Clinical Research

What Is the Clinical Benefit of Common Orthopaedic Procedures as Assessed by the PROMIS Versus Other Validated Outcomes Tools?

Aditya V. Karhade MD, MBA[1,2] ⓘ, **David N. Bernstein MD, MBA, MEI[1,2]** ⓘ, **Vineet Desai BS[1]**,
Hany S. Bedair MD[1], **Evan A. O'Donnell MD[1]**, **Miho J. Tanaka MD[1]**, **Christopher M. Bono MD[1]**,
Mitchel B. Harris MD[1], **Joseph H. Schwab MD, MS[1]**, **Daniel G. Tobert MD[1]**

[1]Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[2]Harvard Combined Orthopaedic Residency Program, Boston, MA, USA

D. G. Tobert ✉, Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, MA 02114, USA, Email: dtobert@mgh.harvard.edu

Ⓦ Wolters Kluwer

Abstract

Background Patient-reported outcome measures (PROMs), including the Patient-reported Outcomes Measurement Information System (PROMIS), are increasingly used to measure healthcare value. The minimum clinically important difference (MCID) is a metric that helps clinicians determine whether a statistically detectable improvement in a PROM after surgical care is likely to be large enough to be important to a patient or to justify an intervention that carries risk and cost. There are two major categories of MCID calculation methods, anchor-based and distribution-based. This variability, coupled with heterogeneous surgical cohorts used for existing MCID values, limits their application to clinical care.

Questions/purposes In our study, we sought (1) to determine MCID thresholds and attainment percentages for PROMIS after common orthopaedic procedures using distribution-based methods, (2) to use anchor-based MCID values from published studies as a comparison, and (3) to compare MCID attainment percentages using PROMIS scores to other validated outcomes tools such as the Hip Disability and Osteoarthritis Outcome Score (HOOS) and Knee Disability and Osteoarthritis Outcome Score (KOOS).

Methods This was a retrospective study at two academic medical centers and three community hospitals. The inclusion criteria for this study were patients who were age 18 years or older and who underwent elective THA for osteoarthritis, TKA for osteoarthritis, one-level posterior lumbar fusion for lumbar spinal stenosis or spondylolisthesis, anatomic total shoulder arthroplasty or reverse total shoulder arthroplasty for glenohumeral arthritis or rotator cuff arthropathy, arthroscopic anterior cruciate ligament reconstruction, arthroscopic partial meniscectomy, or arthroscopic rotator cuff repair. This yielded 14,003 patients. Patients undergoing revision operations or surgery for nondegenerative pathologies and patients without preoperative PROMs assessments were excluded, leaving 9925 patients who completed preoperative PROMIS assessments and 9478 who completed other preoperative validated outcomes tools (HOOS, KOOS, numerical rating scale for leg pain, numerical rating scale for back pain, and QuickDASH). Approximately 66% (6529 of 9925) of patients had postoperative PROMIS scores (Physical Function,

Mental Health, Pain Intensity, Pain Interference, and Upper Extremity) and were included for analysis. PROMIS scores are population normalized with a mean score of 50 ± 10, with most scores falling between 30 to 70. Approximately 74% (7007 of 9478) of patients had postoperative historical assessment scores and were included for analysis. The proportion who reached the MCID was calculated for each procedure cohort at 6 months of follow-up using distribution-based MCID methods, which included a fraction of the SD (1/2 or 1/3 SD) and minimum detectable change (MDC) using statistical significance (such as the MDC 90 from p < 0.1). Previously published anchor-based MCID thresholds from similar procedure cohorts and analogous PROMs were used to calculate the proportion reaching MCID.

Results Within a given distribution-based method, MCID thresholds for PROMIS assessments were similar across multiple procedures. The MCID threshold ranged between 3.4 and 4.5 points across all procedures using the 1/2 SD method. Except for meniscectomy (3.5 points), the anchor-based PROMIS MCID thresholds (range 4.5 to 8.1 points) were higher than the SD distribution-based MCID values (2.3 to 4.5 points). The difference in MCID thresholds based on the calculation method led to a similar trend in MCID attainment. Using THA as an example, MCID attainment using PROMIS was achieved by 76% of patients using an anchor-based threshold of 7.9 points. However, 82% of THA patients attained MCID using the MDC 95 method (6.1 points), and 88% reached MCID using the 1/2 SD method (3.9 points). Using the HOOS metric (scaled from 0 to 100), 86% of THA patients reached the anchor-based MCID threshold (17.5 points). However, 91% of THA patients attained the MCID using the MDC 90 method (12.5 points), and 93% reached MCID using the 1/2 SD method (8.4 points). In general, the proportion of patients reaching MCID was lower for PROMIS than for other validated outcomes tools; for example, with the 1/2 SD method, 72% of patients who underwent arthroscopic partial meniscectomy reached the MCID on PROMIS Physical Function compared with 86% on KOOS.

Conclusion MCID calculations can provide clinical correlation for PROM scores interpretation. The PROMIS form is increasingly used because of its generalizability across diagnoses. However, we found lower proportions of MCID attainment using PROMIS scores compared with historical PROMs. By using historical proportions of attainment on common orthopaedic procedures and a spectrum of MCID calculation techniques, the PROMIS MCID benchmarks are realizable for common orthopaedic procedures. For clinical practices that routinely collect PROMIS scores in the clinical setting, these results can be used by individual surgeons to evaluate personal practice trends and by healthcare systems to quantify whether clinical care initiatives result in meaningful differences. Furthermore, these MCID thresholds can be used by researchers conducting retrospective outcomes research with PROMIS.

Level of Evidence Level III, therapeutic study.

Introduction

The National Institutes of Health's Patient-reported Outcomes Measurement Information System (PROMIS®) assessments are increasingly used for outcomes assessment in orthopaedic surgery and appear to be replacing first-generation assessment tools [10, 11, 14-18]. Previous studies demonstrate PROMIS assessment responsiveness in common orthopaedic conditions and correlate PROMIS with disease-specific measures such as the Hip Disability and Osteoarthritis Outcome Score (HOOS) [16, 17]. Beyond responsiveness, there is interest in understanding how to measure the value of a healthcare intervention [31]. The minimum clinically important difference (MCID) is one method. The MCID is an effect-size estimate; it represents the smallest increment of improvement after treatment that a patient is likely to consider important [20]. It is important to focus not merely on statistically detectable differences, but on effect sizes because patients perceive effect sizes after treatment (rather than p values) and knowing whether a treatment is likely to result in a clinically meaningful improvement is important when surgeons consider applying interventions that carry risk and cost. Two broad groups of MCID calculation methods exist: distribution-based approaches and anchor-based approaches. A distribution-based approach includes fractioning the SD (1/2 or 1/3 SD) of responses to set MCID thresholds. The growing data repositories of PROMIS scores across healthcare systems that lack anchor scores makes distribution-based methods a necessary alternative [32]. Yet, this mathematical approach to the MCID limits its clinical relevance. In contrast, an anchor-based MCID method uses a separate global rating of change questionnaire (anchor) to calculate MCID, which may improve clinical relevance but is statistically problematic [6].

The variation in MCID calculation techniques leads to a range of MCID values that are difficult to generalize across research settings and difficult to apply clinically. This phenomenon was demonstrated using other validated outcomes tools in a spine cohort where different MCID calculation methods yielded a fivefold difference in MCID values [7]. One advantage of PROMIS compared with historical patient-reported outcome measures (PROMS) is the normalization of scores across an entire population, with a mean score of 50 ± 10. Given that most scores fall between 30 and 70, this provides context for score interpretation [9]. Yet, the existing work on determining MCID values for PROMIS utilize heterogeneous cohorts

Wolters Kluwer

of surgical and nonsurgical orthopaedic patients, further limiting their use in clinical situations [18].

Therefore, the purposes of this study were: (1) to determine MCID thresholds and attainment percentages for PROMIS after common orthopaedic procedures using distribution-based methods, (2) to use anchor-based MCID values from published studies as a comparison, and (3) to compare MCID attainment percentages using PROMIS scores to other validated outcomes tools.

Patients and Methods

Study Design and Setting

Our institutional review board approved this retrospective study of electronic medical records from two tertiary care academic medical centers and three community hospitals.

Participants

The inclusion criteria for this study were age 18 years or older, elective surgery, completion of at least one PROM assessment in the 180 days before surgery, completion of at least one PROM assessment in the 210 days after surgery, and operative intervention with one-level posterior lumbar fusion (PLF) for lumbar spinal stenosis or spondylolisthesis, THA for osteoarthritis, TKA for osteoarthritis, total shoulder arthroplasty (TSA) or reverse

total shoulder arthroplasty for glenohumeral arthritis or rotator cuff arthropathy, arthroscopic anterior cruciate ligament reconstruction (ACLR), arthroscopic partial meniscectomy (APM), or arthroscopic rotator cuff repair (RCR). Revision operations and surgeries performed for traumatic (non–soft tissue), neoplastic, or infectious indications were excluded.

For PROMIS, percentages of postoperative assessment completion were 63% (1394 of 2196) for THA, 68% (1803 of 2650) for TKA, 70% (248 of 352) for PLF, 65% (313 of 478) for TSA, 76% (924 of 1223) for ACLR, 57% (1106 of 1951) for APM, and 69% (741 of 1075) for RCR; across all procedures, the median (interquartile range) of PROMIS scores improved from the preoperative to postoperative assessments (Table 1).

For other validated outcomes tools, percentages of postoperative assessment completion were 71% (1384 of 1960) for THA, 71% (1536 of 2167) for TKA, 75% (253 of 337) for PLF, 72% (445 of 614) for TSA, 85% (1000 of 1183) for ACLR, 70% (1281 of 1818) for APM, and 79% (1108 of 1399) for RCR; across all procedures, the median (interquartile range) of historical scores improved from the preoperative to postoperative assessments (Table 2).

We used the 6-month follow-up to compare MCID attainment across these procedures based on prior research. Manderle et al. [26] previously showed that mean achievement of MCID for RCR was approximately 6 months. Matar et al. [27] previously showed that patients undergoing reverse TSA reach maximal medical

Table 1. Comparison of preoperative and postoperative PROMIS Physical Function and Upper Extremity scores across THA, TKA, TSA, PLF, ACLR, APM, and RCR

Procedure	Assessment	Time	Number	Median (IQR)
THA	PROMIS PF	Preoperative	2196	40 (35-45)
		Postoperative	1394	48 (42-54)
TKA	PROMIS PF	Preoperative	2650	42 (37-45)
		Postoperative	1803	45 (40-51)
PLF	PROMIS PF	Preoperative	352	37 (32-42)
		Postoperative	248	45 (40-48)
TSA	PROMIS UE	Preoperative	478	30 (25-38)
		Postoperative	313	38 (29-44)
ACLR	PROMIS PF	Preoperative	1223	48 (42-54)
		Postoperative	924	51 (48-58)
APM	PROMIS PF	Preoperative	1951	45 (40-51)
		Postoperative	1106	48 (42-54)
RCR	PROMIS UE	Preoperative	1075	31 (28-41)
		Postoperative	741	40 (30-47)

TSA = total shoulder arthroplasty; PLF = posterior lumbar fusion; ACLR = anterior cruciate ligament reconstruction; APM = arthroscopic partial meniscectomy; RCR = rotator cuff repair; PROMIS PF = PROMIS Physical Function; PROMIS UE = PROMIS Upper Extremity.

Wolters Kluwer

Table 2. Comparison of preoperative and postoperative historical assessments across THA, TKA, TSA, PLF, ACLR, APM, and RCR

Procedure	Assessment	Time	Number	Median (IQR)
THA	HOOS	Preoperative	1960	58 (44-66)
		Postoperative	1384	80 (70-91)
TKA	KOOS	Preoperative	2167	56 (49-63)
		Postoperative	1536	66 (58-75)
PLF	Leg pain	Preoperative	337	7 (4-8)
		Postoperative	253	1 (0-4)
TSA	QuickDASH	Preoperative	614	50 (36-64)
		Postoperative	445	32 (16-55)
ACLR	KOOS	Preoperative	1183	65 (56-73)
		Postoperative	1000	75 (66-85)
APM	KOOS	Preoperative	1818	60 (52-66)
		Postoperative	1281	68 (58-78)
RCR	QuickDASH	Preoperative	1399	45 (30-59)
		Postoperative	1108	27 (14-48)

The HOOS and the KOOS range from 0 to 100, with lower scores indicating more severe symptoms; leg pain ranges from 0 to 10, with higher scores indicating more severe symptoms; QuickDASH ranges from 0 to 100, with higher scores indicating more severe symptoms. TSA = total shoulder arthroplasty; PLF = posterior lumbar fusion; ACLR = anterior cruciate ligament reconstruction; APM = arthroscopic partial meniscectomy; RCR = rotator cuff repair; HOOS = Hip Disability and Osteoarthritis Outcome Score; KOOS = Knee Disability and Osteoarthritis Outcome Score.

improvement at 6 months. Canfield et al. [5] previously studied the optimal collection window for PROMs in total joint arthroplasty and showed a consistent plateau in PROMs at 6 months postoperatively. Patients in the Spine Patient Outcomes Research Trial (SPORT) undergoing surgery for lumbar spondylolisthesis similarly showed a plateau in PROMs at 6 months postoperatively [35]. Beletsky et al. [2] showed that timing for achieving MCID in patients undergoing APM meniscectomy was approximately 6 months. Hill et al. [13] studied outcomes after ACLR and showed that Knee Disability and Osteoarthritis Outcome Score (KOOS) symptoms and pain subscale scores were a mean 86 and 90.5 at 6 months compared with 88.9 and 92.7 at 1 year, respectively.

Participants' Baseline Data

Baseline data are presented for patients who completed postoperative PROMIS assessments.

For THAs, the median (interquartile range) patient age was 66 years (59 to 73), 52% (721 of 1394) were women, 92% (1288 of 1394) were White, and 69% (962 of 1394) were married. For TKAs, the median age was 68 years (62 to 73), 58% (1039 of 1803) were women, 90% (1630 of 1803) were White, and 71% (1280 of 1803) were married. For PLF, the median age was 64 years (57 to 72), 57% (140 of 248) were women, 90% (224 of 248) were White, and 69% (171 of 248) were married. For TSA, the median age was 69 years (63 to 75), 47% (148 of 313) were women, 94% (295 of 313) were White, and 73% (228 of 313) were married. For ACLR, the median age was 31 years (24 to 43), 49% (451 of 924) were women, 80% (743 of 924) were White, and 37% (346 of 924) were married. For APM, the median age was 56 years (48 to 62), 47% (523 of 1106) were women, 87% (967 of 1106) were White, and 69% (760 of 1106) were married. For RCR, the median age was 60 years (53 to 66), 37% (275 of 741) were women, 92% (679 of 741) were White, and 71% (525 of 741) were married.

Primary and Secondary Study Outcomes

The primary outcome was the percentage of MCID attainment (based on distribution and anchor-based thresholds as described below) on National Institutes of Health PROMIS assessments. Each PROMIS assessment is population normalized with a mean score of 50 ± 10. Most PROMIS Physical Function and Upper Extremity scores range from 30 to 70, with higher scores indicating better function.

The secondary outcome was the percentage of MCID attainment on legacy PROM instruments. Assessments for patients undergoing PLF were the PROMIS Physical Function (PF) and the numerical rating scale for leg pain. Assessments for patients undergoing THA were the PROMIS Physical Function and the HOOS. Assessments for patients undergoing TKA were the PROMIS Physical Function and the KOOS. Assessments for those undergoing TSA and RCR were PROMIS Upper

Table 3. Comparison of MCID threshold and percentage of MCID attainment for PROMIS physical function and upper extremity across THA, TKA, TSA, PLF, ACLR, APM, and RCR

Procedure	Assessment	Definition	1/3 SD	1/2 SD	MDC 90	MDC 95	MDC 99	Anchor-based
THA (n = 1394)	PROMIS PF	Threshold	2.6	3.9	5.1	6.1	8	7.9
		Percentage	91% (1042 of 1143)	88% (944 of 1071)	85% (847 of 997)	82% (784 of 953)	75% (617 of 828)	76% (646 of 850)
TKA (n = 1803)	PROMIS PF	Threshold	2.4	3.6	4.7	5.6	7.4	7.9
		Percentage	89% (1205 of 1355)	81% (945 of 1168)	80% (933 of 1161)	73% (771 of 1050)	68% (665 of 975)	63% (568 of 905)
PLF (n = 248)	PROMIS PF	Threshold	2.3	3.4	4.5	5.4	7.1	7.4
		Percentage	91% (187 of 206)	84% (150 of 179)	84% (150 of 179)	79% (130 of 165)	78% (125 of 161)	76% (120 of 157)
TSA (n = 313)	PROMIS UE	Threshold	2.5	3.7	6.8	8	10.6	8.1
		Percentage	85% (191 of 225)	80% (172 of 215)	74% (150 of 204)	71% (141 of 200)	61% (110 of 179)	70% (140 of 199)
ACLR (n = 924)	PROMIS PF	Threshold	2.6	3.9	5.2	6.2	8.1	4.5
		Percentage	82% (539 of 654)	74% (446 of 601)	72% (414 of 577)	66% (362 of 552)	52% (255 of 492)	73% (431 of 590)
APM (n = 1106)	PROMIS PF	Threshold	2.5	3.8	5	5.9	7.8	3.5
		Percentage	80% (600 of 754)	72% (484 of 669)	70% (454 of 644)	63% (373 of 589)	56% (300 of 537)	74% (502 of 682)
RCR (n = 741)	PROMIS UE	Threshold	3	4.5	8.1	9.6	12.7	8.1
		Percentage	77% (400 of 518)	72% (358 of 497)	63% (287 of 457)	53% (234 of 438)	42% (171 of 412)	63% (287 of 457)

MDC 90 = 1.65∗ $\sqrt{2}$∗SEM; MDC 95 = 1.96∗ $\sqrt{2}$∗SEM; MDC 99 = 2.58∗ $\sqrt{2}$∗SEM; SEM = standard error of the mean; TSA = total shoulder arthroplasty; PLF = posterior lumbar fusion; ACLR = anterior cruciate ligament reconstruction; APM = arthroscopic partial meniscectomy; RCR = rotator cuff repair; please see Supplementary Fig. 1 (http://links.lww.com/CORR/A795) for the cohort size determination.

Extremity and the QuickDASH. Assessments for patients undergoing ACLR and APM were PROMIS Physical Function and KOOS.

Ethical Approval

Approval for this retrospective review of electronic medical records was obtained from our institutional review board.

Data Analysis

The study population was not stratified by gender for analysis because prior research for predicting likelihood of reaching the MCID did not identify gender as an independent prognostic factor [21].

MCID thresholds were calculated for each population with distribution-based methods. Distribution-based methods calculate the MCID by using the range and spread of baseline scores. For example, the 1/2 SD method was defined as half the SD of the baseline scores. Standard error of the mean was used to calculate the minimum detectable change (MDC), and the standard error of the mean was defined as SD * $\sqrt{1-r}$,

where "r" refers to the test-retest reliability coefficient and SD is defined as the SD of baseline scores. Using prior studies, we determined that test-retest reliability coefficients were 0.9 for the HOOS [25], 0.9 for the KOOS [25], 0.95 for the numeric rating scale score for back pain and leg pain [7], 0.9 for the QuickDASH [1, 28], 0.92 for PROMIS Physical Function [4], 0.90 for PROMIS Pain Intensity [4, 33], 0.91 for PROMIS Pain Interference [4, 33], 0.85 for PROMIS Global Mental Health [22], and 0.85 for PROMIS Upper Extremity [19, 36]. We calculated the distribution-based methods as follows:

A. 1/3 SD = SD/3
B. 1/2 SD = SD/2
C. MDC 90 = 1.65∗ $\sqrt{2}$∗SEM
D. MDC 95 = 1.96∗ $\sqrt{2}$∗SEM
E. MDC 99 = 2.58∗ $\sqrt{2}$∗SEM

Evidence-based Determination of Anchor-based MCIDs

A systematic search of the PubMed/MEDLINE databases was used to identify studies examining PROMIS assessments in orthopaedic surgery patients. Studies were

included in the anchor-based analyses if they calculated anchor-based thresholds for PROMIS Physical Function for THA, TKA, PLF, ACLR, and APM or the PROMIS Upper Extremity for TSA and RCR [3, 10-12, 16, 18, 23, 24, 29, 34, 37].

The percentage of patients reaching the MCID was defined as a change in the postoperative score relative to the preoperative score greater than the MCID threshold. Patients who had follow-up within 150 days and reached the MCID were classified as having reached the MCID. Patients who had follow-up within 150 days but had not reached the MCID were classified as having missing data (Supplementary Fig. 1; http://links.lww.com/CORR/A795). Anaconda Distribution (Anaconda Inc), R (The R Foundation), and Python (Python Software Foundation) were used for data analysis and application development.

Results

PROMIS MCID and Proportion of Patients Achieving MCID

Anchor-based thresholds for PROMIS Physical Function identified from previously published papers varied widely. Within a given distribution-based method, the MCID thresholds were similar across multiple procedures. For the PROMIS Physical Function, MCID thresholds ranged from 3.4 to 3.9 points for the 1/2 SD method and from 5.4 to 6.2 points for the MDC 95 method (Table 3). For anchor-based methods, MCID thresholds on PROMIS Physical Function ranged from 3.5 to 7.9 points. For PROMIS Upper Extremity, MCID thresholds ranged from 3.7 to 4.5 points for the 1/2 SD method and from 8 to 9.6 points for the MDC 95 method.

For THA, TKA, and PLF, the proportions of patients reaching the MCID on anchor-based thresholds were at the lower end of the range for distribution-based methods. In comparison, for ACLR and APM, the proportions of patients reaching the MCID on anchor-based thresholds were in the higher end of the range for distribution-based methods. The percentages of patients reaching the MCID for THA for distribution-based thresholds ranged from 75% to 91% compared with 76% with anchor-based thresholds. For TKA, 63% of patients reached the MCID with anchor-based thresholds compared with a range of 68% to 89% for distribution-based thresholds. For PLF, 76% of patients reached the MCID with anchor-based thresholds compared with a range of 78% to 91% with distribution-based methods. For TSA, 70% of patients reached the MCID with anchor-based thresholds compared with a range of 61% to 85% with distribution-based methods. For ACLR, 73% of patients reached the MCID with anchor-based thresholds compared with a range of 52% to 82% with distribution-based methods. For APM, 74% of patients reached the

MCID with anchor-based thresholds compared with a range of 56% to 80% with distribution-based methods. For RCR, 63% of patients reached the MCID with anchor-based thresholds compared with a range of 42% to 77% with distribution-based methods.

Proportion of Patients Achieving MCID on Historical Outcomes Tools

For all procedures other than TKA, higher percentages of patients reached the MCID with historical outcomes tools compared with PROMIS measures with distribution-based thresholds. With anchor-based thresholds, comparable percentages of patients reached the MCID on TKA and ACLR on PROMIS and historical tools; for all other procedures, similar to the trend for distribution-based thresholds, higher percentages of patients reached MCID with historical outcomes tools in comparison to PROMIS. For THA, 86% of patients reached the MCID with anchor-based thresholds compared with a range of 82% to 95% for distribution-based thresholds (Table 4). For TKA, 62% of patients reached the MCID with anchor-based thresholds compared with a range of 65% to 88% for distribution-based thresholds. For PLF, 84% of patients reached the MCID with anchor-based thresholds compared with a range of 84% to 93% with distribution-based methods. For TSA, 83% of patients reached the MCID with anchor-based thresholds compared with a range of 73% to 90% with distribution-based methods. For ACLR, 72% of patients reached the MCID with anchor-based thresholds compared with a range of 61% to 87% with distribution-based methods. For APM, 81% of patients reached the MCID with anchor-based thresholds compared with a range of 67% to 89% with distribution-based methods. For RCR, 78% of patients reached the MCID with anchor-based thresholds compared with a range of 53% to 87% with distribution-based methods.

Discussion

In the past decade, the PROMIS assessments have gained traction in orthopaedic surgery for clinical research, direct application into clinical care, and to measure healthcare value. Despite the growing enthusiasm for PROMIS measures, the available MCID values remain difficult to integrate into clinical care and research efforts because of heterogeneous pathologies or surgical interventions. Whereas other studies have studied MCID in all patients presenting to an outpatient orthopaedic surgery clinic regardless of diagnosis and operative or nonoperative management, we focused on including patients in this study who were from clinically relevant, operatively managed, elective orthopaedic populations. Compared with other

Wolters Kluwer

Table 4. Comparison of distribution and anchor-based MCID threshold and percentage of MCID attainment for PROMs assessments across THA, TKA, TSA, PLF, ACLR, APM, and RCR

Procedure	Assessment	Definition	1/3 SD	1/2 SD	MDC 90	MDC 95	MDC 99	Anchor-based
THA (n = 1384)	HOOS	Threshold	5.6	8.4	12.5	14.8	19.5	17.5
		Percentage	95% (1122 of 1182)	93% (1053 of 1128)	91% (958 of 1057)	88% (883 of 1000)	82% (717 of 879)	86% (814 of 948)
TKA (n = 1536)	KOOS	Threshold	4.7	7.1	10.5	12.5	16.4	17.5
		Percentage	88% (1022 of 1157)	85% (911 of 1073)	78% (767 of 978)	74% (682 of 925)	65% (533 of 821)	62% (497 of 797)
PLF (n = 253)	Leg pain	Threshold	0.9	1.3	1.4	1.6	2.2	2.1
		Percentage	93% (214 of 230)	87% (188 of 215)	87% (188 of 215)	87% (188 of 215)	84% (168 of 201)	84% (168 of 201)
TSA (n = 445)	QuickDASH	Threshold	5.9	8.9	13.1	15.6	20.5	13
		Percentage	90% (292 of 326)	87% (277 of 317)	83% (243 of 293)	79% (221 of 281)	73% (188 of 259)	83% (243 of 293)
ACLR (n = 1000)	KOOS	Threshold	5.3	7.9	11.7	13.9	18.3	13.3
		Percentage	87% (650 of 747)	82% (576 of 701)	76% (498 of 653)	71% (447 of 628)	61% (359 of 587)	72% (455 of 633)
APM (n = 1281)	KOOS	Threshold	5.1	7.6	11.2	13.3	17.5	10.7
		Percentage	89% (796 of 896)	86% (724 of 838)	80% (592 of 742)	75% (518 of 694)	67% (402 of 603)	81% (607 of 754)
RCR (n = 1108)	QuickDASH	Threshold	6.6	10	14.7	17.5	23	13
		Percentage	87% (694 of 799)	82% (617 of 757)	73% (521 of 710)	68% (470 of 688)	53% (335 of 628)	78% (567 of 732)

MDC = minimum detectable change; MDC 90 = $1.65 * \sqrt{2} * SEM$; MDC 95 = $1.96 * \sqrt{2} * SEM$; MDC 99 = $2.58 * \sqrt{2} * SEM$; SEM = standard error of the mean; TSA = total shoulder arthroplasty; PLF = posterior lumbar fusion; ACLR = anterior cruciate ligament reconstruction; APM = arthroscopic partial meniscectomy; RCR = rotator cuff repair; HOOS = Hip Disability and Osteoarthritis Outcome Score; KOOS = Knee Disability and Osteoarthritis Outcome Score; please see Supplementary Fig. 1 (http://links.lww.com/CORR/A795) for the cohort size determination.

validated outcomes tools, we found the percentage of patients who reached the MCID was lower when using PROMIS. This reflects the generalized nature of the PROMIS forms compared with the other validated outcomes tools that are specific to a pathology or anatomical region. The intersection of historical attainments with PROMIS thresholds provides clinically relevant MCID values. For example, the proportion of patients achieving an MCID on the non-PROMIS tools we assessed was about 85%, which corresponds to a 4-point change on PROMIS Physical Function (Table 3). Additionally, we found the PROMIS MCID thresholds to be similar across varying orthopaedic surgeries for a given MCID method. This consistency provides a benchmark for outcomes research using PROMIS.

Limitations

This study has several limitations. The study was limited to seven common orthopaedic procedures, and the addition of routine trauma, foot and ankle, and hand

procedures would offer a more holistic picture of the PROMIS in orthopaedic surgery. However, we contend the procedures in this study differ enough in anatomical region and surgical magnitude to be appropriately representative of the field. Furthermore, the lack of revision surgery and pathologies such as infection and tumor create a more homogenous cohort to answer the posed research questions. Our institution does not include a global rating of change questionnaire or other anchor instruments required for anchor-based MCID analyses. Therefore, values from prior publications were used to calculate anchor-based attainment percentages, which often included a more heterogenous cohort of patients. In the absence of our own anchor-based data, this represents the best-available alternative, and we propose the addition of the anchor-based analysis was more beneficial than omitting it. The study period included assessments up to 210 days, and it is possible further clinical improvement or deterioration can occur with time. This is particularly true for procedures with longer recovery periods, such as TKA and PLF. However, we believe our results retain validity as prior outcomes research demonstrates a plateau

in improvement for these procedures 6 months after surgery [2, 5, 13, 27, 28, 36]. Additionally, between 25% and 35% of patients did not have both pre- and postoperative scores, which introduces transfer bias. We rationalized using this period as a compromise between more loss to follow-up with a longer time and incomplete clinical improvement with a shorter time. Finally, the volume of questions (PROMIS and historical forms) and frequency with which patients are asked to complete them potentially introduces assessment bias in the form of response fatigue. However, the psychometric validation performed after the development of PROMIS minimizes this potential influence.

Discussion of Key Findings

The PROMIS MCID thresholds using a distribution-based approach ranged from 2 to 8 points (Table 3). This variability is unsurprising and reflects the variability in the mathematical formulae used to derive the values. However, the MCID threshold for each distribution-based method was within a single point regardless of the procedure performed (Table 3). This consistency underscores previously published psychometrics of PROMIS used in orthopaedic surgery. In isolation, these distribution-based MCID thresholds are difficult to apply clinically because they are merely a reflection of the underpinning mathematics and can be easily substituted to achieve higher or lower values. This is the crux of the criticism surrounding distribution-based methods for MCID calculation [6].

Yet, methodologic concerns exist for anchor-based calculations as the anchor is often an unvalidated questionnaire and the processing of anchor responses can lead to variable thresholds. Additionally, there are practical concerns emerging. Our healthcare enterprise has collected more than 10 million scores on PROMIS assessments, with other institutions collecting similar quantities of data. An intolerance for distribution-based methods will exclude large quantities of data that may help build on prior research with sample size limitations. Therefore, an effort to quantify how these different methods affect attainment percentages benefit future research and clinical care in orthopaedic surgery. By including only data on index procedures for degenerative pathology, these results are applicable to the most common clinical scenarios in orthopaedic surgery.

Comparing attainment percentages achieved by other validated outcomes tools to that of distribution-based and anchor-based methods helps bring that parity into focus. For example, most studies evaluating proportions of patients achieving the MCID after THA find that more than 85% of patients reach the MCID [8]. Therefore, using the distribution-based results, one can clinically translate a 3-to 4-point increase on PROMIS Physical Function as reaching the MCID (Table 2). The MCID thresholds calculated for other validated outcomes tools in this study were consistent with the findings of prior studies, providing external validity to these findings. Lyman et al. [25] found MCID thresholds for TKA using KOOS of 12 to 13 (MDC 90) and 14 to 16 (MDC 95), similar to those in our study (Table 4). Copay et al. [7] calculated a threshold of 1.3 SD on leg pain for lumbar spine conditions, equaling the threshold of 1.3 in this study. Moreover, comparative patterns for the highest MCID attainment among orthopaedic interventions were consistent with those of prior studies; for example, Rampersaud et al. [30] compared MCID attainment on the SF-36 in patients undergoing lumbar decompression, THA, and TKA and similarly observed that patients undergoing THA had the highest MCID attainment rate, followed by those undergoing lumbar decompression and TKA. The finding of lower MCID attainment percentages using anchor-based PROMIS thresholds might be explained by the availability of published anchor values. For example, the MCID threshold of 7.4 points was obtained from a study by Hung et al. [18]. This work calculated the MCID using data gathered at multiple spine clinic visits, regardless of operative or nonoperative intervention. A different value may be obtained if applied to patients undergoing a single type of spine surgery; however, it was the only anchor-based MCID threshold for PROMIS Physical Function we uncovered in spine publications.

The finding of lower MCID attainment using PROMIS compared with other validated outcomes tools may be explained by the generalizability of the PROMIS tool. Its development was heralded as a way to measure patient outcomes across diagnoses, but it is possibly less sensitive than other validated outcomes tools. These results suggest that MCID thresholds for PROMIS should be based partly on comparison studies that allow examination of the proportions of patients achieving the MCID with other validated outcome assessment tools. For example, the anchor-based QuickDASH indicated 83% (369 of 445) of patients reached the MCID after TSA, which would correspond to a PROMIS Upper Extremity MCID between 2.5 and 3.7 points.

Conclusion

The MCID thresholds calculated for multiple orthopaedic procedures using PROMIS assessments fell within a narrow range of 3 to 5 points when contextualized to other validated outcome assessment tools. These findings have applications for both research and clinical policy communities. First, the lower attainment percentage trend with PROMIS suggests consideration of other validated

outcomes tools during prospective trial design if outcome measure sensitivity is a concern. Second, for retrospective comparative research utilizing PROMIS, these MCID thresholds can be employed rather than simply using statistical significance to identify differences. From a policy standpoint, this research can be used to determine the impact of clinical care initiatives. With routine collection of PROMIS scores, a healthcare system can analyze score trends in the context of these results to assign clinical relevance to strategies seeking to improve healthcare value.

References

1. Beaton DE, Wright JG, Katz JN; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038-1046.
2. Beletsky A, Gowd AK, Liu JN, et al. Time to achievement of clinically significant outcomes after isolated arthroscopic partial meniscectomy: a multivariate analysis. *Arthrosc Sports Med Rehabil*. 2020;2:e723-e733.
3. Beletsky A, Naami E, Lu Y, et al. The minimally clinically important difference and substantial clinical benefit in anterior cruciate ligament reconstruction: a time-to-achievement analysis. *Orthopedics*. 2021;44:299-305.
4. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of Patient-Reported Outcomes Measurement Information System instruments in osteoarthritis. *Arthritis Care Res (Hoboken)*. 2013;65:1625-1633.
5. Canfield M, Savoy L, Cote MP, Halawi MJ. Patient-reported outcome measures in total joint arthroplasty: defining the optimal collection window. *Arthroplast Today*. 2020;6:62-67.
6. Copay AG. Commentary: the proliferation of minimum clinically important differences. *Spine J*. 2012;12:1129-1131.
7. Copay AG, Glassman SD, Subach BR, et al. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry disability index, medical outcomes study questionnaire short form 36, and pain scales. *Spine J*. 2008;8:968-974.
8. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res*. 2019;477:1267-1279.
9. Franovic S, Gulledge CM, Kuhlmann NA, et al. Establishing normal Patient-Reported Outcomes Measurement Information System physical function and pain interference scores: a true reference score according to adults free of joint pain and disability. *JB JS Open Access*. 2019;4:e0019.
10. Franovic S, Kuhlmann NA, Pietroski A, et al. Preoperative patient-centric predictors of postoperative outcomes in patients undergoing arthroscopic meniscectomy. *Arthroscopy*. 2021;37:964-971.
11. Gordon D, Pines Y, Ben-Ari E, et al. Minimal clinically important difference, substantial clinical benefit, and patient acceptable symptom state of PROMIS upper extremity after total shoulder arthroplasty. *JSES Int*. 2021;5:894-899.
12. Gowd AK, Lalehzarian SP, Liu JN, et al. Factors associated with clinically significant patient-reported outcomes after primary arthroscopic partial meniscectomy. *Arthroscopy*. 2019;35:1567-1575:e1563.
13. Hill GN, O'Leary ST. Anterior cruciate ligament reconstruction: the short-term recovery using the Knee Injury and Osteoarthritis Outcome Score (KOOS). *Knee Surg Sports Traumatol Arthrosc*. 2013;21:1889-1894.
14. Horn ME, Reinke EK, Couce LJ, et al. Reporting and utilization of Patient-Reported Outcomes Measurement Information System®(PROMIS®) measures in orthopedic research and practice: a systematic review. *J Orthop Surg Res*. 2020;15:1-13.
15. Hung M, Baumhauer JF, Latt LD, et al. Validation of PROMIS® physical function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res*. 2013;471:3466-3474.
16. Hung M, Bounsanga J, Voss MW, Saltzman CL. 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. *World J Orthop*. 2018;9:41-49.
17. Hung M, Saltzman CL, Greene T, et al. Evaluating instrument responsiveness in joint function: the HOOS Jr, the KOOS Jr, and the PROMIS PF CAT. *J Orthop Res*. 2018;36:1178-1184.
18. Hung M, Saltzman CL, Kendall R, et al. What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? *Clin Orthop Relat Res*. 2018;476:2027-2036.
19. Iyer S, Koltsov JC, Steinhaus M, et al. A prospective, psychometric validation of National Institutes Of Health Patient-Reported Outcomes Measurement Information System physical function, pain interference, and upper extremity computer adaptive testing in cervical spine patients: successes and key limitations. *Spine (Phila Pa 1976)*. 2019;44:1539-1549.
20. Karhade AV, Bono CM, Schwab JH, Tobert DG. Minimum clinically important difference: a metric that matters in the age of patient-reported outcomes. *J Bone Joint Surg Am*. 2021;103:2331-2337.
21. Karhade AV, Sisodia RC, Bono CM, et al. Surgeon-level variance in achieving clinical improvement after lumbar decompression: the importance of adequate risk adjustment. *Spine J*. 2021;21:405-410.
22. Kasturi S, Szymonifka J, Burket JC, et al. Feasibility, validity, and reliability of the 10-item patient reported outcomes measurement information system global health short form in outpatients with systemic lupus erythematosus. *J Rheumatol*. 2018;45:397-404.
23. Koorevaar RCT, Kleinlugtenbelt YV, Landman EBM, van 't Riet E, Bulstra SK. Psychological symptoms and the MCID of the DASH score in shoulder surgery. *J Orthop Surg Res*. 2018;13:246.
24. Kuo AC, Giori NJ, Bowe TR, et al. Comparing methods to determine the minimal clinically important differences in patient-reported outcome measures for veterans undergoing elective total hip or knee arthroplasty in veterans health administration hospitals. *JAMA Surg*. 2020;155:404-411.
25. Lyman S, Lee Y-Y, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res*. 2018;476:2432.
26. Manderle BJ, Gowd AK, Liu JN, et al. Time required to achieve clinically significant outcomes after arthroscopic rotator cuff repair. *Am J Sports Med*. 2020;48:3447-3453.
27. Matar RN, Gardner TJ, Kassam F, Grawe BM. When do patients truly reach maximal medical improvement after undergoing reverse shoulder arthroplasty? The incidence and clinical significance of pain and patient-reported outcome measure improvement. *JSES Int*. 2020;4:675-679.
28. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the shortened disabilities of the arm, shoulder, and hand

questionnaire (QuickDASH) and numeric pain rating scale in patients with shoulder pain. *J Shoulder Elbow Surg*. 2009;18: 920-926.

29. Parker SL, Adogwa O, Paul AR, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. *J Neurosurg Spine*. 2011;14:598-604.

30. Rampersaud YR, Wai EK, Fisher CG, et al. Postoperative improvement in health-related quality of life: a national comparison of surgical treatment for focal (one-to two-level) lumbar spinal stenosis compared with total joint arthroplasty for osteoarthritis. *Spine J*. 2011;11:1033-1041.

31. Safran DG, Higgins A. Getting to the next generation of performance measures for value-based payment. Health Affairs Blog. https://www.healthaffairs.org/do/10.1377/forefront.20190128.477681/full/. Accessed April 29, 2022.

32. Sisodia RC, Dankers C, Orav J, et al. Factors associated with increased collection of patient-reported outcomes within a large health care system. *JAMA Netw Open*. 2020;3:e202764.

33. Stephan A, Mainzer J, Kümmel D, Impellizzeri FM. Measurement properties of PROMIS short forms for pain and function in orthopedic foot and ankle surgery patients. *Qual Life Res*. 2019;28: 2821-2829.

34. van Kampen DA, Willems WJ, van Beers LW, et al. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMS). *J Orthop Surg Res*. 2013;8:40.

35. Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical compared with nonoperative treatment for lumbar degenerative spondylolisthesis. Four-year results in the Spine Patient Outcomes Research Trial (SPORT) randomized and observational cohorts. *J Bone Joint Surg Am*. 2009;91:1295-1304.

36. Wilkinson JT, Clawson JW, Allen CM, et al. Reliability of telephone acquisition of the PROMIS upper extremity computer adaptive test. *J Hand Surg Am*. 2021;46:187-199.

37. Yedulla NR, Tramer JS, Koolmees DS, et al. Preoperative Patient-Reported Outcomes Measurement Information System computerized adaptive testing (PROMIS CAT) scores predict achievement of minimum clinically important difference following anterior cruciate ligament reconstruction using an anchor-based methodology. *Arthrosc Sports Med Rehabil*. 2021;3:e1891-e1898.