Clinical Research

What is the Minimum Clinically Important Difference for the WOMAC Index After TKA?

Nicholas D. Clement PhD, FRCS Ed (Tr & Orth), Michelle Bardgett BSC, MSc,
David Weir FRCS Ed (Tr & Orth), James Holland FRCS Ed (Tr & Orth),
Craig Gerrand MD, FRCS Ed (Tr & Orth), David J. Deehan MD, MSc, FRCS, DSc

Abstract

Background The WOMAC score is a validated outcome measure for use in patients undergoing TKA. Defining meaningful changes in the WOMAC score is important for sample-size calculations in clinical research and for interpreting published studies. However, inconsistencies among published studies regarding key definitions for changes in the WOMAC score after TKA potentially could result in incorrectly powered studies and the misinterpretation of clinical research results.

Questions/purposes (1) To identify the minimum clinically important difference (MCID) for the total WOMAC score and its components 1 year after TKA using an anchor-based methodology. (2) To define the minimum important change (MIC) and the minimum detectable change with 95% confidence (MDC95) for the total WOMAC score and its components 1 year after TKA.

Methods Between 2003 and 2013, 3641 patients underwent primary TKA at one center. Of those, 460 patients (13%) were excluded from this retrospective study for pre-specified reasons (mainly secondary OA and bilateral surgery), and 592 patients (16%) were either lost to followup or could not be included because of incomplete questionnaires. WOMAC scores were recorded preoperatively and at 1 year postoperatively. Patient demographics and preoperative Short Form-12 and WOMAC scores were no different for the 16% of patients who were lost to followup or failed to complete 1-year questionnaires and the study cohort (n = 2589). At 1 year, patients were asked "How much did the knee replacement surgery improve the quality of your life?" Their responses were recorded as: a great improvement, moderate improvement, little improvement, no improvement at all, or the quality of my life is worse. The MCID was defined as the difference in the mean change in the WOMAC score between patients with no improvement compared with those with little improvement according to the anchor question. The MIC was defined as the change in the WOMAC score relative to the baseline score for patients who reported a little improvement in their quality of life. The MDC is the smallest change for an individual who is likely to be beyond the measurement error of the scoring tool and represents true change rather than variability in the scoring measure; we report it with 95% confidence bounds defining real change rather than variability in the scoring measure (MDC95). We calculated this with distribution-based

N. D. Clement, M. Bardgett, D. Weir, J. Holland, C. Gerrand, D. J. Deehan, Department of Orthopaedics, Freeman Hospital, Newcastle upon Tyne, United Kingdom

N. D. Clement (✉), Department of Orthopaedics, Freeman Hospital, Freeman Road, High Heaton, Newcastle upon Tyne, United Kingdom, NE7 7DD, email: nickclement@doctors.org.uk

methods for the whole cohort. Patients recording a little improvement (n = 211) and no improvement (n = 115) were used as anchor responses to calculate the MCID (using regression analysis to adjust for potential confounding variables such as age, gender, BMI and preoperative Short Form-12 or WOMAC scores) and the MIC (using receiver operative characteristics curves).

Results After adjusting for confounding variables such as age, gender, BMI as well as preoperative Short Form-12 and WOMAC scores, the MCID was 11 for pain, 9 for function, 8 for stiffness and 10 for the total WOMAC score. The MIC was 21 for pain, 16 for function, 13 for stiffness and 17 for the total WOMAC score. The MDC95 was 23 for pain, 11 for function, 27 for stiffness and 12 for the total WOMAC score.

Conclusions The MCID and MIC for the WOMAC score represent the smallest meaningful effect sizes when comparing the outcome of two groups (difference in mean change between the groups) or when assessing a cohort (a change in score for the group) after TKA, respectively, helping the reader to distinguish between a clinically important effect size and a mere statistical difference. We determined that the error in measurement (based on the MDC95) for the function component and total WOMAC scores were less than the MIC, which suggests changes beyond the MIC are clinically real and not due to uncertainty in the score. These parameters are essential to interpret TKA outcomes research and to ensure clinical research studies are amply powered to detect meaningful differences. Future studies using the WOMAC score to assess TKA outcomes should report not only the statistical significance (a p value) but also the clinical importance using the reported MCID and MIC values.

Level of Evidence Level III, diagnostic study.

Introduction

The WOMAC [3], Knee Injury and Osteoarthritis Outcome (KOOS) Score [23] and the Oxford Knee Score (OKS) [10] are the most commonly employed patient-reported outcome measures (PROMs) for assessing TKA outcomes [21]. The WOMAC [3] is a validated patient-administered questionnaire that assesses three components: pain, stiffness, and function [31]. The WOMAC knee score is a joint-specific, subjective patient assessment tool used to assess TKA outcomes [29]. A search of PubMed using the terms "WOMAC", "total", and "knee" illustrates an increasing use of the WOMAC to evaluate the outcome of TKA, resulting in double the number of citations using the OKS and KOOS scores.

P values are often used to determine whether a change in a scoring measure after TKA has occurred, but this metric does not convey effect size. In addition, a "significant" p value may be too small for a patient to notice or consider important [18]. Statistical terms help define meaningful changes in scores (Fig. 1): minimum clinically important difference (MCID), minimum important change (MIC), and the minimum detectable change (MDC). The MCID is the minimum difference in the scoring measure that the patient perceives as beneficial or harmful after treatment or a change in their health status compared with those who perceive no change [2, 16, 17]. The MCID was defined as the difference in the mean change in the WOMAC score between patients with "no" improvement compared to those with "little" improvement according to the anchor question (Fig. 1). The MIC is the change in the scoring measure relative to the baseline score for a group of patients who have perceived a small change in their clinical state [2] and was defined as the change in the WOMAC score relative to the baseline score for patients who reported a little improvement in their quality of life (Fig. 1).
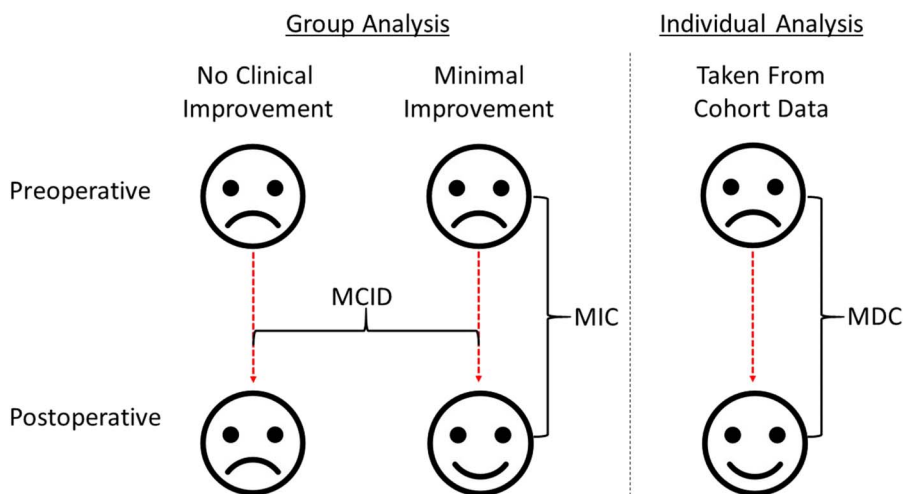

Fig. 1 Diagrammatic representation shows how three meaningful changes are calculated following an operative intervention.

The MDC is the smallest change for an individual who is likely to be beyond the measurement error or uncertainty of the scoring tool and represents true change (Fig. 1) [2]. These definitions are often used interchangeably and can cause confusion [2, 9, 19]. Three previous studies have sought to define the MCID in the WOMAC score after knee surgery. Other studies seem to have defined the MIC [12, 13, 15], which often is larger than the MCID, and so if used for sample-size calculations risks, results in an underpowered study. Another study used a simple "rule of thumb" method (half the SD) to calculate the MCID [25], which is not ideal since it is not derived from patient surveys but rather from data distributions, and so it may not represent a perceptible change as experienced by the patient. In addition, we are not aware of any paper defining the MCID, MIC, and the MDC for the components and total WOMAC score after TKA, or whether the MCID is influenced by preoperative case-mix variables or baseline functional scores.

The primary aim of this study was (1) **t**o identify the MCID for the total WOMAC score and its components 1 year after TKA using an anchor-based methodology. (2) To define the MIC and the minimum detectable change with 95% confidence (MDC95) for the total WOMAC score and its components 1 year after TKA.

Patients and Methods

Between 2003 and 2013, 3641 patients underwent primary TKA at one center. Of those, we excluded 460 patients (13%) from this retrospective study for prespecified reasons (mainly secondary OA and bilateral surgery); an additional 592 patients (16%) were either lost to followup or could not be included because of incomplete questionnaires (Fig. 2). We excluded patients who underwent simultaneous bilateral TKA, we excluded the second knee from analysis in patients who underwent staged-bilateral TKA, and we excluded patients who were diagnosed with periprosthetic joint infection.

During the study period, 2589 patients underwent TKA and had complete pre- and postoperative data that met the inclusion criteria. This included 1187 men and 1402 women, with an overall mean age of 69 years (± 10).

There were no differences in age (2.1 years, 95% confidence intervals [CIs], -3.1 to 7.2; p = 0.96), gender (odds ratio [OR] 0.9, 95% CI 0.8–1.1; p = 0.38), BMI (0.8 kg/m^2, 95% CI -2.8 to 4.4; p = 0.90), preoperative Short Form-12 (SF-12) physical component summary (PCS) (0.7 points, 95% CI -1.8 to 2.5; p = 0.49) or mental component summary (MCS) (0.5 points, 95% CI -1.1 to 2.1; p = 0.71) or WOMAC pain (1.1 points, 95% CI -3.8 to 6.0; p = 0.77), stiffness (2.8 points, 95% CI -4.2 to 9.8; p = 0.92), function (1.8 points, 95% CI -2.0 to 5.6; p = 0.82), or total (1.7 points, 95% CI -1.9 to 5.3; p = 0.60)

scores between those lost to followup and the study cohort.

Outcomes Measured

We used the 3.1 version of the WOMAC preoperatively and 1 year postoperatively, which has been standardized in English for a British population. It consists of 24 questions each answered using a five-point Likert scale (none, mild, moderate, severe, and extreme) [3]. The score can be reported as a total or for three separate subscales assessing pain, physical function, and stiffness. The pain subscale consists of five questions that are each scored from 0 to 4, hence overall scores range from 0 (absent) to 20 (most). If a response was missing, we calculated the average and then multiplied by 5 [5]. The functional subscale consists of 17 questions that are each scored from 0 to 4, and the overall score ranged from 0 (best) to 68 (worst). If there were three or fewer missing responses, we calculated an average and multiplied by 17. The stiffness subscale consists of two questions that are scored from 0 to 4, and the overall score ranged from 0 (absent) to 8 (most). Scores were considered invalid if more than one, more than three, or any were missing from the pain, function, and stiffness subscales, respectively. The scores were standardized as a percentage in a reverse format consisting of a range from 0 (worst) to 100 (best) in accordance to current guidance [24].

The SF-12 is a 12-item self-assessment health questionnaire that evaluates overall generic physical health (PCS) and mental health (MCS) [30]. The PCS and MCS scores range from 0 (worst) to 100 (best). This score was included to account for confounding variables of generic physical and mental health, which have previously been shown to influence PROMs after TKA [7, 8].

One year after TKA, we assessed quality of life improvement by asking "How much did the knee replacement surgery improve the quality of your life?" This was used as the anchor question. A five-point Likert scale was used to record the response to the question: a great improvement, moderate improvement, little improvement, no improvement at all, or the quality of my life is worse. This was used as the anchor question to define a minimum clinically important effect, which is an established technique [9]. There were greater improvements in the components and total WOMAC scores with increasing level of improvement in quality of life gained at 1 year (Table 1). In all, 211 patients declared that they had little improvement, and 115 patients had no improvement in their quality of life.

MCID

The MCID was defined as the difference in the mean change in the WOMAC score between patients with "no"
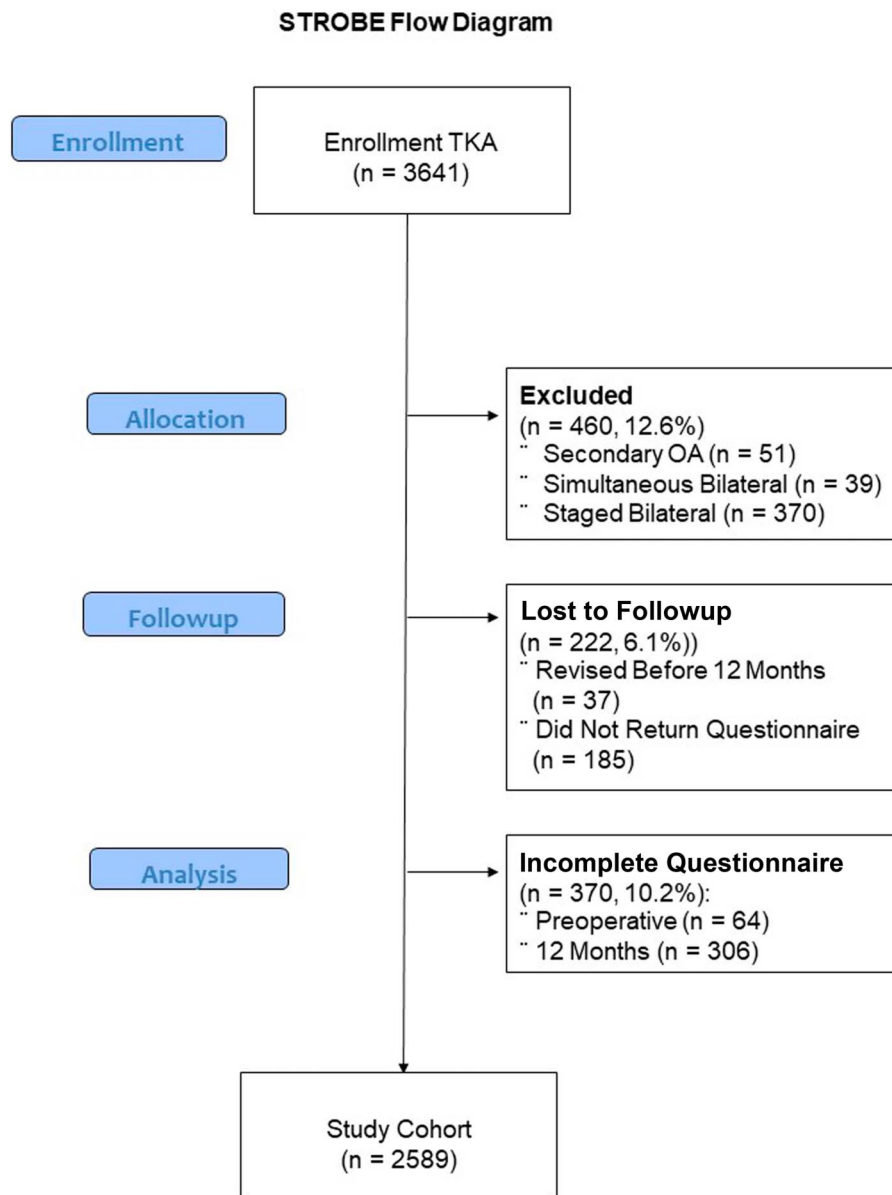
STROBE Flow Diagram

Fig. 2 STROBE flow diagram demonstrates the enrollment of patients, exclusion, and loss to followup for the study cohort.

improvement compared with those with "little" improvement according to the anchor question. This value was calculated as the difference in the mean change in the components and total WOMAC scores for patients who defined their quality of life as having had no improvement with those who reported little improvement on the patient-reported anchor question. The no improvement group was defined as the baseline, and using the anchor question the minimum difference would relate to the next available group, which is patients with little improvement [9]. We used linear regression analysis to adjust for potential preoperative confounding variables (age, gender, BMI, and

preoperative WOMAC score) to identify the MCID for the components and total WOMAC scores [1].

MIC

The MIC was defined as the change in the WOMAC score relative to the baseline score for patients who reported little improvement in their quality of life according to the anchor question. We used receiver operating characteristic (ROC) curve analyses to identify thresholds for the change in the components and total WOMAC scores that predicted

Table 1. Change in the components and total WOMAC score according to the perceived level of improvement in quality of life 1 year after TKA for the study cohort (n = 2589)

Variables	Improvement in quality of life					p value*
	Great	Moderate	Little	No	Worse	
Patients (number, %)	1583 (61.1)	613 (23.7)	211 (8.1)	115 (4.4)	67 (2.6)	
WOMAC (%, mean ± SD)						
Pain	52 ± 20	39 ± 20	29 ± 22	18 ± 19	15 ± 22	< 0.001
Function	45 ± 19	29.5 ± 18.3	21.9 ± 17	11.9 ± 15	7 ± 19	< 0.001
Stiffness	42 ± 25	27 ± 23	18 ± 22	11.5 ± 19	7 ± 27	< 0.001
Total	46 ± 18	31 ± 17	23 ± 17	13.0 ± 14	8 ± 19	< 0.001

*ANOVA = analysis of variance.

patients with a little improvement compared with those with no improvement. The area under the ROC curve ranges from 0.5, indicating a test with no accuracy, to 1.0, which is an accurate test. The threshold is equivalent to the point (WOMAC score) at which the sensitivity and specificity are maximal in predicting a little improvement [14].

MDC95

We used a distribution method based on the standard error of measurement to calculate the MDC95. The 95 signifies that with a 95% CI, a change greater than this is real and not due to intrinsic variability on the WOMAC score. One method of calculating this 95% CI is to use the standard error of measurement, which is the range in which a patient's true score lies (that is, the error associated with the measuring tool used). The standard error of measurement is calculated using the SD and the reliability of the measuring tool: standard error of measurement = SD x $\sqrt{1\text{-reliability}}$. We used Cronbach's alpha to calculate the reliability of the components and total WOMAC scores using the scores for each individual question. Acceptable reliability values of the alpha are thought to range from 0.70 to 0.95 [26]. The MDC was then calculated by multiplying the standard error of measurement by $\sqrt{2}$ (representing two separate occasions in which to measure change). The MDC was then multiplied by a z value representing the chosen CI, and for a 95% CI this value is 1.96, hence: MDC95 = standard error of measurement x $\sqrt{2}$ x 1.96.

Statistical Analysis

We used the Statistical Package for Social Sciences version 17.0 (SPSS Inc, Chicago, IL, USA) for all data analysis. Data was assessed for normality and parametric tests were appropriate. Linear variables were assessed using either an unpaired Student's t-test, or a one-way ANOVA with correction for multiple testing (Bonferroni). We used a chi-square test to

assess gender differences between groups. To adjust for confounding variables influencing the MCID, linear regression analysis (enter methodology was used–forced entry–of all variables) was used identify the actual MCID. Multiple imputation used the assigned random WOMAC scores and level of improvement in quality of life after surgery for those lost to followup (16%). Significance was set as a p value of < 0.05.

Ethics

The institutional review board at our center did not require ethical approval because the collected data was deemed to be part of service evaluation. The arthroplasty database is registered with our institution's audit department (Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK, Project Record Number 3290). The data collection was carried out in accordance with the General Medical Council guidelines for good clinical practice and the Declaration of Helsinki.

Results

MCID

There was no substantial difference in the case-mix variables and preoperative functional scores between patients who had little and no improvement in their quality of life (Table 2). After adjusting for confounding variables (age, gender, BMI, WOMAC and SF-12 scores) between the groups (little and no improvement) the MCID ranged from 8 for stiffness to 11 for pain. Multiple imputation for patients lost to followup increased the MCID to 12 points for the total WOMAC score in the regression model.

MIC

The ROC curve analysis differentiated patients who had a little improvement from those who had no improvement

Wolters Kluwer

Table 2. Demographic case-mix variables and preoperative functional scores according to group

Case-mix variables		Little (n = 211)	No (n = 115)	Difference/odds ratio (OR)	95% CI Lower	Upper	p value*
Age, years (mean ± SD)		68 ± 11	68 ± 10	0.1 difference	-2.4	2.5	0.98
Sex, number (%)	Male	90 (43)	71 (62)	0.83 OR	0.52	1.33	0.44†
	Female	121 (57)	44 (38)				
BMI, kg/m² (mean ± SD)		31 ± 6	31 ± 6	0.2 difference	-1.3	1.4	0.97
WOMAC, % (mean ± SD)							
	Pain	31 ± 16	31 ± 19	0.1 difference	-3.9	4.0	0.97
	Function	31 ± 15	32 ± 18	0.4 difference	-3.1	3.9	0.81
	Stiffness	35 ± 20	33 ± 18	1.1 difference	-3.3	5.5	0.62
	Total	31 ± 15	32 ± 16	2.3 difference	-3.2	3.6	0.90
SF-12 (mean ± SD)							
	PCS	26 ± 7	26 ± 7.4	0.2 difference	-1.4	1.8	0.77
	MCS	43 ± 13	42 ± 14	1.4 difference	-1.6	4.5	0.36

*Student's t-test unless otherwise stated.
†chi-square test; CI = confidence interval; BMI = body mass index; PCS = physical component summary; MCS = mental component summary.

(Fig. 3). The MIC (threshold with maximal specificity and sensitivity) ranged from 13 for stiffness to 21 for pain (Table 3). The MIC for the total WOMAC score was 17, however after multiple imputation for patients lost to followup this increased the MIC to 20 using the ROC curve analysis.

MDC95

The functional component and total WOMAC scores were the most reliable measures with a Cronbach's alpha of greater than 0.9 (Table 3). The MDC95 ranged from 11 for function to 27 for stiffness (Table 3). The MDC95 for the total WOMAC score was 12.

Discussion

Patients do not perceive statistical differences, they perceive effect sizes. The MCID defines the smallest improvement that a patient would describe as clinically important, and should be an important metric that readers and clinician-scientists consider when they evaluate therapeutic claims in clinical research studies [18]. By contrast, the MIC defines the smallest change in health status that a patient would perceive, and the MDC95 ensures that the tool being used to measure MCID and MIC can do so reliably. To our knowledge, this is the first study to define the MCID and MIC for the WOMAC score using an anchor-based method after TKA. The MCID for the total WOMAC score was 10, the MIC was 17 and the MDC95

was 12. Using these defined values, we can state that a 10-point (or more) difference between two groups or a 17-point (or more) improvement in the WOMAC for a cohort of patients (from before an intervention to after that intervention) is perceived by the patient. The MDC95 of 12 demonstrates that the 17-point MIC is measurable using the WOMAC score since the MDC95 is smaller than the MIC.

The major limitation of this study was the fact that we lost 592 patients (16%) to followup. We attempted to assess the influence of the missing data using multiple imputation; however, this increased the MCID and MIC by 2 to 4 points. This was due the random assignment of group (quality of life) and WOMAC score, which is an accepted technique when assessing an overall mean score. This study assessed the difference between two small subgroups (little [8%] and no [4%] improvement in quality of life) to define the MCID and MIC. However, assigning random WOMAC scores and subgroups for the 592 patients (16%) lost to followup resulted in an increase in the MCID and MIC values due to the overall greater mean scores in patients with moderate and great improvement. The aim of this study was to define the "minimum" values for the MCID and MIC, and as such, the values calculated for the cohort, before adjusting for loss to followup, achieve this goal. There were no differences in patient demographics or preoperative scores (WOMAC and SF-12) between those lost to followup and the study cohort. In addition, using the anchor-based methodology to define the MCID and MIC should not be influenced by loss to followup, as we assessed the relationship between specific responses to the anchor question and the WOMAC scores.
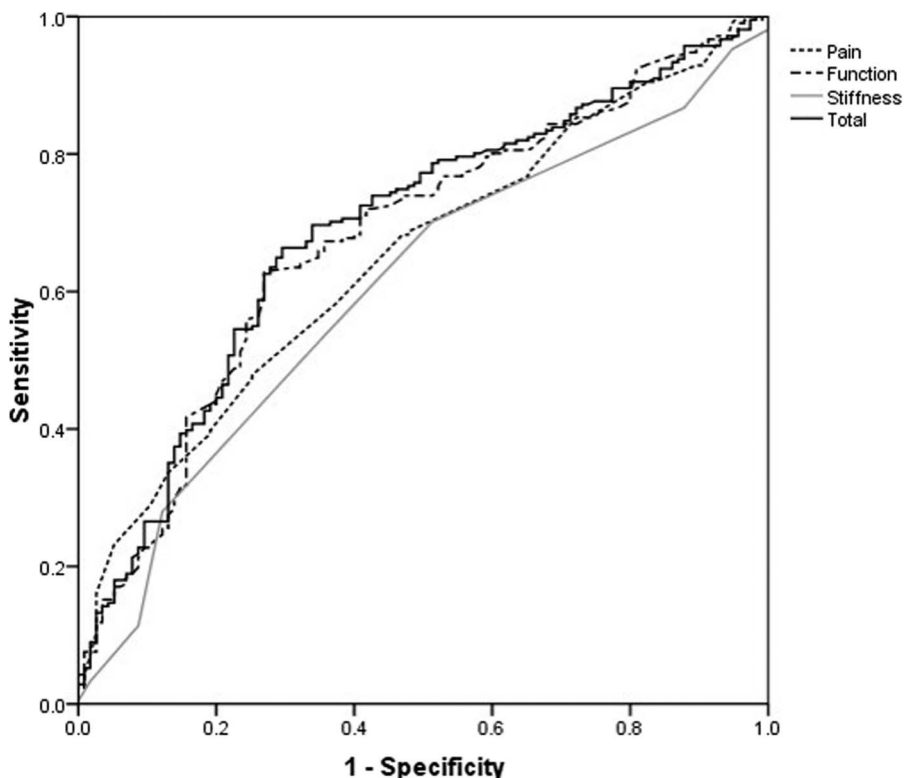
Fig. 3 ROC curves predict little improvement from no improvement for the components and total WOMAC scores.

Another major limitation was the use of a quality-of-life anchor question to define the pain, function, and stiffness components of the WOMAC. In retrospect, these may have been better defined if we had used different anchor questions for each of the components that focused on pain, function, and stiffness. Due to the retrospective design this was not possible; however, the quality-of-life anchor question used may relate more to the total WOMAC score because it is a combination of all three components. Also, the relatively early assessment of the quality of life of patients after their TKA could also be a limitation because their evaluation may change with longer followup. The evidence shows that some patients may change their perceptions about the effectiveness of the arthroplasty after the first year [4, 6]; although different from quality of life assessment, these parameters may be related. Data from the Swedish Joint Registry found the overall level of satisfaction after TKA to be relatively stable after 1 year for patients who had not undergone revision with longer followup [22]. Future studies should assess whether the same MCID and MIC are observed in the longer term.

We believe the MCID identified in our study, after adjusting for confounding variables, represents the most-accurate figures available for the WOMAC score after

Table 3. ROC curve analysis identifying the MIC (threshold), reliability, standard error of the measure (SEM), and MDC95 for the components and the total WOMAC scores

| WOMAC | R^2 | MCID | 95% CI | | p value* |
			Lower	Upper	
Pain	0.34	11	7.2	15.6	< 0.001
Function	0.28	9	5.6	12.7	< 0.001
Stiffness	0.32	8	3.2	11.8	< 0.001
Total	0.29	10	6.1	12.9	< 0.001

*Linear regression analysis: all preoperative variables from Table 2 were included in each model using "enter" methodology; MCID = minimum clinically important difference; CI = confidence interval.

TKA. Used in conjunction with the SD for the change in the scores (Table 1), the MCID could determine sample sizes in future clinical trials. Depending on the primary outcome assessment–pain, function, stiffness, or the total WOMAC score–the study could be powered using the defined values identified in our study, which are independent of case-mix variables and preoperative functional status. The MCID value is crucial to future comparative studies. For example, a power calculation using a previously defined MCID for the functional component of WOMAC of 19 points [13] and a SD of 18 for an alpha of 0.05, a two-way analysis and an 80% power, indicates that 32 patients would need to be randomized. Whereas using the 9-point MCID from the current study shows that 128 patients would need to be randomized. Therefore, the higher MCID would risk underpowering trials and may lead to a type II error.

Escobar et al. [13] first described the MCID for the WOMAC score after TKA. However, it could be argued that they defined the MIC because they did not use the difference between the "equal" and "somewhat better" groups. Instead, they used the absolute improvement in the somewhat group, reporting MIC values similar to our paper. Interestingly, if the MCID is calculated from their data using the definition from our study (the difference in the mean change in the WOMAC score between patients with no improvement compared with those with little improvement according to the anchor question) the MCID is similar to that found in our study (pain 11, function 10, stiffness 7). More recently, the same group used ROC curves to define the MCID to predict the somewhat better group and the equal group using the absolute improvement, but again this was more in keeping with the MIC [12]. Greco et al. [15] aimed to identify the MCID in the WOMAC after knee cartilage surgery but also used ROC curves to identify it, which is more in keeping with MIC. Furthermore, they combined much- and somewhat-better as the improved group and the slight improvement, not improved, and worse as the no-improvement group, which may not represent a minimum important change in the scoring measure. Soohoo et al. [25] used half the SD for change in the WOMAC at 3 months, but combined data for both TKA and THA. Although they found a similar figure for the MCID for the total WOMAC score of 10, just as we did in our study, it is accepted the WOMAC score after THA responds differently [20]. In a recent systematic review of 13 studies, the authors aimed to define the MCID (or MID); they found that the MCID for the pain component was 12 and 13 for the functional component [11]. Their systematic review included heterogenous studies assessing various interventions, from nonoperative to operative, and may have included studies quoting the MIC and not the MID [12,13] and vice versa [27, 28].

The MIC value represents a change in the score beyond which an individual is considered to have experienced a clinical change, and for the current study was defined as the change in the WOMAC score relative to the baseline score for patients who reported a little improvement in their quality of life. This is an important variable when assessing the outcome of a study cohort's change in health status from before to after TKA. For example, depending on a study's sample size and other parameters, a 10-point change in the total WOMAC score may be associated with a p value smaller than 0.05, but our data suggest that such a small change would not be clinically important to the patient (since the MIC we identified was ≥ 17). We are not aware of other evidence defining the MIC for the components or total WOMAC score after TKA. However, as mentioned above, Escobar et al. [13] may have identified the MIC with similar values (pain 23, function 19, and stiffness 15). In the current study, we used ROC curve analysis to identify the MIC, but the AUC was low (at approximately 0.7), which is the lower limit for reliability. The identified MIC allows individual patients to be assessed as to whether they have achieved a clinically relevant change in their WOMAC score(s).

MDC95 represents the smallest change in an individual patient score that is beyond the measurement error of the tool used. This is an important figure when accounting for the MIC, as a MDC95 less than the MIC may not be detectable due to the variation of the measure used. For example, the MIC for the stiffness component of the WOMAC is 13 points, but the MDC95 is 27 and therefore this component is not sensitive enough to differentiate a 13-point difference due to intrinsic score variation. Escobar et al. [13] identified MDC95 values for the WOMAC score after TKA, which were similar to the figures identified in the current study (pain 22, function 13, and stiffness 29). It is interesting to note that the most reliable component of the WOMAC score was the functional assessment and the total WOMAC score, with a Cronbach's alpha greater than 0.90. The MIC for the functional component of 16 was greater than the MDC95, signifying that it can detect the MIC. This may indicate that if future studies use the WOMAC score, they should be powered to either the total or functional component of the WOMAC to ensure reliable results. However, the MDC95 presented may be smaller than when a Chronbach's alpha test-retest data is used, but even a 5-point increase in the MDC95 for the total or functional component of the WOMAC would retain the reliability of measuring MIC.

The MCID and MIC for the WOMAC score represent the smallest meaningful effect sizes when comparing the outcome of two groups (difference in mean change between the groups) or when assessing a cohort (a change in score for the group) after TKA, respectively, helping the

reader to distinguish between a clinically important effect size and a mere statistical difference. We determined that the error in measurement (based on the MDC95) for the function component and total WOMAC scores were less than the MIC, which suggests changes beyond the MIC are clinically real and not due to uncertainty in the score. These parameters are essential to interpret TKA outcomes research and to ensure clinical research studies are amply powered to detect meaningful differences. Future studies using the WOMAC score to assess TKA outcomes should report not only the statistical significance (a p value) but also the clinical importance using the reported MCID and MIC values.

References

1. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol.* 2017;82:128–136.

2. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol.* 2015;68:73–79.

3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–1840.

4. Brander V, Gondek S, Martin E, Stulberg SD. Pain and depression influence outcome 5 years after knee replacement surgery. *Clin Orthop Relat Res.* 2007;464:21–26.

5. Burch FX, Tarro JN, Greenberg JJ, Carroll WJ. Evaluating the benefits of patterned stimulation in the treatment of osteoarthritis of the knee: a multi-center, randomized, single-blind, controlled study with an independent masked evaluator. *Osteoarthritis Cartilage.* 2008;16:865–872.

6. Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. Three groups of dissatisfied patients exist after total knee arthroplasty: early, persistent, and late. *Bone Joint J.* 2018;100-B:161–169.

7. Clement ND, Burnett R. Patient satisfaction after total knee arthroplasty is affected by their general physical well-being. *Knee Surg Sports Traumatol Arthrosc.* 2013;21:2638–2646.

8. Clement ND, MacDonald D, Burnett R. Primary total knee replacement in patients with mental disability improves their mental health and knee function: a prospective study. *Bone Joint J.* 2013;95-B:360–366.

9. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395–407.

10. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br.* 1998;80:63–69.

11. Devji T, Guyatt GH, Lytvyn L, Brignardello-Petersen R, Foroutan F, Sadeghirad B, Buchbinder R, Poolman RW, Harris IA, Carrasco-Labra A, Siemieniuk RAC, Vandvik PO. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ Open.* 2017;7:e015587.

12. Escobar A, Garcia PL, Herrera-Espineira C, Aizpuru F, Sarasqueta C, Gonzalez Saenz de TM, Quintana JM, Bilbao A. Total knee replacement; minimal clinically important differences and responders. *Osteoarthritis Cartilage.* 2013;21:2006–2012.

13. 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.

14. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain.* 2001;94:149–158.

15. Greco NJ, Anderson AF, Mann BJ, Cole BJ, Farr J, Nissen CW, Irrgang JJ. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. *Am J Sports Med.* 2010;38:891–902.

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

17. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:171–184.

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

19. Lydick E, Epstein RS. Interpretation of quality of life changes. *Qual Life Res.* 1993;2:221–226.

20. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage.* 2005;13:1076–1083.

21. Ramkumar PN, Harris JD, Noble PC. Patient-reported outcome measures after total knee arthroplasty: a systematic review. *Bone Joint Res.* 2015;4:120–127.

22. Robertsson O, Dunbar M, Pehrsson T, Knutson K, Lidgren L. Patient satisfaction after knee arthroplasty: a report on 27,372 knees operated on between 1981 and 1995 in Sweden. *Acta Orthop Scand.* 2000;71:262–267.

23. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)–validation of a Swedish version. *Scand J Med Sci Sports.* 1998;8:439–448.

24. Singh J, Sloan JA, Johanson NA. Challenges with health-related quality of life assessment in arthroplasty patients: problems and solutions. *J Am Acad Orthop Surg.* 2010;18:72–82.

25. 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.

26. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ.* 2011;2:53–55.

27. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, De Vet HC. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol.* 2010;63:524–534.

28. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC. Linking measurement error to minimal important change of

patient-reported outcomes. *J Clin Epidemiol.* 2009;62: 1062–1067.

29. Walker LC, Clement ND, Bardgett M, Weir D, Holland J, Gerrand C, Deehan DJ. The WOMAC score can be reliably used to classify patient satisfaction after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2018. [Published online ahead of print February 26, 2018]. DOI: 10.1007/s00167-018-4879-5.

30. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–233.

31. Woolacott NF, Corbett MS, Rice SJ. The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials. *Rheumatology (Oxford).* 2012; 51:1440–1446