**Clinical Orthopaedics
and Related Research®**
A Publication of The Association of Bone and Joint Surgeons®

Systematic Review and Meta-analysis

There are Considerable Inconsistencies Among Minimum Clinically Important Differences in TKA: A Systematic Review

David G. Deckey MD[1], Jens T. Verhey MD[1], Coltin R. B. Gerhart MS[2], Zachary K. Christopher MD[1], Mark J. Spangehl MD[1], Henry D. Clarke MD[1], Joshua S. Bingham MD[1]

Abstract

Background Patient-reported outcome measures (PROMs) are frequently used to assess the impact of total knee arthroplasty (TKA) on patients. However, mere statistical comparison of PROMs is not sufficient to assess the value of TKA to the patient, especially given the risk profile of arthroplasty. Evaluation of treatment effect sizes is important to support the use of an intervention; this is often quantified with the minimum clinically important difference (MCID). MCIDs are unique to specific PROMs, as they vary by calculation methodology and study population. Therefore, a systematic review of calculated MCID values, their respective ranges, and assessment of their applications is important to guide and encourage their use as a critical measure of effect size in TKA outcomes research.

Questions/purposes In this systematic review of MCID calculations and reporting in primary TKA, we asked: (1) What are the most frequently reported PROM MCIDs and their reported ranges in TKA? (2) What proportion of studies report distribution- versus anchor-based MCID values? (3) What are the most common methods by which these MCID values are derived for anchor-based values? (4) What are the most common derivation methods for distribution-based values? (5) How do the reported medians and corresponding interquartile ranges (IQR) compare between calculation methods for each PROM?

Methods Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, a systematic review was conducted using the PubMed, EMBASE, and MEDLINE databases from inception through March 2022 for TKA articles reporting an MCID value for any PROMs. Two independent reviewers screened articles for eligibility, including any article that calculated new MCID values for PROMs after primary TKA, and extracted these data for analysis. Overall, 576 articles were identified, 38 of which were included in the final analysis. These studies had a total of 710,128 patients with a median age of 67.7 years and median BMI of 30.9 kg/m^2. Women made up more than 50% of patients in most studies, and the median follow-up period was 17 months (range 0.25 to 72 months). The overall risk of bias was assessed as moderate using the Jadad criteria for one randomized controlled trial (3 of 5 ideal global score) and the modified Methodological Index for Non-randomized Studies criteria for comparative studies (mean 17.2 ± 1.8) and noncomparative studies (mean 9.6 ± 1.3). There were 49 unique PROMs for which 233

The first two authors contributed equally to this work.

[1]Department of Orthopaedic Surgery, Mayo Clinic Arizona, Phoenix, AZ, USA

[2]Texas Christian University School of Medicine, Fort Worth, TX, USA

D. G. Deckey ✉, Department of Orthopaedic Surgery, Mayo Clinic Arizona, 5777 East Mayo Boulevard, Phoenix, AZ 85054, USA, Email: deckey.david@mayo.edu

Ⓦ Wolters Kluwer

MCIDs were reported. Calculated values were classified as anchor-based, distribution-based, or not reported. MCID values for each PROM, MCID calculation method, number of patients, and study demographics were extracted from each study. Anchor-based and distribution-based MCIDs were compared for each unique PROM using a Wilcoxon rank sum test given non-normal distribution of values.

Results The WOMAC Function and Pain subscores were the most frequently reported MCID value, comprising 9% (22 of 233) and 9% (22 of 233), respectively. The composite Oxford Knee Score (OKS) was the next most frequently reported (9% [21 of 233]), followed by the WOMAC composite score (6% [13 of 233]). The median anchor-based values for WOMAC Function and Pain subscores were 23 (IQR 16 to 33) and 25 (IQR 14 to 31), while the median distribution-based values were 11 (IQR 10.8 to 11) and 22 (IQR 17 to 23), respectively. The median anchor-based MCID value for the OKS was 6 (IQR 4 to 7), while the distribution-based value was 7 (IQR 5 to 10). Thirty-nine percent (15 of 38) used an anchor-based method to calculate a new MCID, while 32% (12 of 38) used a distribution-based technique. Twenty-nine percent of studies (11 of 38) calculated MCID values using both methods. For studies reporting an anchor-based calculation method, a question assessing patient satisfaction, pain relief, or quality of life along a five-point Likert scale was the most commonly used anchor (40% [16 of 40]), followed by a receiver operating characteristic curve estimation (25% [10 of 40]). For studies using distribution-based calculations, all articles used a measure of study population variance in their derivation of the MCID, with the most common method reported as one-half the standard deviation of the difference between preoperative and postoperative PROM scores (45% [14 of 31]). Most reported median MCID values (15 of 19) did not differ by calculation method for each unique PROM (p > 0.05) apart from the WOMAC Function component score and the Knee Injury and Osteoarthritis Outcome Score Pain and Activities of Daily Living subscores.

Conclusion Despite variability of MCIDs for each PROM, there is consistency in the methodology by which MCID values have been derived in published studies. Additionally, there is a consensus about MCID values regardless of calculation method across most of the PROMs we evaluated.

Clinical Relevance Given their importance to treatment selection and patient safety, authors and journals should report MCID values with greater consistency. We recommend using a 7-point increase as the MCID for the OKS, consistent with the median reported anchor-based value derived from several high-quality studies with large patient groups that used anchor-based approaches for MCID calculation, which we believe are most appropriate for most applications in clinical research. Likewise, we recommend

using a 10-point to 15-point increase for the MCID of composite WOMAC, as the median value was 12 (IQR 10 to 17) with no difference between calculation methods. We recommend use of median reported values for WOMAC function and pain subscores: 21 (IQR 15 to 33) and 23 (IQR 13 to 29), respectively.

Introduction

To determine whether a procedure is truly worthwhile, the clinician must ensure that the procedure creates a benefit and whether this benefit is appreciable enough to the patient to justify surgical risk. Frequentist statistics help the clinician to determine whether a treatment is responsible for an effect, but they do not provide information on the magnitude of its effect size. However, the minimum clinically important difference (MCID) is one metric surgeons and researchers can use to identify meaningful change in a patient-perceived outcome. The MCID was first described by Jaeschke et al. [26] as the "smallest difference in a score in the domain of interest which patients perceive as beneficial." The MCID is useful because it defines a threshold for which a change in a patient-reported outcome measure (PROM) is clinically relevant to a patient, and not just of statistical significance. Several methods exist to calculate the MCID, most notably the distribution-based and anchor-based approaches [13, 14]. Distribution-based MCID calculations analyze change in PROM values with a measure of variability, such as the standard deviation or effect size [13, 14]. Anchor-based MCID calculation techniques use an external or secondary subjective patient outcome measure to establish a clinically important difference. Both approaches have been found to be reliable; however, the anchor-based approach is preferred, because the MCID value is calculated using secondary data provided by the patient rather than estimating the value from a measure of variance [39, 43]. Distribution-based approaches are applicable for large datasets when obtaining an anchor value from each patient is not feasible.

Although the MCID is universally accepted as a valuable measure of effect size, the MCID has limitations. Researchers are often relegated to one MCID calculation method based on study size and practicality. Furthermore, variability exists among studies regarding MCIDs, even for the same PROM. This is likely related to a number of factors, including differing patient populations, interventions, follow-up periods, and calculation methods [13, 14]. This same trend of heterogeneity has been described in shoulder arthroplasty [32]. Finally, the aim of arthroplasty is to create large difference in patient outcomes rather than a minimal change, especially considering its major associated risks. Alternatives to the MCID include the patient-acceptable symptom state, a metric assessing

whether the patient is satisfied with his or her current symptoms, and the substantial clinical benefit, which quantifies the outcome improvement needed for a patient to feel substantially better [4]. Despite these limitations, a direct focus on the effect size of treatment in addition to statistical significance is a critical step forward in value-based and patient-centric care [35, 36].

The goal of this systematic review was to consolidate the existing data on newly calculated MCID values for all available PROMs in TKA because, to our knowledge, this has not been done before. In doing so, we aimed to provide future TKA outcomes researchers with a repository of calculated MCIDs for each PROM and their respective ranges, compare the values by their derivation method, and guide effective use of MCIDs with respect to feasibility and practicality.

In this systematic review of MCID calculations and reporting in primary TKA, we asked: (1) What are the most frequently reported PROM MCIDs and their reported ranges in TKA? (2) What proportion of studies report distribution- versus anchor-based MCID values? (3) What are the most common methods by which these MCID values are derived for anchor-based values? (4) What are the most common derivation methods for distribution-based values? (5) How do the reported medians and corresponding interquartile ranges (IQR) compare between calculation methods for each PROM?

Materials and Methods

This systematic review was conducted using Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [44]. PROSPERO was queried for ongoing or unpublished reviews on this topic.

Eligibility Criteria

Studies were included if they calculated new MCID values for PROMs after primary TKA from their study population. Only peer-reviewed articles in the English language were included. Studies in cadavers, animal studies, revision studies, studies on unicompartmental arthroplasty, systematic reviews, meta-analyses, letters to the editor, machine learning studies, technique articles, basic science articles, case reports, conference proceedings without full-text articles, and nonpeer-reviewed preprint server articles were excluded.

Search Strategy

The PubMed, EMBASE, and MEDLINE databases were queried from inception through March 2022. The search

was conducted by two independent reviewers (DGD and JTV); no discrepancies required review by the senior author (JSB). Medical subject headings or keywords were used in varying combinations to broaden the search. The search terms were as follows: (MCID and total knee arthroplasty) or (MCID and TKA) or (minimal clinically important difference and total knee arthroplasty) or (minimal clinically important difference and TKA) or (minimal important difference and TKA) or (minimal important difference and total knee arthroplasty). Reviewers also manually searched the reference lists of the included articles (Fig. 1).

Selection Process and Studies Included

Overall, 576 articles were identified (Fig. 1). The search of the three databases returned 570 articles, and six were identified after a manual search of included articles' references. After removal of duplicates, 313 remained. Preliminary screening of titles and abstracts resulted in the exclusion of 124 articles, based on a priori established inclusion and exclusion criteria. The full texts of 174 articles were then reviewed. Of these, 94 articles cited previously reported MCID values and were excluded, and 38 calculated new MCID values. These 38 studies were included in the final analysis. Eleven studies reported on both TKA and THA PROM MCIDs. Only PROMs specific to patients undergoing TKA were included.

Data Extraction

In the included studies, the technique for calculation was identified as distribution-based, anchor-based, or not reported [13, 14]. Distribution-based MCID calculations analyze the change in PROM values with a measure of variability, such as the standard deviation or effect size [13, 14]. Two frequently used distribution-based techniques include taking half of the standard deviation of the measured change or using the percentage from baseline change method. Anchor-based MCID calculation techniques use an external or secondary subjective patient outcome measure to establish a clinically important difference. With anchor-based methods, even after an anchor has been chosen, the MCID can be calculated in varying ways. Some commonly used methods include the mean change method, the change difference method, or the receiver operating characteristic curve method. Studies reporting outcomes after both THA and TKA were included if their data were separated by procedure—data extraction was limited to only patients undergoing TKA and their respective PROMs and MCIDs. Study characteristics including the type of procedure, indication for the procedure, number of patients, mean age, gender, mean BMI, and length of follow-up were also
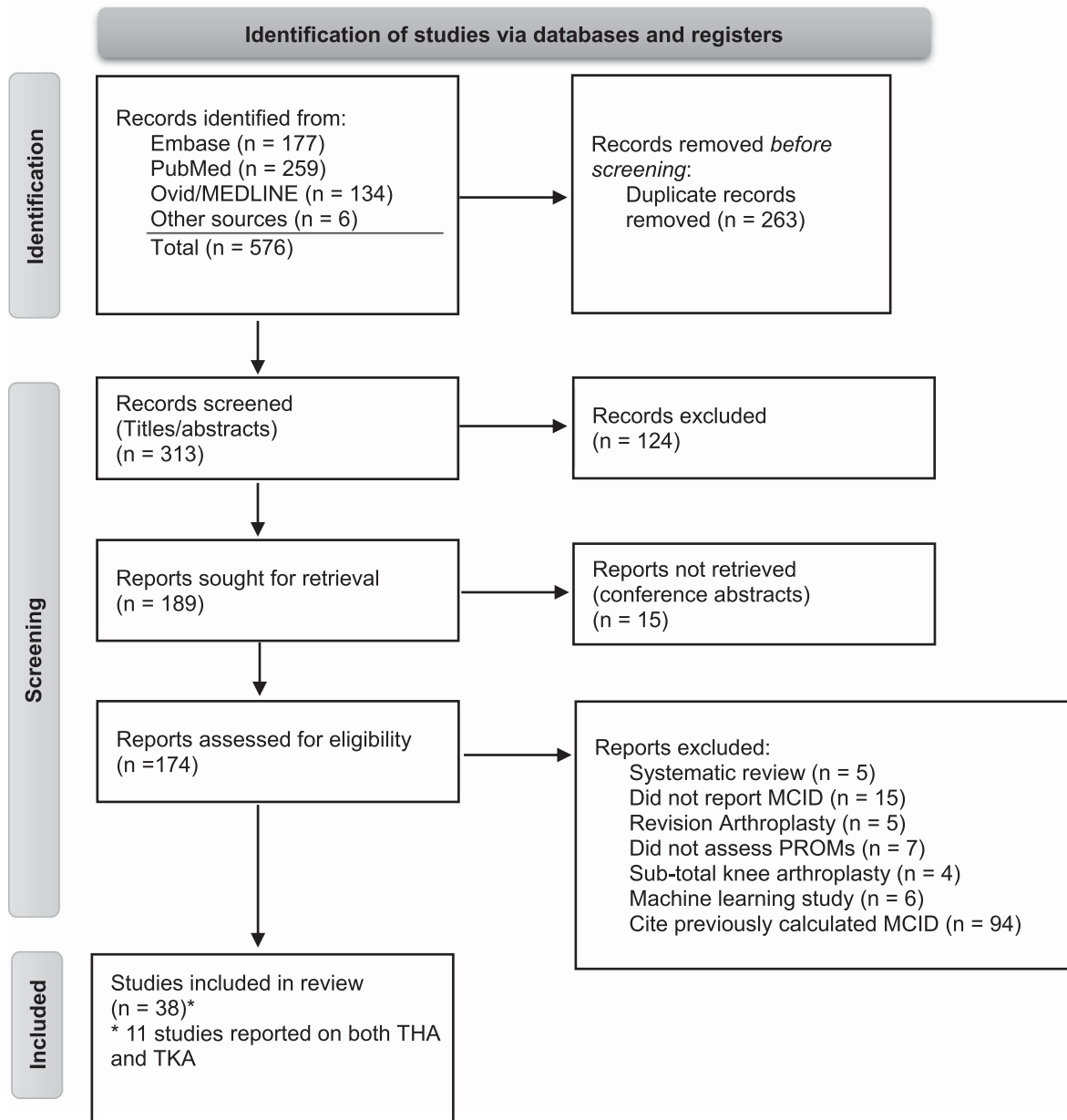
Identification of studies via databases and registers

Identification

| Records identified from:
    Embase (n = 177)
    PubMed (n = 259)
    Ovid/MEDLINE (n = 134)
    Other sources (n = 6)
    Total (n = 576) | → | Records removed *before*
screening:
    Duplicate records
    removed (n = 263) |

Screening

| Records screened
(Titles/abstracts)
(n = 313) | → | Records excluded
(n = 124) |

| Reports sought for retrieval
(n = 189) | → | Reports not retrieved
(conference abstracts)
(n = 15) |

| Reports assessed for eligibility
(n =174) | → | Reports excluded:
    Systematic review (n = 5)
    Did not report MCID (n = 15)
    Revision Arthroplasty (n = 5)
    Did not assess PROMs (n = 7)
    Sub-total knee arthroplasty (n = 4)
    Machine learning study (n = 6)
    Cite previously calculated MCID (n = 94) |

Included

| Studies included in review
(n = 38)*
* 11 studies reported on both THA
and TKA |

Fig. 1 This Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart represents our systematic search of the three databases.

extracted. Median MCID values with their respective interquartile ranges were reported because of a skewed distribution with outliers.

Study Characteristics

The 38 included studies had a total of 710,128 patients (Table 1). There was a median of 575 patients (range 157 to 347,536 patients) in each study, with a median age of 68 years and median BMI of 31 kg/m^2. Women made up 53%

of patients in most studies (median 337 patients), and the median follow-up period was 17 months (range 0.25 to 72 months). Overall, there were 49 unique PROMs for which 233 MCIDs were reported (Fig. 2).

Study Quality and Risk of Bias Assessment

The Methodological Index for Non-randomized Studies criteria for noncomparative and comparative studies was used to evaluate the risk of bias [47]. Additionally,

Table 1. Summary of the demographics of all included studies (n = 38)

Parameter	Value
Total patients, n	710,128
Overall proportion of women, % (n)	53 (376,142)
Number of patients, median (range)	575 (157-347,536)
Age in years, median (range)	68 (63-75)
BMI in kg/m^2, median (range)	31 (26-34)
Median follow-up in months, median (range)	17 (0.25-72)

the Jadad criteria were used to assess the quality of randomized clinical trials [25]. The overall risk of bias was assessed as moderate using the Jadad criteria for one randomized controlled trial (3 of 5 ideal global score) and the modified Methodological Index for Non-randomized Studies criteria for comparative studies (mean 17.2 ± 1.8) and noncomparative studies (mean 9.6 ± 1.3). The Jadad criteria score was 3 of an ideal global score 5 for the single randomized controlled trial in the review.

Primary and Secondary Study Outcomes

Our primary study goal was to review and assess calculated MCID values for each PROM after TKA with their respective ranges. To this end, the MCIDs were extracted from each study that calculated new values based on their study population. Our secondary goals were to describe the corresponding derivation methods of MCIDs used in published studies and determine whether any differences existed between these values. MCIDs were stratified by calculation method, including distribution-based, anchor-based, or other method, and their respective medians and ranges were compared.

Statistical Analysis

Continuous data such as the number of patients, age, number of women, and length of follow-up and MCID values are described using median and range. Categorical variables, including PROMs reported and MCID methods, are described using percentages. Anchor-based and distribution-based MCIDs were compared for each unique PROM using a Wilcoxon rank sum test after determining they followed non-normal distribution of values. Analyses were performed in Excel (Microsoft Corp) and JASP (JASP Team, 2022; Version 0.16.1) [50].

Results

Most-reported MCIDs and Ranges

The WOMAC Physical Function and Pain subscores were the most frequently reported MCID values, comprising 9% (22 of 233) and 9% (22 of 233), respectively. The WOMAC is a 24-item survey divided into three subscales used to evaluate hip and knee osteoarthritis (Table 2). The median anchor-based values for WOMAC Function and Pain subscores were 23 (IQR 16 to 33) and 25 (IQR 14 to 31), while the median distribution-based values were 11 (IQR 10.8 to 11) and 22 (IQR 17 to 23), respectively. The WOMAC median composite score was 15 (IQR 10 to 23) by the anchor-based method and 10 (IQR 10 to 11) by the distribution-based method. Four studies calculated an MCID value of 10 for the WOMAC composite score via both derivation methods (Table 3). The next most-frequently reported PROM MCID was for the Oxford Knee Score (OKS), comprising 9% (21 of 233) of reported values (Fig. 2). The OKS is 12-item survey designed to assess function and pain after TKA and was originally developed by the University of Oxford (Table 2). The median anchor-based MCID value for the OKS was 6 (IQR 4 to 7), while the distribution-based value was 7 (IQR 5 to 10) (Table 4). Three studies reported a calculated value of 5, two of which were derived from the distribution-based method and one from the anchor-based method. The median pooled MCIDs of Knee Society Score (KSS) components of function and knee were 6 (IQR 6 to 9) and 7 (IQR 5 to 7), respectively. The Knee Injury and Osteoarthritis Outcome Score (KOOS) components addressing Pain, Quality of Life (QoL), and Activities of Daily Living (ADL) demonstrated pooled MCIDs of 12 (IQR 10 to 20), 13 (IQR 11 to 13), and 12 (IQR 10 to 14), respectively. Finally, the combined KOOS JR MCID was 9 (IQR 7 to 12) (Table 4).

Proportion of Studies Reporting Distribution- versus Anchor-based MCIDs

Thirty-nine percent (15 of 38) of studies used an anchor-based method to calculate a new MCID, while 32% (12 of 38) used a distribution-based technique (Table 3). Twenty-nine percent of studies (11 of 38) calculated MCID values using both methods. No other calculation methods were reported.

Derivation Methods of Anchor-based MCIDs

For studies reporting an anchor-based calculation method, a question assessing patient satisfaction, pain relief, or quality
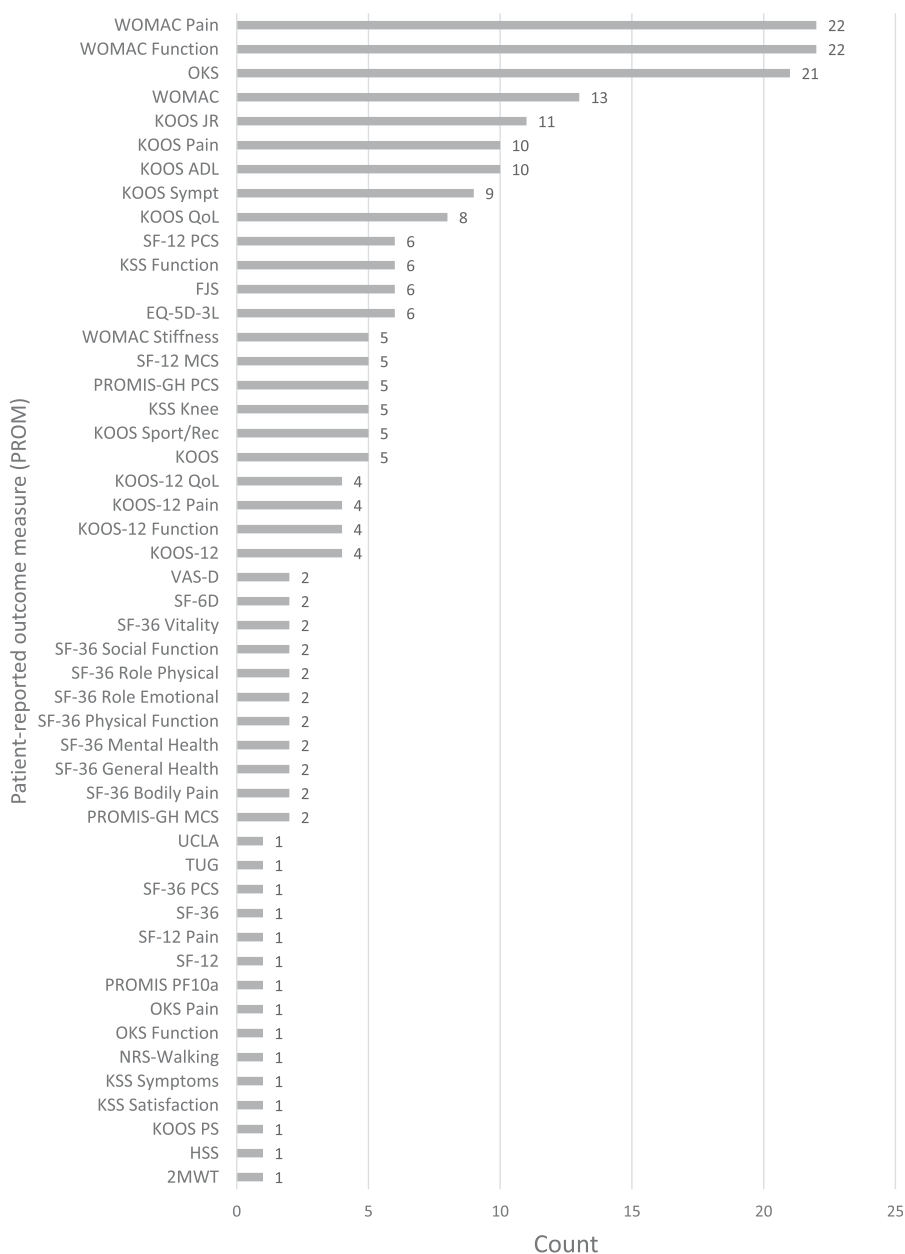
Wolters Kluwer

Fig. 2 This graph shows the frequency of MCID use per PROM for TKA.

of life along a five-point Likert scale was the most commonly used anchor (40% [16 of 40]) (Fig. 3). Studies using an external subjective question measured on the Likert scale calculated the MCID value as the mean change in preoperative to postoperative PROM score in patients who identified themselves to be "somewhat better," "a little better," or "fair" (Fig. 3). Other anchor questions included those assessing a patient's ability to complete activities of daily living (6%), value of the procedure (6%), response to treatment (6%), and awareness of the affected joint (6%) (Supplemental Table 1; http://links.lww.com/CORR/A960). The receiver operating characteristic (ROC) curve

estimation method was used next-most frequently (25% [10 of 40]). Less frequently used derivation methods included the Global Transition Item to classify patients as having improved or not improved after their surgery (5% [two of 40]), the mean change method (5% [two of 40]), and predictive modeling (5% [two of 40]). Additionally, two studies (5%) assessed whether patients would be willing to undergo their TKA again as an external anchor (Fig. 3). All studies that used the anchor-based MCID to answer a separate outcomes question applied the value as a threshold, and characterized patients as either successful or unsuccessful in reaching the threshold.

Wolters Kluwer

Table 2. Review of the top five most commonly used PROMs

PROM	Description	Scoring	Subsections
Oxford Knee Score	A 12-item survey designed to assess function and pain after TKA, originally developed by the University of Oxford	Each question is scored 0 to 4, 4 being best. Overall score can range from 0 to 48. Can convert raw score to 0 to 100 metric, multiply by 3.57.	Pain subscale, function subscale
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)	A 24-item survey divided into three subscales used to evaluate hip and knee osteoarthritis	Each question scored 0 to 4, 4 being worst; 0 to 20 for pain, 0 to 8 for stiffness, 0 to 68 for physical function	Pain domain, stiffness domain, physical function domain
Knee Society Score (KSS)	A 10-question survey designed to evaluate a patient's knee and functional abilities before and after TKA	Both subsections are scored 0 to 100, with lower scores indicating worse outcomes	Knee score domain (seven items), functional score domain (three items)
12-Item Short Form Survey (SF-12)	A 12-question general health survey using questions from each of the eight dimensions of the SF-36 survey	In the United States, the population average for both subdomains is 50 ± 10 points	Physical component domain, mental component domain
Knee Injury and Osteoarthritis Outcome Score	A 42-question knee-specific survey designed to assess changes in knee pathology over time	Scores range from 0 to 100, 0 being worst possible	Pain (nine items), symptoms (seven items), function in daily living (17 items), sports/ recreation (five items), knee-related quality of life (four items)
KOOS-12	KOOS-12: Twelve-question short form		
KOOS JR	KOOS JR: seven-question short form specific to TKA		

PROM = patient-reported outcome measure; KOOS = Knee Injury and Osteoarthritis Outcome Score; KOOS JR = Knee Injury and Osteoarthritis Outcome Score for Joint Replacement.

Derivation Methods of Distribution-based MCIDs

For studies reporting a distribution-based calculation method, all articles used a measure of study population variance, either standard deviation or standard error, to estimate the MCID (100% [31 of 31]) (Fig. 4). A large proportion of studies estimated the MCID by calculating one-half the standard deviation of the difference between preoperative and postoperative PROM scores (45% [14 of 31]) (Supplemental Table 1; http://links.lww.com/CORR/A960). Another portion of studies estimated the MCID by using the minimum detectable change (MDC) of a corresponding PROM as a proxy (29% [nine of 31]). The most used confidence level was 90% (MDC 90; five of nine), followed by 95% (MDC 95; two of nine) and 80% (MDC80; one of nine). Other studies calculated an MCID from one-half the standard deviation of the preoperative PROM score (13% [four of 31]), the standard error of measurement of the preoperative PROM score (3%), and other variance-based methods (3%). All studies that used the distribution-based MCID to answer a separate outcomes question applied the value as a threshold and characterized patients as either successful or unsuccessful in reaching the threshold (12 of 12).

Comparison of MCID Values by Calculation Method

Most reported median MCID values (15 of 19) did not differ by calculation method for each unique PROM (Table 4). The MCIDs reported for the OKS were quite similar between anchor-based and distribution-based calculation methods, with medians of 6 (IQR 4 to 7) and 7 (IQR 5 to 10), respectively. Additionally, the KSS Function and Knee MCID values were similar because medians for each component were within 1 point for both calculation methods. Some variability arose specifically looking at median MCIDs and interquartile ranges by calculation method for other PROMs. Specifically, the WOMAC median composite score via the anchor-based method was 15 (IQR 10 to 23), while the distribution-based value was 10 (IQR 10 to 11). Similar variability existed among the WOMAC subscores of Function and Stiffness, because MCID values were disparate by more than 5 points for each component by calculation method. The KOOS JR anchor-based MCID was 14 (IQR 7 to 18) while the distribution-based value was 8 (IQR 7 to 9). Three MCID PROM values varied by calculation method, including the WOMAC Function component score and KOOS Pain and Activities of Daily Living subscores (Table 4). The median

Table 3. Studies presenting new MCIDs for PROMs in TKA

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Beard et al., 2015 [2]	94,015	OKS: 9	Anchor-based (individual cohort)	6
		OKS: 5	Anchor-based (between study groups)	
		OKS: 7	Anchor-based (ROC)	
		OKS: 4	Distribution-based (MDC90)	
Beard et al., 2020 [1]	264	OKS: 2	Anchor-based (question)	60
Berliner et al., 2017 [3]	562	KOOS: 10	Distribution-based (1/2 SD score change)	12
		SF-12: 5		
Bin Abd Razak et al., 2016 [5]	3062	OKS: 5	Distribution-based (1/2 SD score change)	5
		SF-36 PCS: 10		
Blevins et al., 2019 [6]	228	KOOS ADL: 10	Distribution-based (1/2 SD score change)	24
		KOOS Pain: 10.3		
		KOOS QoL: 13.2		
		KOOS Sports/Recreation: 15.8		
		KOOS Symptoms: 12		
		SF-12 PCS: 5		
		SF-12 MCS: 5.4		
Carender et al., 2022 [7]	12,341	KOOS JR: 9.8	Distribution-based (1/2 SD score change)	12
		KOOS ADL: 11.7		
		KOOS Pain: 11.4		
		KOOS Symptoms: 11.7		
		WOMAC Pain: 11.3		
		WOMAC Function: 10.8		
Chesworth et al., 2008 [8]	1578	WOMAC Pain: 36	Anchor-based (question)	12
		WOMAC Function: 33		
Clement et al., 2014 [10]	505	OKS Pain: 5	Anchor-based (question)	12
		OKS Function: 4.3		
		SF-12 Pain: 4.5		
		SF-12 PCS: 4.8		
Clement et al., 2018 [9]	2589	WOMAC: 10	Anchor-based (question)	12
		WOMAC Function: 9		
		WOMAC Pain: 11		
		WOMAC Stiffness: 8		
		WOMAC: 17	Anchor-based (ROC)	
		WOMAC Function: 16		
		WOMAC Pain: 21		
		WOMAC Stiffness: 13		
		WOMAC: 12	Distribution-based (MDC95)	
		WOMAC Function: 11		
		WOMAC Pain: 23		
		WOMAC Stiffness: 27		

Table 3. continued

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Clement et al., 2019 [12]	2589	SF-12 PCS: 1.8	Anchor-based (question)	12
		SF-12 MCS: 1.5		
		SF-12 PCS: 2.7	Anchor-based (ROC)	
		SF-12 MCS: -1.4		
		SF-12 PCS: 8.9	Distribution-based (MDC90)	
		SF-12 MCS: 13.8		
Clement et al., 2021 [11]	484	FJS: 13.7	Anchor-based (question)	6
		FJS: 17.7	Anchor-based (one group)	
		FJS: 10	Anchor-based (individual)	
		FJS: 12	Distribution-based (MDC90)	
Danoff et al., 2018 [15]	165	VAS-D: 22.6 mm	Anchor-based	0.25
		VAS-D: 16.1 mm	Distribution-based (SEM)	
Darrith et al., 2021 [16]	872	KOOS JR: 6.8	Distribution-based	12
		PROMIS-GH PCS: 2.3		
Eckhard et al., 2021 [17]	352	KOOS-12: 11.1	Anchor-based	12
		KOOS-12 Function: 15.2		
		KOOS-12 Pain: 13.5		
		KOOS-12 QoL: 8		
Escobar et al., 2007 [19]	423	WOMAC: 15	Anchor-based	6
		WOMAC Pain: 23		
		WOMAC Function: 19		
		WOMAC Stiffness: 15		
		SF-36: 10	Anchor-based	
		SF-36 Physical Function: 12		
		SF-36 Role Physical: 12		
		SF-36 Bodily Pain: 17		
		SF-36 General Health: 1		
		SF-36 Social Function: 12		
		SF-36 Role Emotional: 8		
		SF-36 Vitality: 4		
		SF-36 Mental Health: 0		
		SF-6D Utility: 0		
		WOMAC Pain: 22	Distribution-based (MDC95)	
		WOMAC Function: 13		
		WOMAC Stiffness: 29		
		SF-36 Physical Function: 20	Distribution-based (MDC95)	
		SF-36 Role Physical: 27		
		SF-36 Bodily Pain: 38		
		SF-36 General Health: 27		
		SF-36 Social Function: 41		
		SF-36 Role Emotional: 29		
		SF-36 Vitality: 30		
		SF-36 Mental Health: 24		
		SF-6D Utility: 0		

Table 3. continued

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Escobar et al., 2013 [18]	912	WOMAC Pain: 16, 28, 45 WOMAC Function: 17, 33, 45	Anchor-based (by baseline severity tertile)	12
		WOMAC Pain: 10, 23, 35 WOMAC Function: 15, 27, 42	Anchor-based ROC (by baseline severity tertile)	
		WOMAC Pain: 27 WOMAC Function: 21 WOMAC Pain: 26 WOMAC Function: 23	Anchor-based (by two different anchor questions)	
Escobar and Riddle, 2014 [20]	923	WOMAC Function: 47.5, 37.2, 27.6, 15.9 WOMAC Pain: 47.9, 32.2, 25.1, 12.4	Anchor-based (by quartiles of baseline severity)	12
		WOMAC Function: 32 WOMAC Pain: 29	Anchor-based (global mean)	
Fan et al., 2021 [21]	161	HSS: 5.41 NRS-Walking: 1.24	Anchor-based (question)	36
Goodman et al., 2020 [22]	10,775	KOOS Pain: 21 KOOS ADL: 14	Anchor-based	24
Holtz et al., 2020 [23]	199	FJS: 10.8 WOMAC Pain: 7.5 WOMAC Function: 7.2	Anchor-based (binary regression)	12
		FJS: 13 WOMAC Pain: 12.5 WOMAC Function: 14.7	Anchor-based (ROC)	
Humphrey et al., 2022 [24]	314	PROMIS PF10a MCID-Worsening: -1.89	Distribution-based	12
Kang, 2021 [27]	191,379	OKS: 6 EQ-5D-3L: 0.090	Anchor-based (mean change method)	6
		OKS: 7 EQ-5D-3L: 0.036	Anchor-based (Youden index)	
		OKS: 8 EQ-5D-3L: 0.069	Anchor-based (Short distance)	
		OKS: 1.6 EQ-5D-3L: 0.95	Distribution-based (standardized response mean)	
		OKS: 6 EQ-5D-3L: 0.182	Distribution-based (medium effect size)	
		OKS: 9 EQ-5D-3L: 0.292	Distribution-based (large effect size)	
Katakam et al., 2021 [29]	1059	KOOS PS: 8.2	Anchor-based	12
Khalil et al., 2020 [30]	875	PROMIS-GH PCS: 2.3 PROMIS-GH PCS: 2.5	Distribution-based Anchor-based	12

Table 3. continued

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Kim et al., 2021 [31]	422	WOMAC: 23.4	Anchor-based with central sensitization (CS)	24
		WOMAC: 14.7	Anchor-based without CS	
		WOMAC: 29.5	Anchor-based with CS (ROC)	
		WOMAC: 26.5	Anchor-based without CS (ROC)	
		WOMAC: 10.4	Distribution-based with CS	
		WOMAC: 8.6	Distribution-based without CS	
Kuo et al., 2020 [33]	858	KOOS JR: 6.4	Distribution-based (preoperative scores)	24
		KOOS Total: 35.4		
		KOOS Pain: 8.1		
		KOOS Symptoms: 9.5		
		KOOS ADL: 8.7		
		KOOS Sport & Rec: 9.9		
		KOOS QoL: 7.8		
		KOOS JR: 8.7	Distribution-based (score change)	
		KOOS Total: 48.5		
		KOOS Pain: 10.2		
		KOOS Symptoms: 11.1		
		KOOS ADL: 10.3		
		KOOS Sport & Rec: 15.6		
		KOOS QoL: 13.4		
		KOOS JR: 17.5	Anchor-based (total score)	
		KOOS Total: 82.9		
		KOOS Pain: 22.3		
		KOOS Symptoms: 10.8		
		KOOS ADL: 22.1		
		KOOS Sport & Rec: 17.5		
		KOOS QoL: 12.5		
		KOOS JR: 20.8	Anchor-based (itemized score)	
		KOOS Total: 91.8		
		KOOS Pain: 25.0		
		KOOS Symptoms: 14.3		
		KOOS ADL: 24.6		
		KOOS Sport & Rec: 17.5		
		KOOS QoL: 12.5		
Lee et al., 2017 [34]	550	KSS Function: 6.1	Anchor-based (satisfaction)	24
		KSS Knee: 5.3		
		KSS Function: 6.5	Anchor-based (OKS)	
		KSS Knee: 5.4		

Table 3. continued

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Lizaur-Utrilla et al., 2020 [37]	507	KSS Knee: 7.2	Anchor-based (question)	24
		KSS Knee: 8.9	Anchor-based (ROC)	
		KSS Knee: 7.2	Distribution-based	
		KSS Function: 9.7	Anchor-based (question)	
		KSS Function: 10.3	Anchor-based (ROC)	
		KSS Function: 6.3	Distribution-based	
Lyman et al., 2018 [38]	2630	KOOS Pain: 8	Distribution-based (1/2 SD score change)	12
		KOOS Symptoms: 9		
		KOOS ADL: 9		
		KOOS QoL: 8		
		KOOS JR: 6		
		KOOS Pain: 13	Distribution-based (MDC90)	
		KOOS Symptoms: 12		
		KOOS ADL: 13		
		KOOS QoL: 12		
		KOOS JR: 9		
		KOOS Pain: 18	Anchor-based	
		KOOS Symptoms: 7		
		KOOS ADL: 16		
		KOOS QoL: 17		
		KOOS JR: 14		
Most et al., 2022 [40]	26720	Preoperative OKS < 19: 19.5	Anchor-based (stratified by preoperative OKS)	12
		Preoperative OKS 20-27: 14.5		
		Preoperative OKS > 28: 8.5		
Neuprez et al., 2018 [41]	280	WOMAC: 10	Anchor-based	12
Nishitani et al., 2019 [42]	344	KSS Symptoms: 1.9	Anchor-based	6
		KSS Satisfaction: 2.2		
		KSS Function: 4.1		
Sabah et al., 2022 [45]	347536	OKS: 5	Anchor-based (mean change method)	6
		OKS: 10.5	Anchor-based (ROC)	
		OKS: 7	Anchor-based (predictive modeling)	
		OKS: 3.9	Distribution-based (1/2 SD)	
		OKS: 2.6	Distribution-based (SEM)	
		OKS: 6	Distribution-based (MDC90)	
Shaw et al., 2021 [46]	1340	R-TKA	Distribution-based (1/2 SD preoperative score)	6
		KOOS JR: 6.59		
		PROMIS-GH MCS: 4.46		
		PROMIS-GH PCS: 3.39		
		M-TKA	Distribution-based (1/2 SD preoperative score)	
		KOOS JR: 6.79		
		PROMIS-GH MCS: 3.84		
		PROMIS-GH PCS: 3.39		

Table 3. continued

Author, year	Study Size, n	PROM: MCID	MCID calculation method	Mean follow-up in months
Soh et al., 2022 [48]	1931	KOOS-12: 20.1	Anchor-based (mean change method)	3
		KOOS-12 Pain: 17.5		
		KOOS-12 Function: 21		
		KOOS-12 QoL: 21.8		
		KOOS-12: 20.8	Anchor-based (ROC)	
		KOOS-12 Pain: 15.6		
		KOOS-12 Function: 19.8		
		KOOS-12 QoL: 21.9		
		KOOS-12: 22.5	Anchor-based (predictive modeling)	
		KOOS-12 Pain: 22.1		
		KOOS-12 Function: 21.3		
		KOOS-12 QoL: 24.4		
SooHoo et al., 2014 [49]	229	SF-12 PCS: 4.97	Distribution-based	12
		SF-12 MCS: 5.11		
		WOMAC: 10.21		
		UCLA: 0.92		
Unnanuntana et al., 2018 [51]	157	2-minute walk test: 12.7 meters	Distribution-based	72
		Timed up and go test: 9.5 seconds		
Vina et al., 2016 [52]	269	WOMAC: 9.4	Anchor-based	24

PROM = patient-reported outcome measure; MCID = minimum clinically important difference; OKS = Oxford Knee Score; ROC = receiver operating characteristic; MDC = minimum detectable change; SF-12 = 12-Item Short Form Survey; SF-36 PCS = 36-Item Short Form Survey Physical Component Score; KOOS = Knee Injury and Osteoarthritis Outcome Score; ADL = Activities of Daily Living; QoL = Quality of Life; PCS = Physical Component Score; MCS = Mental Component Score; WOMAC = Western Ontario and McMaster University Arthritis Index; FJS = Forgotten Joint Score; VAS-D = Visual Analog Score Drawn; SEM = standard error of measurement; KOOS JR = Knee Injury and Osteoarthritis Outcome Score for Joint Replacement; PROMIS-GH PCS = Patient-Reported Outcomes Measurement Information Systems Global Health Assessment Physical Component Score; HSS = Hospital for Special Surgery Score; NRS = Numeric Rating Scale; PROMIS PF10a = Patient-Reported Outcomes Measurement Information Systems 10-item Physical Function; EQ-5D-3L = EuroQual 5D-3L Test; KSS = Knee Society Score; UCLA = University of California Los Angeles Score

anchor-based MCIDs were more often larger than the distribution-based values (13 of 19). Additionally, the anchor-based MCIDs tended to have larger IQRs than the distribution-based values for each respective PROM (Table 4).

Discussion

The MCID is an important measure of an effect size minimally noticeable to a patient. Use of the MCID for outcome measure comparison is a critical evolution from frequentist statistical analysis, which does not provide information on effect size of a treatment. Indeed, an effect size that is minimally noticeable to the patient is the very least that a clinician should offer in the context of any surgical risk. There is expected heterogeneity in MCID values, which are

known to be unique to specific PROMs, treatment types, and study populations. The rationale for the current study was to create a repository of calculated MCIDs for total knee arthroplasty, compare values based on calculation method, and provide practical recommendations on MCID values and their derivation methods which can be employed in future TKA outcomes research. We found that despite their variability, MCID values were clustered for many of the most frequently reported PROMs, regardless of the calculation method by which they were derived. Additionally, there were predominate calculation methods used for both anchor- and distribution-based MCIDs. We strongly recommend that surgeons use reference-calculated MCID values to evaluate TKA outcome measures against a clinically relevant effect size. Patients deserve treatment modalities which have been critically evaluated in terms of an effect noticeable to them.

Wolters Kluwer

Table 4. Summary of newly calculated MCIDs reported in TKA, separated by calculation method

PROM	Anchor-based		Distribution-based		Combined		p value[a]
	Median	IQR (Q1 to Q3)	Median	IQR (Q1 to Q3)	Median	IQR (Q1 to Q3)	
OKS	6	4 to 7	7	5 to 10	6	5 to 9	0.29
WOMAC	15	10 to 23	10	10 to 11	12	10 to 17	0.20
Function	23	16 to 33	11	10.8 to 11	20	15 to 33	0.04
Pain	25	14 to 31	22	17 to 23	23	13 to 29	0.33
Stiffness	13	11 to 14	28	28 to 29	15	12 to 28	0.20
KSS							
Function	7	6 to 10	6	6	6	6 to 9	1.0
Knee	6	5 to 8	7	7	7	5 to 7	1.0
KOOS	87	85 to 90	35	23 to 42	49	35 to 83	0.20
Pain	22	20 to 23	10	9 to 11	12	10 to 20	0.01
QoL	17	13 to 15	12	8 to 13	13	11 to 13	0.37
ADL	18	14 to 23	10	9 to 11	12	10 to 14	0.01
Symptoms	11	9 to 13	11	10 to 12	11	10 to 12	0.90
Sports/recreation	18	18	16	13 to 16	16	16 to 18	0.14
KOOS JR	14	7 to 18	8	7 to 9	9	7 to 12	0.25
KOOS-12	20	18 to 21	NR	NR	20	18 to 21	n/a
Function	20	19 to 21	NR	NR	20	19 to 21	n/a
Pain	17	15 to 19	NR	NR	17	15 to 19	n/a
QoL	22	18 to 23	NR	NR	22	18 to 23	n/a
SF-12 PCS	3	2 to 4	5	5 to 7	5	3 to 5	0.10
SF-12 MCS	0.1	-0.7 to 0.8	5	5 to 10	5	2 to 5	0.20
SF-36 composite	10	10	NR	NR	10	10	n/a
SF-36 PCS	NR	NR	10	10	10	10	n/a
VAS-D, mm	23	23	16	16	19	18 to 21	n/a
NRS Walking	1	1	NR	NR	1	1	n/a
PROMIS-GH PCS	3	2.9 to 3.4	2.3	2.3	2.5	2.3 to 3.4	0.12
PROMIS-GH MCS	4	3.0 to 4.3	NR	NR	4	3.0 to 4.3	n/a
FJS	13	11 to 14	12	12	13	11 to 14	1.0
EQ-5D-3L	0.07	0.06 to 0.08	0.29	0.24 to 0.62	0.14	0.07 to 0.26	0.10
UCLA	NR	NR	0.9	0.9	0.9	0.9	n/a
HSS	5	5	NR	NR	5	5.4	n/a

[a]Wilcoxon rank sum nonparametric test comparison of anchor-based and distribution-based medians with assumption of non-normal distribution. MCID = minimum clinically important difference; NR = not reported; IQR = interquartile range; Q1 = quartile 1; Q3 = quartile 3; n/a = not applicable secondary to insufficient data for comparison; OKS = Oxford Knee Score; KSS = Knee Society Score; KOOS = Knee Injury and Osteoarthritis Outcome Score; PF = physical functioning; RP = role limitations because of physical health; BP = bodily pain; GH = general health; MCS = Mental Component Summary; VT = vitality; SF = social functioning; RE = role limitations because of emotional problems; MH = mental health; NRS = Numeric Rating Scale; PROMIS-GH PCS = Patient-Reported Outcomes Measurement Information System Global Health Physical Component Summary; FJS= Forgotten Joint Score; EQ-5D = EuroQal 5-Dimensions; VR-12 PCS = Veteran Rands 12-Item Health Survey Physical Component Summary; VR-12 MCS = Veteran Rands 12-Item Health Survey Mental Component Summary; UCLA = University of California Los Angeles Score.

Limitations

This systematic review had several limitations. Although this is a systematic review, articles describing other TKA PROM MCIDs could have been missed. We think the likelihood of this is very low because we strictly adhered to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for our comprehensive three-database search. After the current study was accepted for publication, but before final publication, Karhade et al. [28] published a retrospective study examining MCID thresholds and attainment for PROMIS scores. Because that
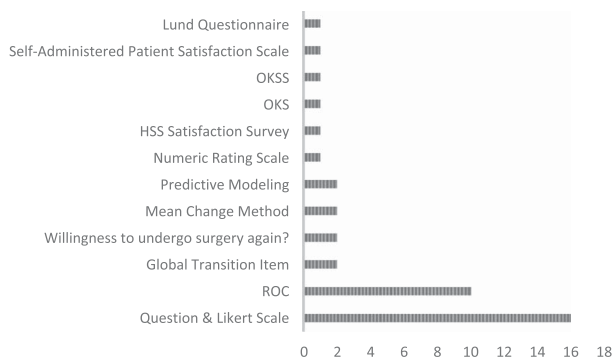
Fig. 3 This graph shows the frequency of times each anchor was used for anchor-based calculation methods.

study was published after our search dates, it was not included in this analysis. Our summary of the data is limited by the amount of information provided by the authors of each article. Authors are frequently vague, or details are limited regarding the version of specific instance of a given survey. This lack of detail is particularly problematic with the Patient-Reported Outcome Measure Information System (PROMIS) surveys because there are many different versions of those scores. Moreover, the use of MCIDs as a metric of improvement is an inherent limitation. In an editorial, Leopold and Porcher [35] aptly pointed out that ideally, a procedure's impact would be more than minimal. We did not address the concept of substantial clinical benefit or other patient-centric improvement metrics. The substantial clinical benefit quantifies the amount of improvement a patient feels is clinically important. Bernstein et al. [4] recommended a tiered approach to grading PROM results for orthopaedic procedures and argued that the MCID, patient-acceptable symptom state, and substantial clinical benefit should not be considered mutually exclusive metrics. Rather, the MCID, patient-acceptable symptom state, and substantial clinical benefit quantify the minimum, intermediate, and substantial effect sizes of a procedure, respectively. Despite this limitation,
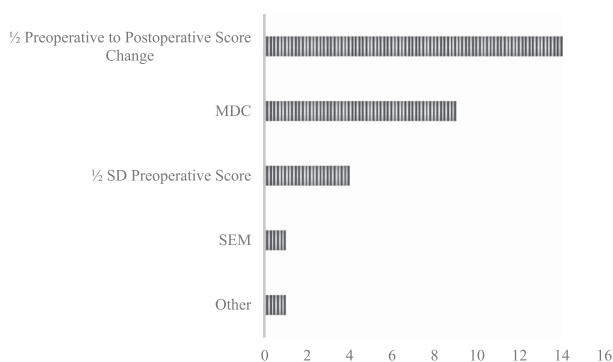

Fig. 4 This graph shows the frequency by which different distribution-based calculation methods were used.

the MCID remains a critically important threshold capturing the minimum effect size necessary to deem a TKA outcome acceptable to the patient and surgeon.

Recommended MCID Values and Use

We recommend using a 7-point increase as the MCID for the OKS, consistent with the median anchor-based value derived from several high-quality studies with large patient groups. Multiple studies [2, 27, 45] reported MCID values clustered around a 7-point increase (IQR 5 to 8). Most importantly, this anchor-based value demands the patient perspective be part of its derivation and is consequently a more reliable value than the distribution-based median. Additionally, we reported a median WOMAC MCID of 12 with a range of 10 to 17, with no difference between calculation methods. The median MCID for the subscore of Function was 20, with an IQR from 15 to 33, and the median value for the Pain subscore was 23, with an IQR from 13 to 29. These wide ranges of values likely represent differing patient populations, follow-up lengths, or the MCID calculation methods. Furthermore, multiple studies stratified WOMAC MCID values based on preoperative PROM scores [18, 20]. The largest MCID values in the WOMAC composite and subscores are derived from studies in patient groups with very poor preoperative scores, positively skewing the dataset. Therefore, we recommend using a 10-point to 15-point increase as the MCID for the WOMAC composite because most values were clustered at this magnitude (seven of 13). Importantly, selecting a smaller value for an MCID within the reported range demands a smaller clinical improvement from a treatment in question. By contrast, selecting a relatively larger MCID value will require a larger effect size because fewer patients in a given study population may not reach the MCID threshold. Therefore, we caution future authors to choose an MCID cutoff considering the overall surgical risk and preoperative disease severity of the study population in question; the riskier or more invasive the procedure, the larger the MCID that should be used.

MCID Calculation Methods

Thirty-nine percent (15 of 38) of studies used an anchor-based method to calculate a new MCID, while 32% (12 of 38) used a distribution-based technique. Twenty-nine percent of studies (11 of 38) calculated MCID values using both methods. The ideal method of calculating MCIDs has not been established. Caution must be taken, because insisting on individual MCIDs for each study will push authors to use distribution-based methods. Distribution-based methods may be convenient, but they do not

incorporate a patient's perception of improvement. The lack of incorporating the patient's appreciation of improvement into the MCID detracts from its meaningfulness to clinicians. Anchor-based methods may be preferred because they are based on patient-related indicators rather than statistical criteria alone. Despite the large variety of anchor questions described in the current review, previous authors calculating anchor-based values have found that the MCID did not depend on the anchor question asked [10, 18]. Specifically, Clement et al. [10] reported that the calculated MCID value for the OKS and SF-12 did not vary between two different anchor questions addressing patient satisfaction and pain, respectively. Moreover, another prevalent issue is the lack of appropriate survey identification. Because many surveys have had multiple versions or iterations, such as the PROMIS surveys, it is crucial for authors to specifically list what version or iteration of a survey they are using. At a minimum, authors should report the full version number and date of the PROM form.

The substantial variation of MCIDs in other studies demands attention. Zuckerman [53] concisely summarized the three main problems that confound MCID reporting: the substantial variability in quantifying and reporting values, the application of MCID values to a different treatment, and the lack of a consistent protocol or calculation method. By applying MCID values without careful attention to the demographics and surgical risk of the patient population from which the MCID was derived, surgeons and policymakers may find that a procedure appears to be clinically beneficial when it does not deliver a clinically important improvement to a specific patient group. The converse is also true. As PROMs become linked to reimbursement, it is paramount that we are using the correct metrics to evaluate our performance. We encourage future authors to include the applicable survey in an appendix accompanying their article to ensure transparency and consistency in outcomes reporting. Authors should use anchor-based calculations, when feasible, to fully capture a patient's perspective. In keeping with most calculated anchor-based values, we urge authors to use an external anchor question assessing pain relief, quality of life, or satisfaction along a Likert scale. With this method, the MCID value is obtained from the mean difference in PROM score between preoperatively and postoperatively, corresponding to patients who respond they are "somewhat better" after TKA. Likewise, when anchor-based calculations are not feasible, we encourage the use of one-half the standard deviation of the difference between preoperative and postoperative PROM scores for the distribution-based method for consistency with previously published research. When deriving a new MCID from a study population is not feasible, we strongly support the use of a referenced MCID value to ensure a TKA outcome is measured in effect size rather than mere statistical

significance. Care should be taken to use a reference MCID value from a similar study population, along with demographic and surgical risk parameters. A higher MCID value should be used when evaluating results in study populations in which individuals have higher comorbidity burdens, more invasive procedures, or milder levels of preoperative symptoms. Each journal should insist on these practices in order to better consolidate PROMs data. These changes would help improve our understanding of the PROMs used in TKA and improve the quality of care for our patients.

Conclusion

MCIDs represent an essential threshold for surgeons to consider offering arthroplasty to patients in the context of surgical risk. Despite substantial variability in MCIDs for each PROM, some consensus exists, regardless of calculation method. We believe anchor-based values are preferable, as the MCID value is calculated using secondary data provided by the patient rather than estimating the value from a measure of variance. Therefore, we recommend using a 7-point increase as the MCID for the OKS, consistent with the median reported anchor-based value derived from several high-quality studies. Likewise, we recommend using a 10-point to 15-point increase for the MCID of composite WOMAC, as the median value was 12 with no difference between calculation methods. Given the importance of MCID values to treatment selection and patient safety, authors and journals should report these values with greater consistency. Prioritization of MCID use in outcomes research will support improved patient selection for arthroplasty, allow the surgeon to practice more safely and efficiently, and may ultimately guide public policy for procedure reimbursement.

References

1. Beard DJ, Davies LJ, Cook JA, et al. Total versus partial knee replacement in patients with medial compartment knee osteoarthritis: the TOPKAT RCT. *Health Technol Assess*. 2020; 24:1-98.
2. Beard DJ, Harris K, Dawson J, et al. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68:73-79.
3. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res*. 2017;475:149-157.
4. Bernstein DN, Nwachukwu BU, Bozic KJ. Value-based health care: moving beyond "minimum clinically important difference" to a tiered system of evaluating successful clinical outcomes. *Clin Orthop Relat Res*. 2019;477:945-947.
5. Bin Abd Razak HR, Tan CS, Chen YJ, et al. Age and preoperative Knee Society Score are significant predictors of outcomes among Asians following total knee arthroplasty. *J Bone Joint Surg Am*. 2016;98:735-741.

6. Blevins JL, Chiu YF, Lyman S, et al. Comparison of expectations and outcomes in rheumatoid arthritis versus osteoarthritis patients undergoing total knee arthroplasty. *J Arthroplasty*. 2019; 34:1946-1952.e1942.

7. Carender CN, Glass NA, De A, Bozic KJ, Callaghan JJ, Bedard NA. Outcomes vary significantly using a tiered approach to define success after total knee arthroplasty. *J Arthroplasty*. 2022; 37:1266-1272.

8. Chesworth BM, Mahomed NN, Bourne RB, Davis AM; OJRR Study Group. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol*. 2008;61:907-918.

9. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. What is the minimum clinically important difference for the WOMAC index after TKA? *Clin Orthop Relat Res*. 2018;476: 2005-2014.

10. Clement ND, MacDonald D, Simpson AH. The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. 2014;22:1933-1939.

11. Clement ND, Scott CEH, Hamilton DF, MacDonald D, Howie CR. Meaningful values in the Forgotten Joint Score after total knee arthroplasty. *Bone Joint J*. 2021;103:846-854.

12. Clement ND, Weir D, Holland J, Gerrand C, Deehan DJ. Meaningful changes in the Short Form 12 physical and mental summary scores after total knee arthroplasty. *Knee*. 2019;26: 861-868.

13. Copay AG, Chung AS, Eyberg B, Olmscheid N, Chutkan N, Spangehl MJ. Minimum clinically important difference: current trends in the orthopaedic literature, part I: upper extremity: a systematic review. *JBJS Rev*. 2018;6:e1.

14. Copay AG, Eyberg B, Chung AS, Zurcher KS, Chutkan N, Spangehl MJ. Minimum clinically important difference: current trends in the orthopaedic literature, part II: lower extremity: a systematic review. *JBJS Rev*. 2018;6:e2.

15. Danoff JR, Goel R, Sutton R, Maltenfort MG, Austin MS. How much pain is significant? Defining the minimal clinically important difference for the visual analog scale for pain after total joint arthroplasty. *J Arthroplasty*. 2018;33:S71-S75.e72.

16. Darrith B, Khalil LS, Franovic S, et al. Preoperative Patient-Reported Outcomes Measurement Information System global health scores predict patients achieving the minimal clinically important difference in the early postoperative time period after total knee arthroplasty. *J Am Acad Orthop Surg*. 2021;29: e1417-e1426.

17. Eckhard L, Munir S, Wood D, et al. Minimal important change and minimum clinically important difference values of the KOOS-12 after total knee arthroplasty. *Knee*. 2021;29:541-546.

18. Escobar A, Garcia Perez L, Herrera-Espineira C, et al. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage*. 2013;21:2006-2012.

19. Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage*. 2007;15:273-280.

20. Escobar A, Riddle DL. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage*. 2014;22: 1107-1110.

21. Fan XY, Ma JH, Wu X, et al. How much improvement can satisfy patients? Exploring patients' satisfaction 3 years after total knee arthroplasty. *J Orthop Surg Res*. 2021;16:389.

22. Goodman SM, Mehta BY, Mandl LA, et al. Validation of the Hip Disability and Osteoarthritis Outcome Score and Knee Injury and Osteoarthritis Outcome Score pain and function subscales for use in total hip replacement and total knee replacement clinical trials. *J Arthroplasty*. 2020;35:1200-1207.e1204.

23. Holtz N, Hamilton DF, Giesinger JM, Jost B, Giesinger K. Minimal important differences for the WOMAC osteoarthritis index and the Forgotten Joint Score-12 in total knee arthroplasty patients. *BMC Musculoskelet Disord*. 2020;21:401.

24. Humphrey TJ, Katakam A, Melnic CM, Bedair HS. Defining failure in primary total joint arthroplasty: the minimal clinically important difference for worsening score. *J Arthroplasty*. 2022; 37:630-636.e631.

25. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1-12.

26. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-415.

27. Kang S. Assessing responsiveness of the EQ-5D-3L, the Oxford Hip Score, and the Oxford Knee Score in the NHS patient-reported outcome measures. *J Orthop Surg Res*. 2021;16:18.

28. Karhade AV, Bernstein DN, Desai V, et al. What is the clinical benefit of common orthopaedic procedures as assessed by the PROMIS versus other validated outcomes tools? *Clin Orthop Relat Res*. 2022;480:1672-1681.

29. Katakam A, Bragdon CR, Chen AF, Melnic CM, Bedair HS. Elevated body mass index is a risk factor for failure to achieve the Knee Disability and Osteoarthritis Outcome Score-physical function short form minimal clinically important difference following total knee arthroplasty. *J Arthroplasty*. 2021;36:1626-1632.

30. Khalil LS, Darrith B, Franovic S, Davis JJ, Weir RM, Banka TR. Patient-Reported Outcomes Measurement Information System (PROMIS) global health short forms demonstrate responsiveness in patients undergoing knee arthroplasty. *J Arthroplasty*. 2020; 35:1540-1544.

31. Kim MS, Koh IJ, Choi KY, Seo JY, In Y. Minimal clinically important differences for patient-reported outcomes after TKA depend on central sensitization. *J Bone Joint Surg Am*. 2021;103:1374-1382.

32. Kolin DA, Moverman MA, Pagani NR, et al. Substantial inconsistency and variability exists among minimum clinically important differences for shoulder arthroplasty outcomes: a systematic review. *Clin Orthop Relat Res*. 2022;480:1371-1383.

33. 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.

34. Lee WC, Kwan YH, Chong HC, Yeo SJ. The minimal clinically important difference for Knee Society Clinical Rating System after total knee arthroplasty for primary osteoarthritis. *Knee Surg Sports Traumatol Arthrosc*. 2017;25:3354-3359.

35. Leopold SS, Porcher R. Editorial: the minimum clinically important difference-the least we can do. *Clin Orthop Relat Res*. 2017;475:929-932.

36. Leopold SS, Porcher R. Editorial: threshold p values in orthopaedic research-we know the problem. What is the solution? *Clin Orthop Relat Res*. 2018;476:1689-1691.

37. Lizaur-Utrilla A, Gonzalez-Parreno S, Martinez-Mendez D, Miralles-Munoz FA, Lopez-Prats FA. Minimal clinically important differences and substantial clinical benefits for Knee Society Scores. *Knee Surg Sports Traumatol Arthrosc*. 2020;28:1473-1478.

38. Lyman S, Lee YY, 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;476:2432-2441.

39. Maltenfort M, Diaz-Ledezma C. Statistics in brief: minimum clinically important difference-availability of reliable estimates. *Clin Orthop Relat Res*. 2017;475:933-946.

40. Most J, Hoelen TA, Spekenbrink-Spooren A, Schotanus MGM, Boonen B. Defining clinically meaningful thresholds for patient-reported outcomes in knee arthroplasty. *J Arthroplasty*. 2022;37: 837-844.e833.

41. Neuprez A, Neuprez AH, Kaux JF, et al. Early clinically relevant improvement in quality of life and clinical outcomes 1 year postsurgery in patients with knee and hip joint arthroplasties. *Cartilage*. 2018;9:127-139.

42. Nishitani K, Yamamoto Y, Furu M, et al. The minimum clinically important difference for the Japanese version of the new Knee Society Score (2011KSS) after total knee arthroplasty. *J Orthop Sci*. 2019;24:1053-1057.

43. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*. 2003;41:582-592.

44. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

45. Sabah SA, Alvand A, Beard DJ, Price AJ. Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty. *J Clin Epidemiol*. 2022;143:159-168.

46. Shaw JH, Lindsay-Rivera KG, Buckley PJ, Weir RM, Banka TR, Davis JJ. Minimal clinically important difference in robotic-assisted total knee arthroplasty versus standard manual total knee arthroplasty. *J Arthroplasty*. 2021;36: S233-S241.

47. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological Index for Non-randomized Studies (MINORS): development and validation of a new instrument. *ANZ J Surg*. 2003;73:712-716.

48. Soh SE, Harris IA, Cashman K, et al. Minimal clinically important changes in HOOS-12 and KOOS-12 scores following joint replacement. *J Bone Joint Surg Am*. 2022;104: 980-987.

49. SooHoo NF, Li Z, Chenok KE, Bozic KJ. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. *J Arthroplasty*. 2015;30:176-191.

50. Team J. *JASP* (Version 0.16.3). 2022.

51. Unnanuntana A, Ruangsomboon P, Keesukpunt W. Validity and responsiveness of the two-minute walk test for measuring functional recovery after total knee arthroplasty. *J Arthroplasty*. 2018;33:1737-1744.

52. Vina ER, Hannon MJ, Kwoh CK. Improvement following total knee replacement surgery: exploring preoperative symptoms and change in preoperative symptoms. *Semin Arthritis Rheum*. 2016; 45:547-555.

53. Zuckerman JD. CORR insights(R): substantial inconsistency and variability exists among minimum clinically important differences for shoulder arthroplasty outcomes: a systematic review. *Clin Orthop Relat Res*. 2022;480:1384-1386.