

Examining the Minimal Important Difference of Patient-reported Outcome Measures for Individuals with Knee Osteoarthritis: A Model Using the Knee Injury and Osteoarthritis Outcome Score

Kathryn A.G. Mills, Justine M. Naylor, Jillian P. Eyles, Ewa M. Roos, and David J. Hunter

ABSTRACT. Objective. To examine the influence of different analytical methods, baseline covariates, followup periods, and anchor questions when establishing a minimal important difference (MID) for individuals with knee osteoarthritis (OA). Second, to propose MID for improving and worsening on the Knee Injury and Osteoarthritis Outcome Score (KOOS).

Methods. Retrospective analysis of prospectively collected data from 272 patients with knee OA undergoing a multidisciplinary nonsurgical management strategy. The magnitude and rate of change as well as the influence of baseline covariates were examined for 5 KOOS subscales over 52 weeks. The MID for improving and worsening were investigated using 4 anchor-based methods.

Results. Waitlisted for joint replacement and exhibiting unilateral/bilateral symptoms influenced change in KOOS over time. Generally, low correlations between anchors and KOOS change scores limited calculations of MID; thus, they were only proposed for the pain, activities of daily living, and quality of life subscales. The method used to calculate the MID influenced the cutpoint; however, the type of anchor question only influenced the MID when analyzed with a particular mean change method. Depending on patient and clinical characteristics, the subscale, and the analytical approach used, the MID for KOOS improvement ranged from an absolute change of -1.5 to 20.6 points and worsening ranged from -19.17 to 8.5 points.

Conclusion. MID vary with patient and clinical characteristics, KOOS subscale, and analytical approach. Provided the anchor question is relevant to the patient-reported outcome and baseline status is considered, the anchor does not appear to influence the MID for improvement or worsening when using some anchor-based methods. (First Release January 15, 2016; J Rheumatol 2016;43:395–404; doi:10.3899/jrheum.150398)

Key Indexing Terms:

KNEE OSTEOARTHRITIS PATIENT OUTCOME ASSESSMENT QUALITY INDICATORS

Patient-reported outcomes (PRO) are a valuable source of information used to monitor people with knee osteoarthritis (OA). One challenge in applying PRO in OA treatment settings is interpreting changes in scores over time^{1,2,3}. This challenge has sparked the pursuit for the minimal important difference (MID) as well as insight into measurement error when PRO are applied in different patient cohorts, disease states, and treatment settings.

No best practice approach exists for determining the MID. Since first described by Jaeschke, *et al*⁴, recommendations regarding the nomenclature and definition of what constitutes clinically important change have evolved considerably. For example, while terminology was originally introduced as the “minimal clinically important difference”⁴, the “clinically” was subsequently removed to minimize the focus on the clinical arena and to focus on the patients’ experiences

From the Centre for Physical Health, Department of Medicine and Health Sciences, Macquarie University; Orthopaedic Department, Liverpool Hospital; South West Sydney Clinical School, University of New South Wales; Ingham Institute of Applied Medical Research; Department of Rheumatology, Royal North Shore Hospital, Sydney, Australia; Institute of Bone and Joint Research, Kolling Institute, University of Sydney, Australia; Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

Jillian Eyles receives funding from a Ramsay Health Care Allied Health Scholarship and a Royal North Shore Hospital Staff Specialist Award. Professor Hunter is supported by a National Health and Medical Research Council Practitioner Fellowship.

K.A. Mills, BPhy, PhD, Centre for Physical Health, Department of Medicine and Health Sciences, Macquarie University; J.M. Naylor,

BAppSci (Phy), PhD, Orthopaedic Department, Liverpool Hospital, and South West Sydney Clinical School, University of New South Wales, and Ingham Institute of Applied Medical Research; J.P. Eyles, BAppSci (Phy), Department of Rheumatology, Royal North Shore Hospital, and Institute of Bone and Joint Research, Kolling Institute, University of Sydney; E.M. Roos, PhD, PT, Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark; D.J. Hunter, MBBS, PhD, Department of Rheumatology, Royal North Shore Hospital, and Institute of Bone and Joint Research, Kolling Institute, University of Sydney.

Address correspondence to Dr. K. Mills, Discipline of Physiotherapy, Macquarie University, Sydney, Australia, 2109. E-mail: Kathryn.mills@mq.edu.au

Accepted for publication November 10, 2015.

instead⁵. Further, definitions have been summarized as weighing change in health status against the risk of undertaking the intervention, change that would alter the patient's care, and change of a sufficient magnitude to be perceived as important to the patient^{1,6}. Despite these different definitions, the common concept is that the MID is the lowest boundary of change determined to be important in some way. The patient remains at the center and so many believe that an MID is valid only if it is anchored to patients' perception of health^{6,7}.

The Knee injury and Osteoarthritis Outcome Score (KOOS) is a PRO for patients with knee injury and OA. Its responsiveness to change has been demonstrated in a diverse range of patients and pathologies^{8,9,10}. Consisting of 5 subscales, each scored from 0–100 — pain, other symptoms, activities of daily living (ADL), sport and recreation, and knee-related quality of life (QOL) — higher scores indicate less pain and disability.

The minimal detectable change (MDC) and standard error (SE) of the measure have been examined for each of the KOOS subscales in knee OA cohorts^{11,12}. However, the MID for improvement and worsening are yet to be empirically established in cohorts undergoing nonsurgical interventions. These realities make KOOS an ideal model for examining the process of establishing a MID for a PRO that is primarily used for individuals undergoing nonsurgical knee OA management. In doing so, the model permits discussion of issues and complexities involved with defining clinically meaningful change, including determining whether single or multiple methods for establishing a MID is more appropriate, and whether baseline covariates, time, or the type of anchor question influences the result. The aim of our study was to undertake this examination, and in doing so to propose potential MID for the KOOS subscales.

MATERIALS AND METHODS

Ours was a retrospective analysis of prospectively collected data from participants enrolled in the Osteoarthritis Chronic Care Program (OACCP) at a teaching hospital in New South Wales (NSW), Australia. The OACCP was developed by the Agency for Clinical Innovation as a model of nonsurgical management for individuals with hip and knee OA. The OACCP was a 12-month program focusing on self-management, timely access to information, and management from a multidisciplinary team. All patients received individualized care plans that were reviewed at about 12 weeks, 26 weeks, and 52 weeks. While the OACCP served as a clinical program, the NSW Population and Health Services Ethics Committee approved the use of patient data for research purposes (2012/08/413), and patients were informed of this at the time of enrollment.

Inclusion criteria. Individuals with medically diagnosed hip or knee OA were eligible to participate in the OACCP. The face validity of a doctor diagnosis for the presence of OA has previously been demonstrated¹³. Individuals can be referred to the program by general practitioners, orthopedic surgeons, or rheumatologists. In addition to OA, a minimum of 30/100 pain severity in the signal joint on most days of the month was required for inclusion. While there were no exclusion criteria, patients with severe comorbidities may not complete the walking test. In addition, patients with dementia were excluded from PRO and were thus not included here. Because we were interested in changes in KOOS over time in individuals with knee

OA, only data from patients managed for unilateral or bilateral knee OA who had completed their 26-week review were included in the analysis. Of the 280 participants meeting these criteria, data from 272 individuals were analyzed because 6 participants had not completed any KOOS subscale. Of the 191 participants who had completed the 52-week assessment (55 individuals were lost to followup), 7 had not completed the KOOS. Thus, data from 184 individuals were retained (Table 1).

Collected data. Age at entry, unilateral or bilateral pain, and whether the patient was waitlisted for joint replacement were recorded at baseline. At baseline and each followup, the KOOS, index joint pain severity, and level walking test score were recorded. Pain severity over the previous week was measured on a 0–100 mm, worst to best, visual analog scale. Level walking was measured using a 6-min walk test (6MWT) in which the patient was asked to walk as quickly as possible back and forth over a 25-m straight flat walking area for 6 min and distance was recorded in meters.

It has been suggested that global transition scales, or anchors, examining both a patient's index joint and general health were required because they measure different aspects of a patient's outcome¹⁴. At all followups, patients were asked: "Compared with when I started this program, my walking on level ground has..." and "Compared with when I started this program, my knee has..." Patients responded on a 7-point Likert scale: "much improved," "moderately improved," "slightly improved," "not changed," "slightly worse," "moderately worse," and "much worse."

Statistical analysis. Each KOOS subscale was considered separately. Analyses were conducted in RStudio (Version 0.98.1091, RStudio Inc.). First, 4 linear mixed models examining maximum likelihood estimation were fitted to each subscale. This approach was chosen because of its flexible assumptions, ability to increase the complexity of the model with each iteration, and ability to cope with changes in sample size from baseline to 52 weeks. The first 2 models examined whether the magnitude of change in KOOS (dependent variable) differed at each followup period. The third model examined whether the rate of change in KOOS significantly differed between each followup period. These models provided insight into whether changes in the KOOS could be combined across all followup periods or whether each followup period needed independent examination. Adjustment for baseline status was viewed as critically important when making interpretations regarding the MID^{2,7,15,16}. Thus, the final model examined the effect of baseline covariates on an individual's baseline KOOS, magnitude, and rate of change. Baseline characteristics (sex, age, pain severity, 6MWT, waitlisted for joint replacement status, and unilateral/bilateral symptoms) were added to the model using a forced entry method. Variables that improved the fit of the model using log likelihood, Akaike information criterion, and Bayesian information criterion were retained.

Once the appropriate model for each subscale was established, KOOS change scores were calculated by subtracting baseline KOOS from their respective subscale at each followup. Spearman rank correlations were then examined between the anchors and the absolute KOOS change scores. Similar to the first 3 linear mixed models, this analysis addressed whether time to followup influenced the MID. It also served to determine whether the type of anchor question was an important consideration when establishing an MID. Correlation coefficients higher than between 0.3 and 0.5 have been proposed as suitable between the anchor or measurement instruments^{7,17}. Therefore, an *a priori* threshold of 0.4 was set as the minimum correlation. Only the KOOS subscales that met this minimum criterion underwent further analysis for a MID.

The MID for both improvement and worsening in each subscale was calculated using receiver-operating characteristic (ROC) curves and mean change methods². The ROC method is increasingly used to ascertain the MID because it uses all data, which increases the precision and accuracy of estimates¹⁸. The improvement MID, or minimum clinically important improvement (MCII), classified participants who indicated they had slightly, moderately, or much improved as the "improved group" whereas those registering no change or worse were considered the "non-improved group". Similarly, for the minimal clinically important change for worsening (MCIW), those who reported slightly, moderately, or much worse were

Table 1. Demographics at each measurement occasion. Values are mean (SD) unless otherwise specified.

Variables	Baseline	Week 12	Week 26	Week 52
BMI	30.24 (6.1)	30.3 (6.47)	30.24 (6.2)	30.48 (7.08)
Pain, VAS 100 mm	42.61 (23.03)	32.59 (23.82)	31.39 (24.52)	27.96 (21.87)
6MWT distance	418.41 (116.92)	455.39 (113.14)	462.38 (117.59)	465.15 (121.03)
WJR/not WJR, n	119/150	122/150	122/150	78/106
Unilateral/bilateral symptoms, n	174/95	175/97	175/97	118/66
KOOS pain				
No. participants	269	272	272	184
Male/female, n	84/185	84/188	84/188	58/126
Mean (SD)	47.56 (18.66)	53.61 (17.99)	57.01 (18.84)	58.91 (19.41)
KOOS other symptoms				
No. participants	269	272	272	184
Male/female, n	84/185	84/188	84/188	58/126
Mean (SD)	47.31 (19.57)	52.69 (18.54)	53.98 (19.67)	56.19 (19.88)
KOOS ADL				
No. participants	272	272	272	184
Male/female, n	84/188	84/188	84/188	58/126
Mean (SD)	49.37 (20.2)	56.25 (19.61)	58.75 (20.62)	59.87 (20.11)
KOOS sport and recreation				
No. participants	190	203	204	142
Male/female, n	68/123	73/130	70/134	50/32
WJR/not WJR, n	73/118	78/125	74/130	53/89
Unilateral/bilateral symptoms, n	127/64	134/69	128/76	91/51
Mean (SD)	27.9 (21.07)	29.65 (20.39)	31.4 (21.56)	32.29 (21.51)
KOOS quality of life				
No. participants	244	254	260	171
Male/female, n	80/164	80/174	79/181	57/114
WJR/not WJR, n	100/144	109/145	113/147	68/103
Unilateral/bilateral symptoms, n	158/86	166/88	169/91	107/64
Mean (SD)	29.95 (15.8)	37.86 (17.67)	47.32 (18.39)	47.58 (15.79)

BMI: body mass index; VAS: visual analog scale; 6MWT: 6-min walk time; WJR: waitlisted for joint replacement; KOOS: Knee injury and Osteoarthritis Outcome Score; ADL: activities of daily living.

classified as the “worse group” and those reporting no change or improved as the “not-worse group”. The area under the curve (AUC) and 95% CI were calculated as an estimate of each anchor’s ability to predict changes on the KOOS. The respective AUC were then compared using DeLong statistics to determine whether 1 anchor was a superior predictor¹⁹.

The primary ROC approach for determining an appropriate MID was the Youden method²⁰. In this method, the point of the ROC curve that maximized the distance to the identity line was selected as the optimal MID^{17,21}. The secondary method was the 80% specificity rule, whereby the MID is the best sensitivity for response while still achieving at least 80% specificity²². These approaches were chosen because the former provides maximum accuracy at predicting a MID by any combination of sensitivity and specificity²², whereas the latter provides the optimal MID while ensuring that at least 80% or median number of true negatives are correctly classified^{22,23}. CI were established by drawing 500 stratified bootstrap samples^{2,19}.

Two mean change methods were investigated. The first method was as it was originally proposed by Jaeschke, *et al*⁴, whereby the MID for improvement and worsening were calculated as the mean change in scores over time within the subgroup of participants who reported they were slightly improved or slightly worse. The second was as proposed by Redelmeier and Lorig²⁴, where the mean change score for the subgroup identifying as “no change” was subtracted from that of the groups identifying as slightly improved or worse. CI for both methods were calculated as 1.96 multiplied by the SE².

RESULTS

Description of the study population. Based on the rules for

missing items in the KOOS²⁵, baseline subscale scores could be calculated for 98.9% of pain and other symptoms, 100% of ADL, 69.85% of sport and recreation, and 89.7% of knee-related QOL. The number of missing subscale scores reduced over time; the sport and recreation subscale had the most missing items at every timepoint (Table 1).

The highest percentage of responses for the walking on level ground anchor at 26 weeks and 52 weeks was “much improved” (27% and 32%, respectively), and only 2 patients reported they were “much worse”. This pattern of responses was mirrored when patients were asked about their knee health (Figure 1). In general, patients reporting themselves as improved also reported better KOOS at followup; those who were symptomatically worse exhibited worse scores. However, there was significant overlap between categories on both anchors (Figure 2).

KOOS changes over time and influence of baseline covariates. With the exception of the sport and recreation subscale, the rate of change significantly fluctuated between timepoints (Supplementary Data is available from the authors on request). Thus, subsequent MID analyses treated each followup in isolation (baseline to 12 weeks, baseline to 26 weeks, and baseline to 52 weeks). Pain severity and waitlisted

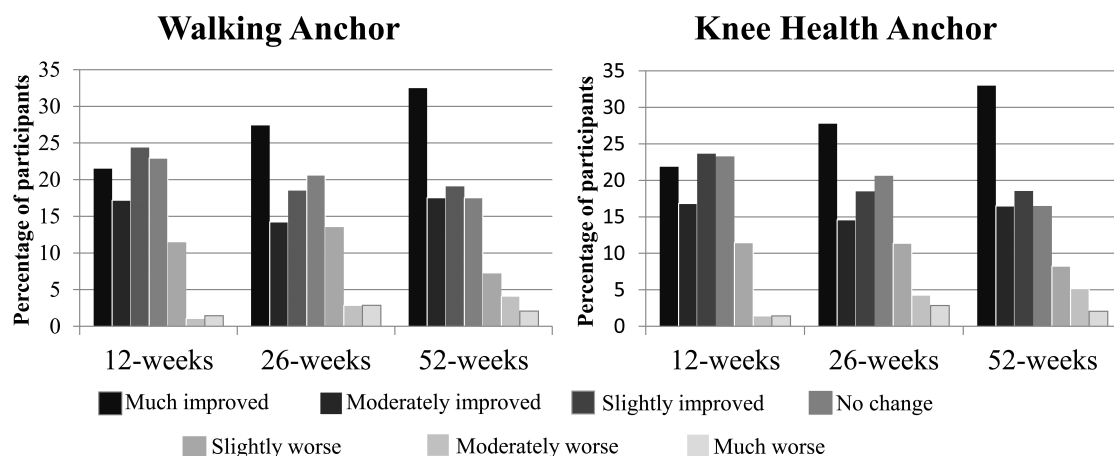


Figure 1. Pattern of responses on the global transition scales for each timepoint.

for joint replacement influenced baseline KOOS for all subscales, with those reporting worse pain and those on the waiting list for joint replacement independently reporting worse scores. Two subscales exhibited significant interactions between predictors and time. Individuals with bilateral knee OA reported 2.39 points (95% CI 0.76–4.01) greater improvement in the pain subscale over 52 weeks than their unilateral counterparts. Individuals who were not waitlisted for joint replacement exhibited 1.93 points (95% CI 0.23–3.64) greater improvement over 52 weeks on the QOL subscale. All baseline covariates except sex and unilateral/bilateral symptoms were significantly correlated; however, all correlation coefficients were small ($0.1 < r < 0.3$; data not shown)²⁶.

Correlations with anchor scales. Despite highly significant correlations between both anchors and the majority of KOOS change scores at followups, there were few instances where correlations exceeded an absolute value of 0.4 (Table 2). Correlations exceeding this threshold, and subsequently entered into the MID analysis, were the change in the pain subscale from baseline to 26 weeks for those with bilateral knee OA and the 52-week change scores for those with unilateral knee OA, the change scores obtained from baseline to 52 weeks for the ADL subscale, and QOL subscales from baseline to 52 weeks only for patients waitlisted for joint replacement.

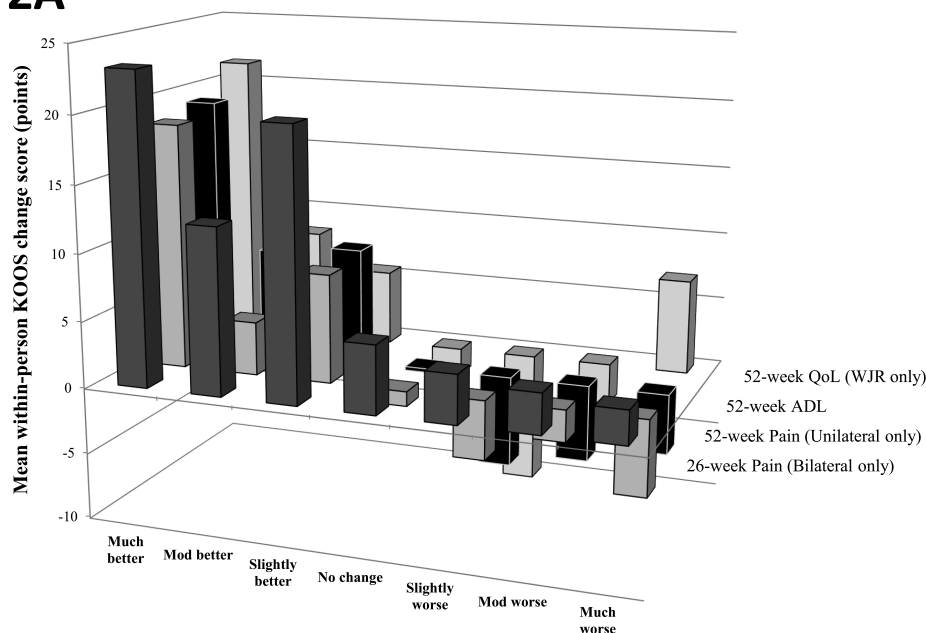
Influence of anchor question on MID. The AUC indicated that both anchors were fair to good predictors of change in KOOS subscales (Table 3) with no statistical difference between their predictive ability (Figure 3). Further, the MID established using the ROC methods were comparable between the anchor questions (Table 3), with the exception of the MCIW for pain at 52 weeks where the difference was 11 points. The comparability between anchors was also observed using the Jaeschke, *et al*⁴ mean change method, where the largest discrepancy between anchors was exhibited

for the MCII of the pain subscale at 26 weeks (4.04 points). In contrast, using the Redelmeier and Lorig²⁴ method, MID proposed for pain and QOL subscales at 26 weeks and 52 weeks exhibited differences ranging from 6.61 points to 12.64 points between anchor questions.

MID thresholds. The range of MCII thresholds were comparable between the analytical methods when considered across all subscales. Of the mean change methods, the Jaeschke, *et al*⁴ method produced the highest MCII of 20.6 points (8.91–31.21) for the pain subscale for bilateral patients at 26 weeks and the Redelmeier and Lorig²⁴ method produced the lowest MCII at –5.04 points (–23.06 to 12.98) for the 52-week QOL subscale (waitlisted for joint replacement patients only). Of the ROC methods, the 80% specificity produced the highest and lowest MCII: 18 points for the pain subscale at 26 weeks for patients with bilateral knee symptoms and 1.5 points for the ADL subscale at 52 weeks.

For the MCIW, the lowest thresholds depended on the analytical method used. Mean change-based MCIW ranged from –19.17 points (–31.96 to –6.38) when using the general knee anchor on the QOL subscale at 52 weeks for those who were waitlisted for joint replacement, to 3.64 points (–0.99 to 8.28) when using the walking anchor for the pain subscale at 26 weeks for bilateral knee pain patients. ROC curve-based methods resulted in MID ranging from –7 points for pain at 52 weeks for those with unilateral symptoms to 8.5 points for pain at 26 weeks. Importantly, all but 3 MID established using the Redelmeier and Lorig²⁴ method and 10 of the 16 MID established using the Jaeschke, *et al*⁴ method exhibited CI that included 0, indicating that some participants reported no change or change in the opposite direction on the KOOS compared with the anchor questions. Similarly, 10 of the 16 MID established using the 80% specificity ROC exhibited CI for their sensitivity that were $< 50\%$, indicating that these thresholds were possibly no better than chance at correctly classifying individuals who reported improvement or worsening.

2A



2B

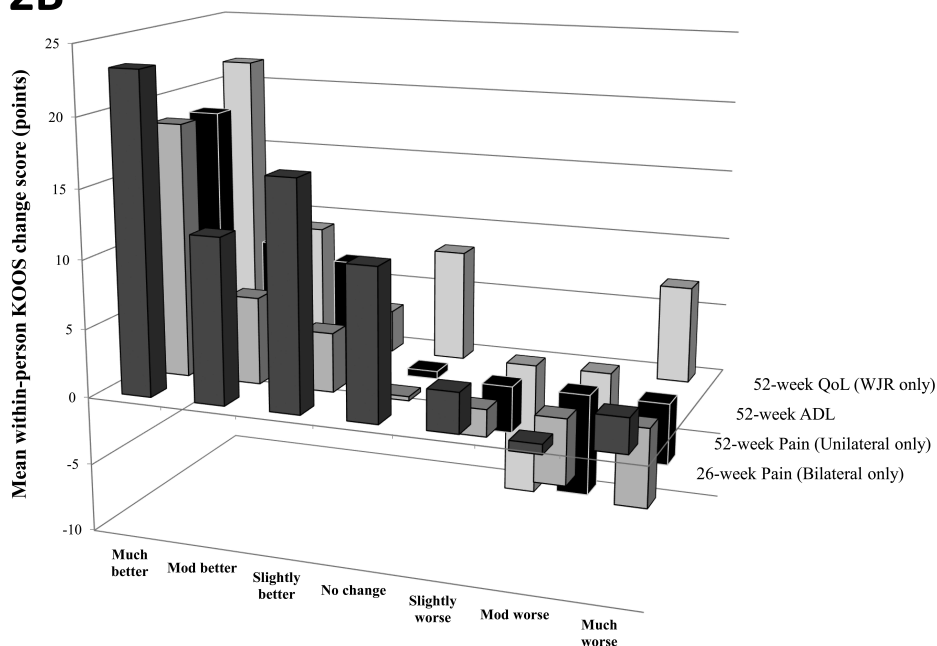


Figure 2. Mean KOOS change scores at 26 weeks and 52 weeks of OACCP participation within transition groups of the (A) walking anchor and (B) general knee health anchor. KOOS: Knee injury and Osteoarthritis Outcome Score; OACCP: Osteoarthritis Chronic Care Program; QOL: quality of life; WJR: waitlisted for joint replacement; ADL: activities of daily living.

DISCUSSION

The importance of defining meaningful change in outcomes used in clinical decision making and clinical trials is widely accepted. The objectives of our study were to examine the complexities of establishing a MID in a knee OA population

undergoing nonsurgical multidisciplinary management as well as to understand its influences. To do this, we used the KOOS as a model for PRO. The initial modeling determined that the magnitude and rate of KOOS change over time depended on patient and clinical characteristics and differed

Table 2. Spearman rank correlation between KOOS change scores and the anchor questions reported at 12, 26, and 52 weeks.

Variables	Walking Anchor	Knee Health Anchor
Pain		
12-week unilateral	-0.391**	-0.373**
12-week bilateral	-0.331**	-0.287**
26-week unilateral	-0.376**	-0.375**
26-week bilateral	-0.453**	-0.455**
52-week unilateral	-0.412**	-0.437**
52-week bilateral	-0.349**	-0.332**
Other symptoms		
12 week	-0.229**	-0.209**
26 week	-0.277**	-0.288**
52 week	-0.314**	-0.325**
ADL		
12 week	-0.291**	-0.279**
26 week	-0.335**	-0.345**
52 week	-0.502**	-0.471**
Sport and recreation		
12 week	0.046	0.066
26 week	-0.128	-0.138
52 week	-0.168	-0.144
Knee-related QOL		
12-week WJR	-0.195	-0.163
12-week non-WJR	-0.288**	-0.283**
26-week WJR	-0.253*	-0.233*
26-week non-WJR	-0.222**	-0.278**
52-week WJR	-0.584**	-0.494**
52-week non-WJR	-0.366**	-0.392**

* Correlation significant at 0.05 level. ** Correlation significant at 0.01 level. KOOS: Knee injury and Osteoarthritis Outcome Score; ADL: activities of daily living; QOL: quality of life; WJR: waitlisted for joint replacement.

for each subscale. Incorporating these findings and the analytical method, the MID for KOOS improvement ranged from an absolute change of -1.5 to 20.6 points and worsening ranged from -19.7 to 8.5 points. This highlights a growing consensus among clinimetric research: there is an inability to apply a single MID across all populations and followup periods.

Our study suggests that the influence of specific anchor questions on MID is controversial and may depend on the PRO/subscale examined and analytical method. While the correlation coefficients and AUC of the subscales that underwent MID analysis in our study are considered acceptable^{27,28}, this was not the case for all the subscales and timepoints assessed. Even in the subscales where the correlation to the anchor questions met our *a priori* threshold, there were several instances where the groups rating themselves as “slightly improved” or “slightly worse” exhibited greater KOOS change than those in the moderate categories. This confirms that the constructs measured by a single-item question and those measured using a multidimensional scale are not identical. These findings reinforce the importance of establishing the correlation between the anchor and PRO prior to any MID analysis and limited the number of MID that could be proposed in our current study.

Of the subscales assessed, the anchors used by the

OACCP produced similar MID for pain and ADL when either ROC method or the Jaeschke, *et al*⁴ approach were used. Further, both the AUC and DeLong statistic indicated the anchors had similar predictive value for KOOS change on these subscales. Therefore, it appears these analytical methods support the findings of Tubach, *et al*³, whereby provided the MID is relevant, the actual anchor has a low effect on the definition of success in clinical trials. However, Redelmeier and Lorig’s²⁴ mean change method resulted in substantial differences (1.29 to 12.64 points) between MID proposed for the same subscale when different anchors were used. This questions the application of MID established using methods that are highly influenced by the anchor question when there are alternative methods available, unaffected by this issue.

Regardless of the anchor, the method used to calculate the MID will result in different values^{2,7}. For both the pain subscale at 26 weeks and QOL subscale at 52 weeks, there was up to a 16-point difference between MICW and an 11.5-point difference between MCII proposed using different analytical methods. Despite this, of the 64 proposed MID, 58 did not exceed the ± 13 points required to exceed the MDC obtained from comparable cohorts to our current study^{11,12}. Therefore, the applicability of the proposed MID for KOOS is questionable. It may be that any change beyond measurement error represents clinically meaningful change. Further validation is required to confirm this.

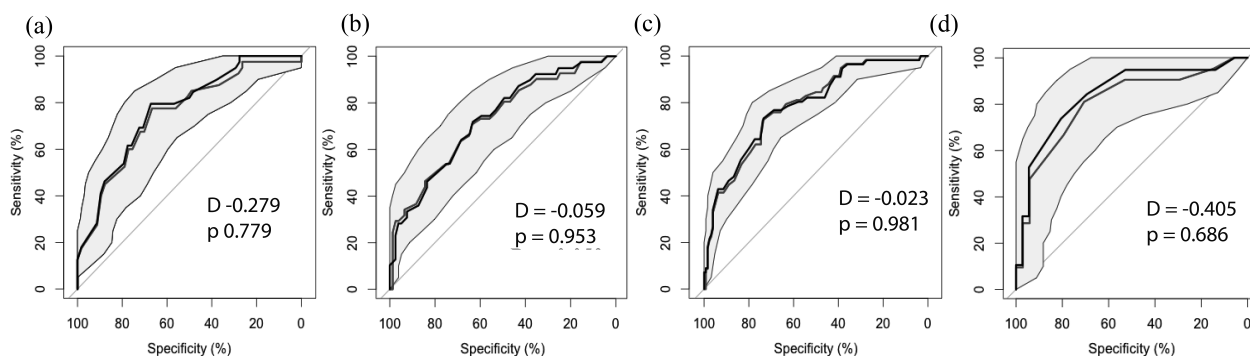
The differences exhibited between the MID analytical methods also suggest the use of each method needs to be considered when applied in clinical and research settings. An advantage of applying MID calculated from ROC analytical methods is that they remain the same when applied to groups and individual patients because ROC curves are a diagnostic analysis technique²¹. In contrast, MID calculated from the mean change methods are only applicable at group levels²⁹. This implies that ROC methods may have greater use in clinical settings whereas either method could be used for research purposes investigating group effects. However, in our analysis, ROC curve MID exhibited lower sensitivity than specificity. This means they performed better when identifying patients who need more or different interventions compared with those responding well to the current strategy. To gauge the most precise distribution of patients’ outcomes, researchers and clinicians would need to be able to calculate and apply both the MCII and MCIW. Mean change methods avoid this double analysis issue because they only handle data from individuals who are slightly improved or worse, rather than all data. This carries its own limitations because sample sizes may be small, leading to low power and unreliable MID. This was the case in our analysis, as evidenced by wide 95% CI. Therefore, while the triangulation of multiple MID approaches is recommended^{2,3,7,30}, our findings suggest that consideration of the context and intended interpretation of the thresholds are required to maximize power and generalizability.

Table 3. MID analysis for each subscale using 3 different methods to determine the threshold.

Variables	Transition Scale	KOOS Subscale	AUC, % (95% CI)	Youden Method (95% CI)		80% Specific Method (95% CI)		Mean Change Method, Jaeschke, et al ⁴		Mean Change Method, Redelmeier and Lorig ²⁴	
				MID	Sensitivity, % (95% CI)	Specificity, % (95% CI)	MID	Sensitivity, % (95% CI)	Specificity, % (95% CI)	MID (95% CI)	Mean Change Method, Redelmeier and Lorig ²⁴ MID (95% CI)
Improvement*	Walking	Pain, 26 weeks***	77.03 (67.67–86.39)	11.5	67.34 (55.17–79.31)	79.48 (66.67–92.31)	18.0	52.34 (34.4–74.33)	82.05 (66.72–93.96)	20.06 (8.91–31.21)	14.95 (2.05–27.84)
		Pain†	73 (63.37–82.62)	4.0	63.3 (51.9–73.42)	71.79 (56.41–84.62)	13.0	46.04 (26.26–65.07)	82.3 (65.72–93.71)	8.15 (0.69–15.91)	9.25 (–1.88 to 20.38)
		ADL	78.68 (71.52–85.83)	2.5	73.44 (66.0–80.47)	73.21 (62.5–83.93)	–1.5	60.71 (47.28–73.21)	80.0 (73.44–85.94)	8.67 (2.85–14.18)	8.93 (0.74–17.13)
		QOL‡	84.14 (72.49–95.78)	3.0	73.68 (52.63–92.24)	80.56 (67.99–91.67)	3.0	73.68 (52.63–92.24)	80.56 (67.99–91.67)	11.62 (1.46–21.78)	7.6 (–9.72 to 24.92)
Worsening**	Walking	Pain, 26 weeks***	75.09 (65.18–85)	11.5	66.67 (54.39–77.19)	77.5 (65.0–90.0)	17.0	50.88 (38.60–63.16)	82.5 (70.0–95.0)	16.56 (9.89–23.23)	5.56 (–7.73 to 18.86)
		Pain†	72.58 (62.86–82.31)	4.0	63.64 (52.56–73.41)	70.73 (56.1–82.93)	12	48.83 (32.74–66.83)	80.5 (64.69–93.16)	4.3 (–3.43 to 12.06)	4.03 (–7.73 to 15.44)
		ADL	78.56 (71.56–85.55)	2.5	72.41 (60.34–82.76)	73.81 (65.46–81.75)	–1.5	58.62 (46.55–70.69)	80 (73.02–86.51)	8.17 (2.24–14.1)	7.64 (–0.79 to 16.06)
		QOL‡	80.6 (68.07–93.13)	6.5	70.59 (55.88–85.29)	80.95 (64.17–95.24)	6.5	70.59 (55.88–85.29)	80.95 (64.17–95.24)	10.33 (–3.05 to 23.71)	–5.04 (–23.06 to 12.98)
	General knee	Pain, 26 weeks*	72.05 (60.75–83.34)	8.5	64.93 (54.55–76.01)	80 (60.0–95.0)	8.5	64.93 (54.55–76.01)	80 (60.0–95.0)	3.64 (–0.99 to 8.28)	–1.47 (–9.47 to 6.53)
		Pain**	70.71 (58.7–82.71)	5.5	53.53 (43.91–63.64)	78.95 (60.39–94.74)	–7.0	36.84 (15.79–57.89)	82.83 (76.77–89.9)	–4.5 (–21.47 to 12.47)	–3.4 (–19.29 to 12.49)
		ADL	79.1 (70.36–87.84)	2.5	65.82 (58.86–73.42)	80.77 (65.38–92.31)	2.5	65.82 (58.86–73.42)	80.77 (65.38–92.31)	–6.83 (–16.45 to 2.79)	–6.57 (–17.57 to 4.44)
		QOL***	80.68 (65.67–96.68)	–3.0	61.54 (30.77–84.62)	90.48 (90.96–97.62)	–3.0	61.54 (30.77–84.62)	90.48 (90.96–97.62)	–9.8 (–21.59 to 1.99)	–7.8 (–22.13 to 6.52)
	General knee	Pain, 26 weeks*	73.97 (62.96–84.98)	8.5	65.33 (54.67–74.67)	80.95 (61.0–95.24)	8.5	65.33 (54.67–74.67)	80.95 (61.0–95.24)	2.9 (–2.59 to 8.44)	–8.08 (–21.86 to 5.7)
		Pain**	74.37 (63.51–85.23)	–5.5	52.17 (30.43–69.51)	81.05 (72.63–87.37)	–4.5	52.17 (30.43–69.51)	80 (72.63–87.37)	–2.0 (–19.18 to 15.18)	–2.3 (–18.28 to 13.67)
		ADL	78.81 (70.44–87.18)	2.5	66.88 (59.09–73.38)	80.0 (63.33–93.33)	2.5	66.88 (59.09–73.38)	80.0 (63.33–93.33)	–3.46 (–12.69 to 5.76)	–4.0 (–14.18 to 6.19)
		QOL***	83.36 (69.33–97.39)	–3.0	64.29 (35.72–85.71)	92.68 (84.09–100)	–3.0	64.29 (35.72–85.71)	92.68 (84.09–100)	–11.0 (–22.51 to 0.51)	–19.17 (–31.96 to –6.38)

* Clinically relevant improvements exceed the MID cutoff. ** Clinically relevant worsening is lower than the MID cutoff. *** 26-week change score for those with bilateral knee OA. † 52-week change score for those with unilateral knee OA. ‡ Only for those waitlisted for joint replacement. MID: minimal important difference; OA: osteoarthritis; KOOS: Knee injury and OA Outcome Score; ADL: activities of daily living; QOL: quality of life.

MCII



MCIW

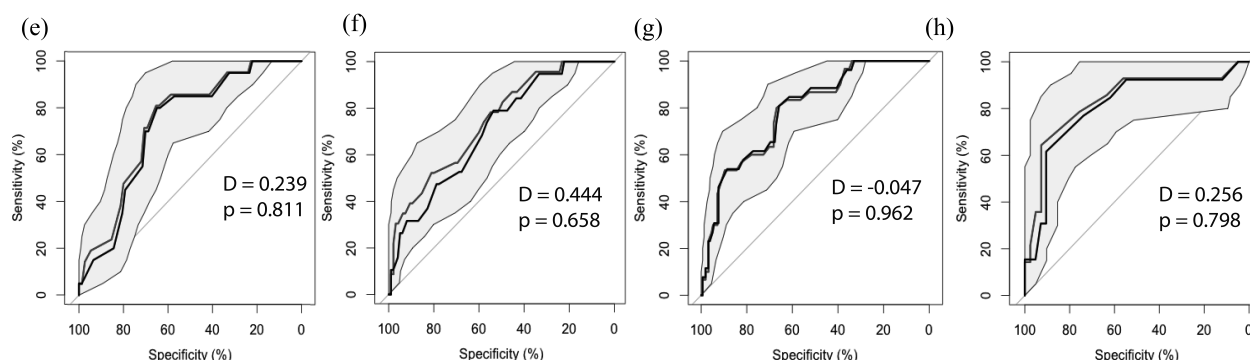


Figure 3. ROC curves depicting MCII and MCIW. Black and shading indicate the ROC curve and 95% CI for the knee health anchor. Grey lines depict the ROC curve for the level-walking anchor. DeLong statistics tests the hypothesis that there are significant differences in AUC between ROC curves. KOOS subscales are (a) and (e) pain at 26 weeks for those with bilateral symptoms, (b) and (f) pain at 52 weeks for those with unilateral symptoms, (c) and (g) ADL at 52 weeks, and (d) and (h) knee-related QOL at 52 weeks for those waitlisted for joint replacement. ROC: receiver-operating characteristic; MCII: minimum clinically important improvement; MCIW: minimal clinically important change for worsening; D: DeLong statistics; AUC: area under the curve; ADL: activities of daily living; QOL: quality of life; KOOS: Knee injury and Osteoarthritis Outcome Score.

In addition to the low number of reliably determined MID, our study has several limitations. First, we used absolute change scores rather than relative change. The latter is frequently used to adjust for baseline covariates. Our choice of approach is based on the potential of a ceiling effect. When the PRO is designed for higher scores to indicate better health status, using percent change scores increases the association with baseline scores, inflates the MID, and risks eliminating true differences in meaning along with the potential error^{1,28,31}. There is no consensus on the optimal adjustments for baseline covariates when establishing a MID and using absolute change may not be the optimal approach. However, by isolating analysis based on covariates that influence KOOS change over time and only proposing MID for subscales that were moderately correlated with the anchor, we believe that we appropriately adjusted for factors that may influence baseline differences.

Limitations of all MID research is the presence of negative

values for MCII and positive values for MCIW, as occurred in the current and previous studies³². Further, while the magnitude of improvement or worsening is quantified, rarely is there an attempt to clarify whether this magnitude of change is important to the patient. Rather, as was the case in our study, “importance” is assumed to be implicit whenever a response category other than “not changed” is selected. It is possible for individuals with chronic conditions to recalibrate their perception of their condition and the importance of change over time, or “response shift”³³. It is also possible that a recall bias has occurred, as commonly occurs in PRO³⁴. We did not investigate either of these issues, and we did not determine which category on the anchor question should be consistently used for the MID. These issues need further investigation and could be as simple as performing a “then-test” and asking patients if their reported change is important to them.

Our study is the first, to our knowledge, to focus on the

issues and recommendations for evaluating the MID for the KOOS in a cohort of individuals with knee OA undergoing nonsurgical management. Patient and clinical characteristics, specifically time to followup and clinical disease severity as measured by laterality of symptoms and being waitlisted for joint replacement, affected the MID for the KOOS. This further confirms there is no such thing as 1 MID for a specific PRO. Additionally, the association between the anchor and PRO appeared to fluctuate over time and limited the MID's clinical usefulness. The result was a limited series of MID for pain, ADL, and knee-related QOL at 26 weeks and 52 weeks that is lower than previously reported MDC. Thus, for a change in the KOOS to be interpreted as clinically meaningful, it needs to exceed both the MID and MDC. Because of the discrepancies in MID thresholds resulting from different analytical approaches, triangulation of multiple methods continues to be more appropriate than using a single method to establish an MID. However, clinicians and researchers are urged to think about the context in which the MID will be applied when determining which methods to triangulate. Future research needs to validate this finding, clarify "important change," and account for recall bias and response shift. By establishing and using MID, clinicians and researchers will be better served to interpret the meaning of PRO.

ACKNOWLEDGMENT

Data used in the presented analysis were obtained with the permission of the Agency for Clinical Innovation (ACI). The authors acknowledge the work of the ACI's Musculoskeletal Network in developing and supporting implementation of the Osteoarthritis Chronic Care Program in New South Wales. We thank the ACI Musculoskeletal Network, especially Robyn Speerin, for assistance with accessing data. We are very grateful to Matthew Williams for assistance with Royal North Shore Hospital data collection.

REFERENCES

1. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002;14:109-14.
2. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524-34.
3. Tubach F, Giraudeau B, Ravaud P. The variability in minimal clinically important difference and patient acceptable symptomatic state values did not have an impact on treatment effect estimates. *J Clin Epidemiol* 2009;62:725-8.
4. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
5. Schünemann HJ, Guyatt GH. Commentary—goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res* 2005;40:593-7.
6. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11:171-84.
7. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9.
8. Orqvist M, Iversen MD, Janarv PM, Broström EW, Roos EM. Psychometric properties of the Knee injury and Osteoarthritis Outcome Score for Children (KOOS-Child) in children with knee disorders. *Br J Sports Med* 2014;48:1437-46.
9. Roos EM, Roos HP, Ek Dahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)—validation of a Swedish version. *Scand J Med Sci Sports* 1998;8:439-48.
10. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes* 2003;1:17.
11. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage* 2008;16:423-8.
12. Naylor JM, Hayen A, Davidson E, Hackett D, Harris IA, Kamalasena G, et al. Minimal detectable change for mobility and patient-reported tools in people with osteoarthritis awaiting arthroplasty. *BMC Musculoskelet Disord* 2014;15:235.
13. Jordan K, Clarke AM, Symmons DP, Fleming D, Porcheret M, Kadam UT, et al. Measuring disease prevalence: a comparison of musculoskeletal disease using four general practice consultation databases. *Br J Gen Pract* 2007;57:7-14.
14. 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* 2013;22:1933-9.
15. Escobar A, Riddle DL. Concordance between important change and acceptable symptom state following knee arthroplasty: the role of baseline scores. *Osteoarthritis Cartilage* 2014;22:1107-10.
16. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29-33.
17. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol* 2010;63:28-36.
18. Turner D, Schünemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol* 2009;62:374-9.
19. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
20. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5.
21. de Vet HC, Terluin B, Knol DL, Roorda LD, Mokkink LB, Ostelo RW, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol* 2010;63:37-45.
22. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. *Arthritis Rheum* 2009;61:313-20.
23. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: identifying reversible and irreversible components. *Arthritis Rheum* 2006;54:2784-92.
24. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements: an illustration in rheumatology. *Arch Int Med* 1993;153:1337-42.
25. Roos EM. KOOS Scoring 2012. [Internet. Accessed November 26, 2015.] Available from: www.koos.nu/KOOSscoring2012.pdf
26. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Spor Exerc* 2009;41:3-13.
27. Hopkins WG. A new view of statistics. [Internet. Accessed

- November 26, 2015.] Available from:
www.sportsci.org/resource/stats
28. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7:541-6.
 29. Wells G, Beaton D, Shea B, Boers M, Simon L, Strand V, et al. Minimal clinically important differences: review of methods. *J Rheumatol* 2001;28:406-12.
 30. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;49:1215-9.
 31. Fayers PM, Hays RD. Don't middle your MID: regression to the mean shrinks estimates of minimally important differences. *Qual Life Res* 2013;23:1-4.
 32. Rouquette A, Blanchin M, Sébille V, Guillemin F, Côté SM, Falissard B, et al. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *J Clin Epidemiol* 2014;67:433-40.
 33. Schwartz CE, Andresen EM, Nosek MA, Krahn GL; RRTC Expert Panel on Health Status Measurement. Response shift theory: important implications for measuring quality of life in people with disability. *Arch Phys Med Rehabil* 2007;88:529-36.
 34. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. *Health Qual Life Outcomes* 2010;8:65.