# Interpretation threshold values for the Oxford Knee Score in patients undergoing unicompartmental knee arthroplasty

Lasse K HARRIS [1], Anders TROELSEN [1], Berend TERLUIN [2], Kirill GROMOV [1], Andrew PRICE [3], and Lina H INGELSRUD [1]

[1] Department of Orthopaedic Surgery, Copenhagen University Hospital Hvidovre, Copenhagen Denmark; [2] Department of General Practice, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, The Netherlands; [3] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK
Correspondence: lasse.kindler.harris@regionh.dk

**Background and purpose** — Developing meaningful thresholds for the Oxford Knee Score (OKS) advances its clinical use. We determined the minimal important change (MIC), patient acceptable symptom state (PASS), and treatment failure (TF) values as meaningful thresholds for the OKS at 3-, 12-, and 24-month follow-up in patients undergoing unicompartmental knee arthroplasty (UKA).

**Patients and methods** — This is a cohort study with data from patients undergoing UKA collected at a hospital in Denmark between February 2016 and September 2021. The OKS was completed preoperatively and at 3, 12, and 24 months postoperatively. Interpretation threshold values were calculated with the anchor-based adjusted predictive modeling method. Non-parametric bootstrapping was used to derive 95% confidence intervals (CI).

**Results** — Complete 3-, 12-, and 24-month postoperative data was obtained for 331 of 423 (78%), 340 of 479 (71%), and 235 of 338 (70%) patients, median age of 68–69 years (58–59% females). Adjusted OKS MIC values were 4.7 (CI 3.3–6.0), 7.1 (CI 5.2–8.6), and 5.4 (CI 3.4–7.3), adjusted OKS PASS values were 28.9 (CI 27.6–30.3), 32.7 (CI 31.5–33.9), and 31.3 (CI 29.1–33.3), and adjusted OKS TF values were 24.4 (CI 20.7–27.4), 29.3 (CI 27.3–31.1), and 28.5 (CI 26.0–30.5) at 3, 12, and 24 months postoperatively, respectively. All values statistically significantly increased from 3 to 12 months but not from 12 to 24 months.

**Interpretation** — The UKA-specific measurement properties and clinical thresholds for the OKS can improve the interpretation of UKA outcome and assist quality assessment in institutional and national registries.

Unicompartmental knee arthroplasty (UKA) is deemed a viable alternative to total knee arthroplasty (TKA) for patients with severe knee osteoarthritis with a certain wear pattern (1). Patient-reported outcome measures (PROMs) are increasingly used to evaluate treatment effectiveness and quality of care from a patient-centered perspective (2). The Oxford Knee Score (OKS) is frequently used to assess pain and functional limitations after knee arthroplasty, on a scale ranging from 0 to 48 (worst to best) (3,4). However, meaningful interpretation of PROM data is challenging, as statistically significant improvements are not necessarily clinically meaningful (5). To help assign meaning to PROM scores, 3 interpretation threshold concepts have been suggested. The minimal important change (MIC) concept defines the smallest change score that is deemed important by the average patient (6). The patient acceptable symptom state (PASS) concept defines the score above which patients consider themselves well (7). The treatment failure (TF) concept defines the score below which patients consider their treatment to have failed (8).

Interpretation threshold values are considered context-specific, highlighting the importance of investigating possible differences across patient populations (5,9,10). No previous studies have determined MIC or TF values, but a PASS value of 41.5 points for the OKS in people undergoing UKA at 24 months postoperatively has been suggested (11). Additionally, in people undergoing TKA, a MIC value of 6.9 for the OKS at 6 months' follow-up and TF values of 27 at 12 and 24 months has been presented (12,13). Although UKA is deemed a viable alternative to TKA it remains to be investigated as to whether PROM scores should be interpreted alike across both patient populations. There is a potential variability in interpretation threshold values across patient populations that is

unclear. Therefore, we determined the MIC, PASS, and TF for the OKS at 3-, 12-, and 24-month follow-up after undergoing UKA.

## Patients and methods

### *Study design and setting*

This is a cohort study using data from a Danish hospital's local arthroplasty registry. Between February 2016 and September 2021, all patients with scheduled UKA were asked to complete an electronic questionnaire during their preoperative visit to the hospital. Electronic follow-up questionnaires were emailed to patients at 3, 12, and 24 months postoperatively. 2 reminder emails with a 2-week interval and ultimately a paper version of the questionnaire were sent by postal mail to the patients if they failed to complete the electronic questionnaire or were without an email address. The yearly use of UKA at the hospital increased from 9% to 58% during the study period, mainly because more surgeons have adopted the surgical procedure and they all follow the same indication for UKA recommended by Hamilton et al. (**14**).

### *Participants*

The data was from patients undergoing medial UKA, which has been routinely collected at the Danish hospital. The inclusion criterion was patients undergoing primary surgery for knee osteoarthritis. The exclusion criterion was patients undergoing revision surgery. If patients were registered as having bilateral medial UKAs within the study period, we selected the first to be included.

### *Questionnaires*

The OKS is a 12-item questionnaire assessing degree of pain and function summed to a total score between 0 and 48 (worst–best) (**3**). Adequate validity, reliability, and responsiveness characteristics for the OKS in patients undergoing knee arthroplasty has been reported (**15**). Additionally, at each postoperative time-point, 3 anchor questions were responded to (Table 1, see Supplementary data). First, patients were asked whether they had experienced overall changes in symptoms since the knee surgery: "How are your knee problems now compared with prior to your operation?" Response options were on a 7-point scale (**16**). Patients answering "better, an important improvement" or "somewhat better, but enough to be an important improvement were classified as being importantly improved. Patients answering "worse, an important deterioration" or "somewhat worse, but enough to be an important deterioration" were classified as being importantly deteriorated. Patients answering "about the same" or "very small improvement/deterioration, not enough to be an important improvement/deterioration" were classified as unchanged. Second, patients were asked: "Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?" (yes/no) (**7**). Finally, if the patients responded in terms of not having a satisfactory symptom state, they were asked: "Would you consider your current state as being so unsatisfactory that you think the treatment has failed?" (yes/no) (**8**).

### *Statistics*

Patient characteristics were reported as mean (SD) or median (IQR) for continuous variables, and as frequency and percentage distribution for categorical variables. The OKS change score distributions across anchor response options were investigated using boxplots.

An anchor-based approach was used to calculate interpretation threshold values. This approach involved anchoring the OKS to anchor question responses. We used the predictive modeling method developed to estimate MIC thresholds because of the reported methodological advantages compared with the commonly used receiver operating characteristic (ROC) method (**17**). The predictive modeling method is centered on a logistic regression using the dichotomized anchor response as the dependent variable and the change in OKS for MIC improvement, or the postoperative OKS for PASS and TF, as the independent variable. The thresholds were the OKS corresponding to a likelihood ratio of 1, which means that the postoperative odds of being importantly improved or having a satisfactory symptom state are the same as the preoperative odds for improvement or having a satisfactory symptom state (**17**). The predictive modeling method, and the ROC method, is biased if the proportion being importantly improved or having a satisfactory symptom state differs from 50%. This biases results in overestimation of the threshold if the proportion is greater than 50% or underestimation if the proportion is smaller than 50%. Consequentially, we used an adjustment to the threshold for unequal proportions of patients with the following equation proposed by Terluin et al. (**18**):

$$MIC_{adjusted} = MIC_{pred} - (0.090 + 0.103 * Cor) * SD_{change} * \text{log-odds(imp)}.$$

In this equation, Cor is the point biserial correlation between postoperative OKS and the anchor, $SD_{change}$ is the SD of the OKS change score, and log-odds(imp) is the natural logarithm of (proportion improved/[1 – proportion improved]). Similarly, we used the equation fitted for PASS and TF thresholds to adjust for unequal proportions of people with a satisfactory symptom state. Additionally, we used bootstrapping (n = 1,000) to obtain 95% confidence intervals (CI) reported as 0.025–0.975 quantiles. Furthermore, we tested whether threshold values differed statistically between the 3 follow-up timepoints by evaluating the 95% CI around the mean differences between the 1,000 bootstrap samples for each timepoint, calculated as the 0.025–0.975 quantiles of differences.

We additionally calculated interpretation threshold values using the ROC method, enabling the comparison of the predictive modeling method with this traditional method. Opti-

mal values were identified using the Youden index (19).

Correlations were calculated to assess anchor validity. Point-biserial correlation was calculated for dichotomized MIC anchors and the change in OKS, and for both PASS and TF anchors and the postoperative scores. Polyserial correlation was additionally calculated for the 7-level MIC anchor responses and change, plus preoperative and postoperative OKS.

We investigated baseline dependency of MIC values by randomly splitting the OKS item set into 2 separate scales, using one scale to stratify into low and high baseline subgroups, and the other scale to calculate MIC values (20). Splitting the OKS is a workaround to create an independent second baseline measurement to avoid redistributing measurement error by stratifying on the baseline score. Because some variation may occur depending on the exact division of the items, we repeated the random splitting of OKS 5 times and estimated the average MIC value for each baseline group, as recommended (20). Baseline dependency of PASS and TF values was investigated by calculating these values on median-split datasets. Finally, we tested whether threshold values were statistically significantly different between the baseline groups by performing item-split (MIC) and median-split (PASS and TF) analyses on 1,000 bootstrap samples. We calculated mean differences and reported 95% CI as 0.025–0.975 quantiles of the mean differences. For all analyses, R version 4.1.2 (http://www.r-project.org/) was used.

### Ethics, funding, and potential conflicts of interest

This study was carried out in accordance with the Helsinki declaration. The local arthroplasty registry has been approved by the Danish Data Protection Agency (Journal number HVH-2012-048). In Denmark, register-based studies using only questionnaire data require no approval from the ethical committee. The Department of Orthopaedic Surgery at the hospital fully funded this project. No potential conflicts of interest are declared by the authors in relation to this study.

## Results

### Participants

Complete data was obtained for 331 of 423 (78%), 340 of 479 (71%), and 235 of 338 (70%) patients at 3-, 12-, and 24-month
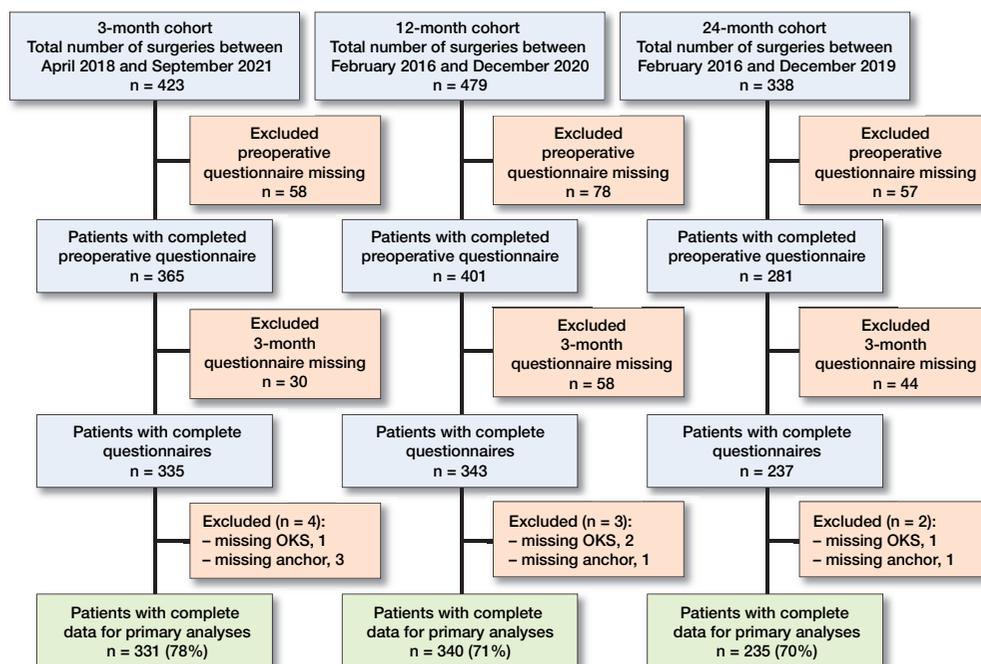


Figure 1. Flowchart of patients enrolled. OKS, Oxford Knee Score.

Table 2. Patient demographics and preoperative characteristics. Values are count (%) unless otherwise specified

| Factor | 3 months n = 331 | 12 months n = 340 | 24 months n = 235 |
|---|---|---|---|
| Age [a] | 69 (61–74) | 68 (61–74) | 68 (60–74) |
| Female sex | 194 (59) | 196 (58) | 138 (59) |
| BMI [a] | 29 (26–34) | 29 (25–33) | 28 (25–32) |
| ASA | | | |
| 1 | 20 (6) | 26 (8) | 23 (10) |
| 2 | 257 (78) | 255 (75) | 176 (75) |
| 3 | 53 (16) | 58 (17) | 36 (15) |
| 4 | 1 (0) | 1 (0) | – |
| KL grade | | | |
| 2 | 3 (1) | 10 (3) | 13 (5) |
| 3 | 79 (24) | 92 (27) | 68 (29) |
| 4 | 249 (75) | 238 (70) | 154 (66) |
| OKS [a] | 23 (17–28) | 24 (19–28) | 24 (19-29) |
| EQ5D index [a] | 0.66 (0.59–0.72) | 0.72 (0.62–0.72) | 0.72 (0.63–0.72) |
| EQ5D VAS [a] | 70 (50–80) | 70 (50–80) [b] | 70 (51–80) [b] |

[a] Values are median (0.025–0.975 quantile range).
[b] Missing data, n = 1.
KL grade: Kellgren & Lawrence classification. OKS: Oxford Knee Score. EQ5D: EuroQol 5-Dimension. VAS: visual analog scale.

follow-up, respectively (Figure 1). At surgery, patients responding at the follow-up timepoints had a median age of 68–69 years and 58–59% were female (Table 2). Patients with complete data and patients with missing data had similar age characteristics. Patients with missing data were more often male in the 3-month group, had higher BMI in the 12-month group, and had worse preoperative OKS and lower overall self-rated health in the 12- and 24-month groups compared with patients with complete data (Table 3, see Supplementary data).

Table 4. Proportion of patient responses to minimal important change anchor question at 3, 12, and 24 months after surgery. Values are count (%)

| Factor | 3 months n = 331 | 12 months n = 340 | 24 months n = 235 |
|---|---|---|---|
| Importantly improved | | | |
| Better, an important improvement | 224 (68) | 257 (76) | 183 (78) |
| Somewhat better, but enough to be an important improvement | 64 (19) | 44 (13) | 25 (11) |
| Unchanged | | | |
| Very small change, not enough to be an important improvement | 21 (6) | 14 (4) | 8 (3) |
| About the same | 7 (2) | 7 (2) | 4 (2) |
| Very small change, not enough to be an important deterioration | 2 (1) | 2 (1) | 2 (1) |
| Importantly deteriorated | | | |
| Somewhat worse, but enough to be an important deterioration | 9 (3) | 9 (2) | 12 (5) |
| Worse, an important deterioration | 4 (1) | 7 (2) | 1 (0) |

Table 5. Proportions of patients achieving a satisfactory symptom level, considering treatment failure, or neither at 3, 12, and 24 months after surgery. Values are count (%)

| Factor | 3 months n = 331 | 12 months n = 340 | 24 months n = 235 |
|---|---|---|---|
| Satisfactory symptom level | 271 (82) | 282 (83) | 199 (85) |
| Neither satisfactory symptoms nor treatment failure | 47 (4) | 32 (9) | 14 (6) |
| Treatment failure | 13 (4) | 26 (8) | 22 (9) |

At 3 months postoperatively, the overall percentage of patients reporting important improvements was 87%, while 4% reported being importantly deteriorated. 89% of patients reported important improvements at 12 and 24 months, while 4% and 5% reported being importantly deteriorated, respectively (Table 4).

Postoperative OKS change scores were generally higher for patients feeling importantly improved, in comparison with those feeling importantly deteriorated or unchanged in symptoms (Figure 2).

At 3 months postoperatively, 82% considered themselves to have satisfactory symptoms, while 4% considered their symptoms state as being so unsatisfactory that they considered the treatment to have failed. At 12 and 24 months the proportion of patients satisfied with their symptom level was 83% and 85%, while 8% and 9% considered the treatment to have failed, respectively (Table 5).

Postoperative OKS were generally higher for patients considering their symptom level to be satisfactory, in comparison with those considering the treatment to have failed or neither (Figure 3)

The point-biserial correlations between the dichotomized MIC anchor and the change in OKS were 0.43, 0.49, and 0.56 at 3, 12, and 24 months. Correlations between PASS and TF
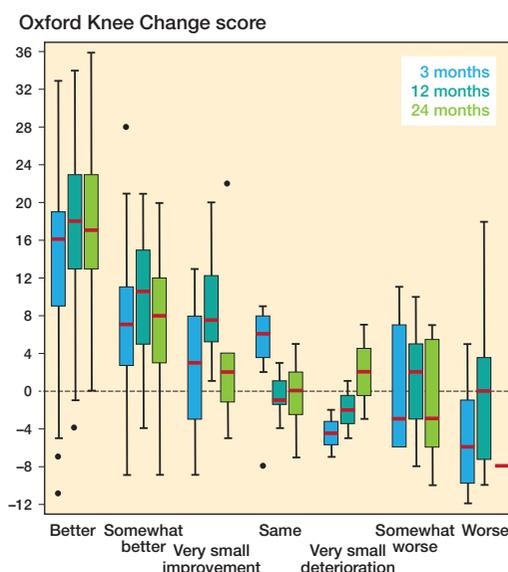


Figure 2. Oxford Knee Change scores at 3, 12, and 24 months postoperatively by minimal important change anchor question response categories ranging from "better, an important improvement" to "worse, an important deterioration." Horizontal bars present the median, the box the interquartile range (IQR), the whiskers the maximum and minimum scores within 1.5 * IQR from the box, and • represents outliers.
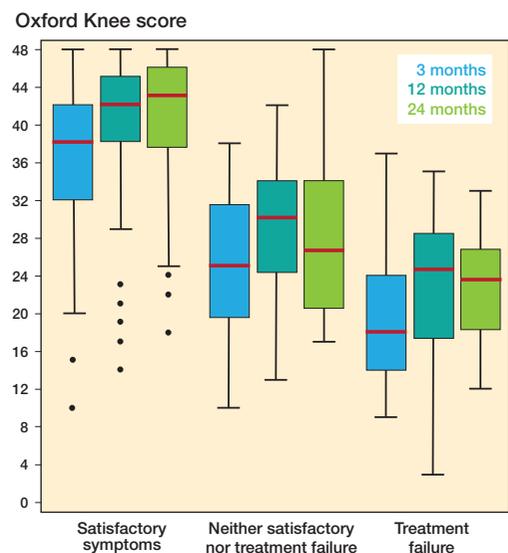


Figure 3. Oxford Knee Score distribution at 3, 12, and 24 months postoperatively for patients with satisfactory symptoms, considering the treatment to have failed, or neither. See Figure 2 for boxplot interpretation.

anchor questions and the postoperative OKS were 0.55 and 0.33 at 3 months, 0.67 and 0.53 at 12 months, and 0.67 and 0.59 at 24 months, respectively. Polyserial correlations for MIC anchor responses and change, preoperative and postoperative OKS, as well as point-biserial correlation for PASS and TF anchor responses and preoperative and postoperative OKS are presented in supplementary data (Table 6, see Supplementary data).

Table 7. Minimal important change (MIC), patient acceptable symptom state (PASS), and treatment failure (TF) cut-off values calculated with the adjusted predictive modeling method for the Oxford Knee Score at 3, 12, and 24 months after unicompartmental knee arthroplasty

| Follow-up | n | MIC value (CI) [a] | PASS value (CI) [a] | TF value (CI) [a] |
|---|---|---|---|---|
| 3 months | 331 | 4.7 (3.3–6.0) | 28.9 (27.6–30.3) | 24.4 (20.7–27.4) |
| 12 months | 340 | 7.1 (5.2–8.6) | 32.7 (31.5–33.9) | 29.3 (27.3–31.1) |
| 24 months | 235 | 5.4 (3.4–7.3) | 31.3 (29.1–33.3) | 28.5 (26.0–30.5) |

[a] 95% confidence intervals (CI) are the 0.025–0.975 quantiles of the 1,000 bootstrap threshold values.

### Interpretation threshold values

When MIC values were adjusted for the high proportion of improved patients the OKS threshold values were 4.7 (CI 3.3–6.0) at 3 months, 7.1 (CI 5.2–8.6) at 12 months, and 5.4 (CI 3.4–7.3) at 24 months postoperatively. When PASS values were adjusted for the high proportion having satisfactory symptoms the OKS values were 28.9 (CI 27.6–30.3), 32.7 (CI 31.5–33.9), and 31.3 (CI 29.1–33.3) at 3, 12, and 24 months, respectively. When TF values were adjusted for the small proportion considering their treatment to have failed the OKS values were 24.4 (CI 20.7–27.4) at 3 months, 29.3 (CI 27.3–31.1) at 12 months, and 28.5 (CI 26.0–30.5) at 24 months, respectively (Table 7).

The interpretation threshold values increased statistically from 3 to 12 months, but not from 12 to 24 months postoperatively (Table 8, see Supplementary data).

The interpretation threshold values were consistently higher for patients in the low baseline subgroup than in the high baseline subgroup for all postoperative timepoints (Table 9, see Supplementary data).

Interpretation threshold values calculated with the adjusted predictive modeling method were lower and the CIs were generally narrower compared with the ROC method (Table 10, see Supplementary data).

## Discussion

This cohort study from a Danish public hospital estimated interpretation threshold values for the OKS at 3-, 12-, and 24-month follow-up in patients undergoing UKA. Adjusted MIC values were 4.7, 7.1, and 5.4 points, adjusted PASS values were 28.9, 32.7, and 31.3 points, and adjusted TF values were 24.4, 29.3, and 28.5 points at 3, 12, and 24 months postoperatively, respectively. All values increased statistically from 3 to 12 months but not from 12 to 24 months.

The adjusted OKS MIC values we found lie in the range of previously published values. No studies have previously determined these values in patients undergoing UKA exclusively. 2 studies using the same methodological approach as ours found values of 7 and 8 points at 6 and 12 months, respec-

tively, in patients undergoing TKA (12,16). However, other TKA studies found values of 9 points at 6 months, and 5 points at 12 months, but by using different anchor questions and statistical approaches (21,22). These findings suggest that the postoperative OKS MIC scores in general are similar in patients undergoing UKA and TKA, but the values may depend on the statistical method used (12,16,21,22). We found a small but statistically higher adjusted OKS MIC value between 3 and 12 months postoperatively, suggesting that patients' expectations of pain-levels and knee function increase with time after undergoing UKA. However, the non-statistical difference between 12- and 24-month values also suggests that these patient expectations may stabilize after 12 months.

To our knowledge, only 1 study has previously determined adjusted OKS PASS values in people undergoing UKA at 24 months postoperatively (11). That study proposed a cut-off value of 41.5 points which is 10.2 points higher compared with our finding. The large difference could be explained by the ROC analysis used on a population where the proportion improved was very high (92.7%), possibly causing an upward biased value in the comparative study (11). However, our adjusted OKS PASS and TF values found at 3 and 12 months postoperatively are within 3 points of the proposed cut-offs previously suggested in a study using the same method for patients undergoing TKA (13). These findings suggest that the OKS PASS and TF scores are similar in patients undergoing UKA and TKA. We found that both adjusted OKS PASS and TF values increased from 3 to 12 months, suggesting that patients accept a higher symptom level early after surgery, while requiring better functional status at 12 months postoperatively. Additionally, we found that the adjusted OKS TF threshold values were between 2.8 and 4.5 points below PASS thresholds, suggesting that the area where people neither consider their symptom levels satisfactory nor consider their treatment to have failed is narrow and perhaps redundant. However, the low number of patients considering their treatment to have failed at 3 (n = 13 [4%]), 12 (n = 26 [8%]), and 24 months (n = 22 [9%]) makes these assumptions uncertain.

We demonstrated that using different statistical approaches yields different interpretation threshold values. First, the predictive modeling method derived cut-offs with greater precision (i.e., CIs were narrower) compared with the ROC method (17). Second, we demonstrated how the adjusted predictive modeling method altered the cut-offs as the proportion of patients being importantly improved, feeling satisfactory symptoms, or feeling treatment failure differed greatly from 50% (18). These findings align with previous studies, and emphasize the preference of the predictive modelling method above the ROC method (12,13,16).

Preoperative symptom status impacts on the interpretation threshold values. We demonstrated baseline dependency of the threshold values at all postoperative timepoints except

for TF at 3 months. Likewise, previous studies determined baseline dependency of OKS PASS and TF values, using a similar methodological approach, in patients undergoing TKA (13,23). For the MIC, previous results are sparse and conflicting with different methods used to evaluate baseline dependency (16,24). We demonstrated baseline dependency also of MIC values, using a newly developed method that avoids redistributing measurement error (20). The adjusted predictive model cut-offs for the low baseline subgroups were from 4.0 to 6.4 points lower than the high baseline subgroups. These findings support the notion that patients who are in a poor health condition need greater improvement to consider their change important, but are concurrently willing to accept an overall worse outcome than patients who are in a better health condition (25). The implication of baseline dependency is that when applying the threshold values, it is important to select the value that derives from a patient population with comparable preoperative status as the population under study.

Providing meaningful interpretation threshold values for the OKS has both scientific and clinical implications. They can help improve the interpretation of studies using OKS as an outcome measure. Additionally, arthroplasty registries collecting the OKS are provided with a tool to monitor quality of treatment from the patient-centered perspective. Furthermore, from a clinical perspective, the values at 3, 12, and 24 months postoperatively may be used as reference values for what the "average" patient undergoing UKA would deem as an important improvement, a satisfactory symptom state, and a state feeling that their treatment has failed. If the OKS is used in clinical practice, these interpretation thresholds could lead to greater understanding and better applicability for clinicians and patients in the shared decision-making process. Our study suggests that PROM scores can be interpreted using the same interpretation values across both UKA and TKA populations.

This study has limitations. The data having been collected at 1 public hospital in Denmark possibly limits the generalizability of the interpretation threshold values found in this study. Furthermore, between 70% and 78% of the patients receiving a UKA provided complete data, possibly introducing selection bias, further lowering the generalizability of the findings. It could be that patients answering the follow-up questionnaires are those generally feeling satisfied with their treatment result. However, considering hospital uptake area, coverage of both urban and rural geographical areas, and patient characteristics depicting the nationwide Danish Knee Arthroplasty Register supports the representativeness of our study population in a Danish context (26). Additionally, because the adjusted predictive modeling method requires normally distributed scores and change scores, this study could potentially provide biased values. Skewness in either direction may cause downward bias for the MIC and if the skew is right- or left-sided it causes downward or upward bias for the PASS and TF, respectively. Nonetheless, before the suggested values are applicable in other countries and cultures, they must be compared with similar data derived from preferably large-scale international registries.

In conclusion, we believe the development of UKA-specific measurement properties and clinical thresholds for the OKS may guide the interpretation of UKA studies using this PROM. Additionally, all values increased from 3 to 12 months postoperatively, implying that patients have higher expectations regarding their knee pain and function long term. Similar studies should investigate the external validity of these values.

1.  **Liddle A D, Judge A, Pandit H, Murray D W.** Adverse outcomes after total and unicompartmental knee replacement in 101 330 matched patients: a study of data from the National Joint Registry for England and Wales. Lancet 2014; 384(9952): 1437-45. doi: 10.1016/S0140-6736(14)60419-0

2.  **Weldring T, Smith S M S.** Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Heal Serv Insights 2013; 6: 61-8. doi: 10.4137/HSI.S11093

3.  **Murray D W, Fitzpatrick R, Rogers K, Pandit H, Beard D J, Carr A J, et al.** The use of the Oxford hip and knee scores. J Bone Joint Surg Br 2007; 89(8): 1010-14. doi: 10.1302/0301-620X.89B8.19424

4.  **Liddle A D, Pandit H, Judge A, Murray D W.** Patient-reported outcomes after total and unicompartmental knee arthroplasty. Bone Joint J 2015; 97-B(6): 793-801. doi: 10.1302/0301-620X.97B6.35155

5.  **King M T.** A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011; 11(2): 171-84. doi: 10.1586/erp.11.9

6.  **de Vet H C, Terwee C B, Ostelo R W, Beckerman H, Knol D L, Bouter L M.** Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006; 4:54. doi: 10.1186/1477-7525-4-54

7.  **Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al.** Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Ann Rheum Dis 2005; 64(1): 34-7. doi: 10.1136/ard.2004.023028

8.  **Ingelsrud L H, Granan L-P, Terwee C B, Engebretsen L, Roos E M.** Proportion of patients reporting acceptable symptoms or treatment failure and their associated KOOS values at 6 to 24 months after anterior cruciate ligament reconstruction: a study from the Norwegian Knee Ligament Registry. Am J Sports Med 2015; 43(8): 1902-7. doi: 10.1177/0363546515584041

9.  **Guyatt G H, Osoba D, Wu A W, Wyrwich K W, Norman G R.** Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002; 77(4): 371-83. doi: 10.4065/77.4.371

10. **Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson D T, et al.** Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007; 34(5): 1188-93.

11. **Goh G S, Liow M H L, Chen J Y, Tay D K-J, Lo N-N, Yeo S-J.** The patient acceptable symptom state for the Knee Society Score, Oxford Knee Score and Short Form-36 following unicompartmental knee arthroplasty. Knee Surgery, Sport Traumatol Arthrosc 2021. doi: 10.1007/s00167-021-06592-x

12. **Sabah S A, Alvand A, Beard D J, Price A J.** Minimal important changes and differences were estimated for Oxford hip and knee scores following primary and revision arthroplasty. J Clin Epidemiol 2021; 143: 159-68. doi: 10.1016/j.jclinepi.2021.12.016

13. **Ingelsrud L H, Terluin B, Gromov K, Price A, Beard D, Troelsen A.** Which Oxford Knee Score level represents a satisfactory symptom state after undergoing a total knee replacement? Acta Orthop 2021; 92(1): 85-90. doi: 10.1080/17453674.2020.1832304

14. **Hamilton T W, Pandit H G, Lombardi A V, Adams J B, Oosthuizen C R, Clavé A, et al.** Radiological Decision Aid to determine suitability for medial unicompartmental knee arthroplasty: development and preliminary validation. Bone Joint J 2016; 98-B(10 Supple B): 3-10. doi: 10.1302/0301-620X.98B10.BJJ-2016-0432.R1

15. **Harris K, Dawson J, Gibbons E, Lim C R, Beard D J, Fitzpatrick R, et al.** Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. Patient Relat Outcome Meas 2016; 7: 101-8. doi: 10.2147/PROM.S97774

16. **Ingelsrud L H, Roos E M, Terluin B, Gromov K, Husted H, Troelsen A.** Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. Acta Orthop 2018; 89(5): 541-7. doi: 10.1080/17453674.2018.1480739

17. **Terluin B, Eekhout I, Terwee C B, de Vet H C W.** Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol 2015; 68(12): 1388-96. doi: 10.1016/j.jclinepi.2015.03.015

18. **Terluin B, Eekhout I, Terwee C B.** The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol 2017; 83: 90-100. doi: 10.1016/j.jclinepi.2016.12.015

19. **Hanley J A.** Receiver operating characteristic (ROC) methodology: the state of the art. Crit Rev Diagn Imaging 1989; 29(3): 307-35.

20. **Terluin B, Roos E M, Terwee C B, Thorlund J B, Ingelsrud L H.** Assessing baseline dependency of anchor-based minimal important change (MIC): don't stratify on the baseline score! Qual Life Res 2021; 30(10): 2773-82. doi: 10.1007/s11136-021-02886-2

21. **Clement N D, MacDonald D, Simpson A H R W.** 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(8): 1933-9. doi: 10.1007/s00167-013-2776-5

22. **Beard D J, Harris K, Dawson J, Doll H, Murray D W, Carr A J, et al.** Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. J Clin Epidemiol 2015; 68(1): 73-9. doi: 10.1016/j.jclinepi.2014.08.009

23. **Arden N K, Kiran A, Judge A, Biant L C, Javaid M K, Murray D W, et al.** What is a good patient reported outcome after total hip replacement? Osteoarthr Cartil 2011; 19(2): 155-62. doi: 10.1016/j.joca.2010.10.004

24. **Most J, Hoelen T-C A, Spekenbrink-Spooren A, Schotanus M G M, Boonen B.** Defining clinically meaningful thresholds for patient-reported outcomes in knee arthroplasty. J Arthroplasty 2022; 37(5): 837-844.e3. doi: 10.1016/j.arth.2022.01.092

25. **Farrar J T, Young J P J, LaMoreaux L, Werth J L, Poole M R.** Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001; 94(2): 149-58. doi: 10.1016/S0304-3959(01)00349-9

26. **Odgaard A, Emmeluth C, Schrøder H, Østgaard S E, Madsen F, Troelsen A, et al.** Danish Knee Arthroplasty Register Annual Report; 2019.

# Supplementary data

Table 1. Anchor items used to determine minimal important change, patient acceptable symptom state, and treatment failure criteria

| Anchor questions |
| --- |
| Anchor response and classification of response options |

**Minimal important change (MIC)**
*How are your knee problems now compared with prior to your operation?*
    Importantly improved
        1: Better, an important improvement
        2: Somewhat better, but enough to be an important improvement
    Unchanged
        3: Very small improvement, not enough to be an important improvement
        4: About the same
        5: Very small deterioration, not enough to be an important deterioration
    Importantly deteriorated
        6: Somewhat worse, but enough to be an important deterioration
        7: Worse, an important deterioration

**Patient acceptable symptom state (PASS)**
*Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider that your current state is satisfactory?*
    1: Yes
    2: No

**Treatment failure (TF)**
*If you answered "no" to the previous question, would you consider your current state as being so unsatisfactory that you think the treatment has failed?*
    1: Yes
    2: No

Table 3. Patient demographics and preoperative characteristics. Values are count (%) unless otherwise specified

| Factor | Non-responders 3 months n = 92 | Responders 3 months n = 331 | p-value | Non-responders 12 months n = 139 | Responders 12 months n = 340 | p-value | Non-responders 24 months n = 103 | Responders 24 months n = 235 | p-value |
|---|---|---|---|---|---|---|---|---|---|
| Age [a] | 67 (59–74) | 69 (61–74) | 0.3 | 67 (58–74) | 68 (61–74) | 0.2 | 68 (60–74) | 68 (60–74) | 0.9 |
| Female sex | 41 (45) | 194 (59) | 0.01 | 72 (52) | 196 (58) | 0.3 | 53 (52) | 138 (59) | 0.3 |
| BMI [a] | 30 (26–34) | 29 (26–34) | 0.1 | 30 (26–35) | 29 (25–33) | 0.01 | 29 (26–34) | 28 (25–32) | 0.2 |
| ASA | | | | | | | | | |
| 1 | 3 (3) | 20 (6) | 0.03 | 12 (9) | 26 (8) | 0.4 | 7 (7) | 23 (10) | 0.1 |
| 2 | 62 (68) | 257 (78) | | 95 (68) | 255 (75) | | 71 (69) | 176 (75) | |
| 3 | 27 (29) | 53 (16) | | 32 (23) | 58 (17) | | 25 (24) | 36 (15) | |
| 4 | – | 1 (0) | | – | 1 (0) | | – | – | |
| KL grade | | | | | | | | | |
| 2 | 2 (2) | 3 (1) | 0.03 | 10 (7) | 10 (3) | 0.4 | 7 (7) | 13 (5) | 0.4 |
| 3 | 24 (26) | 79 (24) | | 40 (29) | 92 (27) | | 23 (22) | 68 (29) | |
| 4 | 66 (72) | 249 (75) | | 89 (64) | 238 (70) | | 73 (71) | 154 (66) | |
| OKS [a] | 21 (16–28) [b] | 23 (17–28) | 0.3 | 19 (15–24) [d] | 24 (19–28) | 0.01 | 20 (16–26) [g] | 24 (19–29) | 0.01 |
| EQ5D index [a] | 0.69 (0.57–0.77) [b] | 0.66 (0.59–0.72) | 0.5 | 0.66 (0.50–0.72) [e] | 0.72 (0.62–0.72) | 0.1 | 0.66 (0.34–0.72) [h] | 0.72 (0.63–0.72) | 0.09 |
| EQ5D VAS [a] | 70 (50–80) [c] | 70 (50–80) | 1 | 60 (39–80) [e] | 70 (50–80) [f] | 0.05 | 50 (41–79) [h] | 70 (51–80) [f] | 0.01 |

Abbreviations: see Table 1.
P-values calculated with Wilcoxon signed rank test for continuous variables and chi-square test for dichotomous variables.
[a] Numbers are median (0.025–0.975 quantile range).
[b–h] Missing data, [b] n = 58, [c] n = 59, [d] n = 79, [e] n = 78, [f] n = 1, [g] n = 58, [h] n = 57.

Table 6. Polyserial correlation coefficient for minimal important change (MIC) and point-biserial correlation coefficient for patient acceptable symptom state (PASS) and treatment failure (TF) anchor questions and OKS

| OKS follow-up | Anchor question and OKS correlation coefficient | | | | | | |
|---|---|---|---|---|---|---|---|
| | MIC | | | PASS | | TF | |
| | Preop | Change | Postop | Preop | Postop | Preop | Postop |
| 3-month | 0.05 | 0.51 | 0.58 | 0.12 | 0.55 | 0.08 | 0.33 |
| 12-month | 0.07 | 0.55 | 0.68 | 0.13 | 0.67 | 0.08 | 0.53 |
| 24-month | 0.07 | 0.61 | 0.76 | 0.09 | 0.67 | 0.04 | 0.59 |

Abbreviations: OKS: Oxford Knee Score; Preop: before operation; Postop: after operation.

Table 8. Mean difference in minimal important change (MIC), patient acceptable symptom state (PASS), and treatment failure (TF) across follow-up time-points after unicompartmental knee arthroplasty for the OKS. Values are mean difference obtained from adjusted predictive modeling. Values in parentheses are 95% confidence interval, calculated using 1,000-replication bootstrapping and reported as 0.025–0.975 quantiles

| OKS follow-up between months | MIC | PASS | TF |
|---|---|---|---|
| 3 and 12 | –2.3 (–4.3 to –0.2) | –3.8 (–5.6 to –2.0) | –5.1 (–8.9 to –1.4) |
| 12 and 24 | 1.6 (–1.0 to 4.2) | 1.4 (–1.0 to 3.9) | 0.8 (–2.1 to 3.8) |
| 3 and 24 | –0.7 (–3.1 to 1.9) | –2.4 (–4.6 to 0.1) | –4.3 (–8.4 to –0.3) |

OKS: Oxford Knee Score.

Table 9. Baseline dependency of minimal important change (MIC), patient acceptable symptom state (PASS), and treatment failure (TF) cut-off OKS values calculated with adjusted predictive modeling using a baseline dependency method. Values are adjusted predictive value with 95% confidence interval in parentheses calculated using 1,000-replication bootstrapping and reported as 0.025–0.975 quantiles

| OKS follow-up | Low baseline | High baseline | Difference |
|---|---|---|---|
| MIC | | | |
| 3-month | 7.2 (5.6–8.8) | 2.5 (0.3–4.1) | 4.7 (2.7–7.3) |
| 12-month | 9.1 (7.0–11.2) | 5.1 (2.6–6.8) | 4.0 (1.6–7.2) |
| 24-month | 8.0 (5.3–10.7) | 2.7 (-0.5–4.4) | 5.3 (2.8–9.6) |
| PASS | | | |
| 3-month | 26.6 (24.8–28.5) | 31.4 (29.5–33.0) | 4.7 (2.0–7.2) |
| 12-month | 30.0 (28.1–31.9) | 35.3 (33.4–36.7) | 5.4 (2.5–7.7) |
| 24-month | 29.0 (24.9–32.0) | 33.9 (31.2–35.6) | 5.0 (0.6–9.2) |
| TF | | | |
| 3-month | 21.5 (16.0–25.8) | 27.9 (21.5–31.9) | 6.4 (−0.9–13.5) |
| 12-month | 26.4 (23.1–29.1) | 31.8 (28.9–33.8) | 5.4 (1.3–9.1) |
| 24-month | 26.0 (21.7–28.8) | 31.8 (28.1–33.8) | 5.8 (1.0–10.2) |

OKS: Oxford Knee Score.

Table 10. Minimal important change (MIC), patient acceptable symptom state (PASS), and treatment failure (TF) OKS thresholds obtained from adjusted predictive modeling, unadjusted predictive modeling, and ROC statistics. Values in parentheses are 95% confidence interval, calculated using 1,000-replication bootstrapping and reported as 0.025–0.975 quantiles

| | Modeling approach | | |
| OKS follow-up | Adjusted predictive | Unadjusted predictive | ROC statistics |
|---|---|---|---|
| MIC | | | |
| 3-month | 4.7 (3.3–6.0) | 7.0 (5.7–8.1) | 11.5 (3.5–13.5) |
| 12-month | 7.1 (5.2–8.6) | 9.6 (8.0–11.0) | 10.5 (5.5–10.5) |
| 24-month | 5.4 (3.4–7.3) | 8.3 (6.4–9.9) | 7.5 (5.5–7.5) |
| PASS | | | |
| 3-month | 28.9 (27.6–30.3) | 30.8 (29.7–32.0) | 30.5 (27.5–36.5) |
| 12-month | 32.7 (31.5–33.9) | 34.8 (33.7–35.8) | 35.5 (34.5–36.5) |
| 24-month | 31.3 (29.1–33.3) | 33.7 (31.7–35.4) | 34.5 (29.0–34.5) |
| TF | | | |
| 3-month | 24.4 (20.7–27.4) | 27.8 (24.5–30.5) | 28.5 (18.5–35.5) |
| 12-month | 29.3 (27.3–31.1) | 32.3 (30.6–33.9) | 35.5 (29.5–35.5) |
| 24-month | 28.5 (26.0–30.5) | 31.5 (29.3–33.3) | 33.5 (27.5–33.5) |

OKS: Oxford Knee Score; ROC: receiver operating characteristic.