033901 DOI: 10.1007/s11136-006-9109-9]

50 Man-Son-Hing M, Laupacis A, O’Rourke K, Molnar FJ, Mahon J, Chan KB, Wells G. Determination of the clinical importance of study results. J Gen Intern Med 2002; 17: 469-476 [PMID: 1213163]

51 Berliner J, Brodkie DJ, Chan V, SoolHoo NF, Bozic KJ. John Charnley Award: Preoperative Patient-reported Outcome Measures Predict Clinically Meaningful Improvement in Function After THA. Clin Orthop Relat Res 2016; 474: 321-329 [PMID: 26201420 DOI: 10.1097/s1199-015-4350-6]

52 Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and Clinically Important Improvement Thresholds for Osteoarthritis Pain and Function scales: A Multicenter study. J Rheumatol 2014; 41: 509-515 [PMID: 24429183 DOI: 10.3899/jrheum.130609]

53 Beaton DE, van Eerd D, Smith P, van der Velde G, Cullen K, Kennedy CA, Hogg-Johnson S. Minimal change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. J Clin Epidemiol 2011; 64: 487-496 [PMID: 21109396 DOI: 10.1016/j.cej.2010.07.012]
