Predictors of clinical outcome in total hip and knee replacement : a methodological appraisal of implants and patient factors
Keurentjes, J.C.

Citation
Keurentjes, J. C. (2014, September 30). *Predictors of clinical outcome in total hip and knee replacement : a methodological appraisal of implants and patient factors*. Retrieved from https://hdl.handle.net/1887/28958

Cover Page

The handle http://hdl.handle.net/1887/28958 holds various files of this Leiden University dissertation

Author: Keurentjes, Johan Christiaan
Title: Predictors of clinical outcome in total hip and knee replacement : a methodological appraisal of implants and patient factors
Issue Date: 2014-09-30

Patient Acceptable Symptom States After Total Hip or Knee Replacement at Mid-term Follow-up

JC Keurentjes[1], FR Van Tol[1], M Fiocco[2], C So-Osman[3], R Onstenk[4], AW Koopman-Van Gemert[5], RG Pöll[6], RG Nelissen[1]

1 Orthopaedic Surgery, Leiden University Medical Center.

2 Medical Statistics and BioInformatics, Leiden University Medical Center.

3 Sanquin Blood Supply, South West Region.

4 Orthopaedic Surgery, Groene Hart Hospital, Gouda.

5 Anaesthesiology, Albert Schweitzer Hospital, Dordrecht.

6 Orthopaedic Surgery, Slotervaart Hospital, Amsterdam.

Abstract

Objectives To define Patient Acceptable Symptom State (PASS) thresholds for the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) at mid-term follow-up.

Methods In a prospective multicenter cohort study, OHS and OKS were collected at an average of 3 years follow-up, combined with a Numeric Rating Scale of Satisfaction (NRSS) and an external validation question (VQ): willingness to undergo surgery again.

Results 550 patients underwent THR and 367 underwent TKR. Receiver operating characteristic (ROC) curves identified a PASS threshold of 42 for the OHS after THR and 37 for the OKS after TKR. THR patients with an OHS$\geq$42 and TKR patients with an OKS$\geq$37 had a higher Numeric Rating Scale of Satisfaction and a larger odds of being willing to undergo surgery again.

Conclusions PASS thresholds appear larger at mid-term follow-up than at 6 months after surgery. Without external validation, we would advise against using these PASS thresholds as absolute thresholds in defining whether or not a patients has attained an acceptable symptom state after THR/TKR.

Introduction

Several distinct types of outcome measures are of interest in orthopaedic surgery. The time to a certain event, such as revision surgery, has historically been the principal outcome of interest in joint replacement patients.[116] In recent years, patient-reported outcome measures (PROMs) have become popular, allowing the assessment of the clinical outcome of joint replacement from the patient's perspective.[205] PROMs can be summarised in numerous ways. In the orthopaedic literature, mean scores of the study population are frequently presented. The mean pre-operative score provides information on the "average" patient before surgery. Similarly, the mean post-operative score provides information on the "average" patient after surgery, and the mean change in these scores provides information on the improvement (or deterioration) experienced by the "average" patient, who does improve substantially after joint replacement.[1] However, a large proportion of joint replacement patients suffer from persisting pain,

or are dissatisfied with the surgical results.[11, 15, 166] Data regarding the mean improvement after joint replacement mainly report the improvement of many patients with successful outcomes, but can neglect patients with suboptimal outcomes, making it of limited use for individual patients encountered in clinical practice.

Patient Acceptable Symptom States (PASS) and Minimal Clinically Important Differences (MCIDs) are two complementary constructs, which allow a more individualised approach to the analysis of PROMs.[193, 206, 207] PASS is defined as an outcome score threshold of the post-operative score, above which a patient is defined as experiencing a satisfactory outcome, and below which an unsatisfactory outcome is experienced. MCID is defined as the minimum amount of improvement between pre- and post-operative scores that a patient should experience after a specific intervention in order to have achieved a minimally important difference. PASSs and MCIDs allow estimation of the probability of a satisfactory outcome or a relevant improvement. These probabilities are relevant for individual patients, encountered in clinical practice, who either do or do not achieve an acceptable state or experience a relevant improvement.[191] Recently, PASSs have been estimated for the Oxford Hip Score (OHS)[208] and Oxford Knee Score (OKS)[209] at short-term follow-up.[210] An important issue is whether the chosen follow-up period of six months after joint replacement is adequate. A recent systematic review has suggested that patients may not have fully recovered at six months after THR.[149] Thresholds that define whether or not patients have achieved an acceptable symptom state, such as the PASS, may therefore differ between patients who are still recovering from their surgery and patients who have recovered fully. Therefore, we questioned whether PASS thresholds are different at mid-term compared with short-term follow-up. We questioned whether the OHS and OKS are correlated to patient satisfaction at mid-term follow-up. Additionally, we questioned whether responders (i.e., patients who have an acceptable symptom state according to the PASS) are more satisfied than non-responders. Finally, we questioned whether responders were more likely to be willing to undergo surgery again, compared with non-responders.

Materials and Methods

The current study is part of a multicentre cohort study of health-related quality of life (HRQoL) after THR/TKR (NTR2190), performed from August 2010 to August 2011.[21, 27, 28] Institutional Review Board approval was obtained from all participating centres and all patients gave written informed consent (CCMO-Nr: NL29018.058.09; MEC-Nr: P09.189). It concerned the clinical follow-up of a multi-centre randomised controlled clinical trial, comparing the use of the drug erythropoietin and two re-infusion techniques of autologous blood in order to decrease allogenic blood transfusions (Netherlands Trial Register: NTR303). In this trial, 2442 primary and revision hip or knee replacements in 2257 patients were included between 2004 and 2009. All patients who participated in the randomised controlled trial completed pre-operative HRQoL questionnaires, underwent primary THR or TKR and who were alive at the time of inclusion for the present follow-up study, were eligible for inclusion. The first joint replacement was selected for inclusion in the follow-up study for patients who participated more than once in the previous study. Records of the financial administration of all participating centres were checked in order to ascertain that all eligible patients were alive before being approached by the first author (JCK). For the present follow-up study, all eligible patients were first sent an invitation letter signed by their treating orthopaedic surgeon, an information brochure and a reply card. Patients who indicated that they were willing to participate were sent a questionnaire. Patients who did not respond to the first invitation within four weeks were sent another invitation letter. Those who did not respond to this second invitation were contacted by telephone by the first author. Patients who did not return their questionnaire within four weeks were also contacted by telephone by the first author. The data used in this report constitute a subset of patients who completed post-operative questionnaires.

Outcome measures: We measured the overall satisfaction with the outcome of surgery on a numeric rating scale (NRS), which ranged from 0 (extremely dissatisfied) to 10 (extremely satisfied). We added a validation question to the questionnaire, which took the following form: "Knowing what your hip or knee replacement surgery did for

you, would you still have undergone this surgery?", with dichotomous answers of 'yes' vs 'no'. This validation question was previously used in a similar validation study of clinically important differences after THR and TKR.[203]

Joint-specific PROMs were measured using the OHS for THR patients and the OKS for patients undergoing TKR, both of which were translated and validated in Dutch.[211, 212] Each questionnaire comprises 12 questions regarding pain and functioning of the hip or knee during the previous four weeks. Each question is answered on a five-point Likert scale, and an overall score is calculated by summarising the responses to each of the 12 questions. This sum score ranges from 0 to 48, where 0 indicates the most severe symptoms and 48 the least severe symptoms.

Potential confounders included age at joint replacement, gender, body mass index (BMI), indication for joint replacement (osteoarthritis (OA) vs other), patient-reported Charnley classification of comorbidity (A, patients in which the index operated hip or knee are affected only; B, patients in which the other hip or knee is affected as well; C, patients with a hip or knee replacement and other affected joints and/or a medical condition which affects the patients' ability to ambulate)[136, 137] and pre-operative HRQoL. HRQoL was measured pre-operatively using the Short-Form (SF-)36,[197] which is translated and validated in the Dutch language.[133] The 36 items cover eight domains (physical function, role physical, bodily pain, general health, vitality, social function, role emotional, and mental health), for which a sub-scale score is calculated (100 indicating no symptoms and 0 indicating extreme symptoms). These sub-scales can be summarised in a physical component summary (PCS) and a mental component summary (MCS).

Statistical Analysis: We performed descriptive analyses of the patients' baseline characteristics. All analyses were performed separately for THR and TKR, as MCIDs have been shown to differ considerably between these surgical interventions.[141] The correlation between OHS or OKS and NRS for satisfaction was calculated using Pearson's correlation coefficient. Receiver operating characteristic (ROC) curves were used to identify thresholds for OHS/OKS scores at mid-term follow-up, which are

associated with an acceptable level of patient satisfaction with joint replacement. An acceptable level of patient satisfaction was defined as a NRS for satisfaction ≥ 5, which is the equivalent of a visual analogue scale satisfaction score ≥ 50.[18, 205] This particular threshold has been used previously to compare satisfied and dissatisfied patients after joint replacement.[210] The chosen PASS thresholds were equivalent to the point at which sensitivity and specificity were closest.[213] The 95% confidence intervals (CI) around the thresholds were estimated using percentile bootstrap methods, based on 1000 random samples with replacement from the original data. In order to explore whether the found thresholds are consistent across subgroups, we identified separate thresholds for subgroups based on the following variables: length of follow-up (< 3 years vs ≥ 3 years), gender, age (< 70 years vs ≥ 70 years), BMI (< 30 kg/m2 vs ≥ 30 kg/m2), Charnley classification (A/B vs C), SF-36 PCS (< 50 vs ≥ 50) and SF-36 MCS (< 50 vs ≥ 50).

Based on the overall PASS thresholds, we divided patients into responders (those with an OHS or OKS $\geq$ the PASS threshold) and non-responders (OHS/OKS $<$ PASS threshold). We compared the mean NRS for satisfaction between responders and non-responders separately for THR and TKR patients, using three different models. In the first model, we calculated the mean NRS for satisfaction of all responders and the mean NRS of all non-responders, stratified by centre. In the second model, we performed linear mixed model regression analyses, with age and gender as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length. The final model consisted of linear mixed model regression analyses, with age, gender, BMI, Charnley classification, indication (OA vs other), and pre-operative SF-36 PCS and MCS as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length.

Finally, we compared the odds of responding the validation question positively between responders and non-responders, using three different models. In the first model, we calculated crude odds ratios. In the second model, we performed logistic mixed model regression analyses, with age and gender as fixed effects and the centre as a random

effect, while stratifying for quartile of follow-up length. In the final model, we performed logistic mixed model regression analyses, with age, gender, BMI, Charnley classification, indication (OA vs other), and pre-operative SF-36 PCS and MCS as fixed effects and the centre as a random effect, while stratifying for quartile of follow-up length.

All analyses were performed using R v2.14.1.[43]

Results

A total of 550 patients underwent THR and 367 underwent TKR (see study flowchart in figure 1.1 (p. 5)). Patient characteristics are described in Table 8.1 on the following page. The mean follow-up was similar for THR and TKR patients, at 3.2 years (1.5 to 6.0) and 3.2 years (1.3 to 6.0), respectively. THR patients were slightly younger at joint replacement surgery than TKR patients. The proportion of males was similar. TKR were more often obese or morbidly obese. The majority of THR and TKR patients underwent joint replacement for primary OA.

 The mean and median OHS scores at mid-term follow-up were 41.5 (sd 7.93) and 44 (interquartile range (IQR) 39 to 47), respectively. The mean NRS for satisfaction was 8.55 (sd 2.19) and 94.7% (521 of 550) of all THR patients were satisfied (defined as NRS $\geq$5). The mean and median OKS scores at mid-term follow-up were 39.1 (sd 9.04) and 42 (IQR 35 to 46), respectively. The mean NRS for satisfaction was 8.07 (sd 2.61) and 90.7% (333 of 367) of all TKR patients were satisfied.

The NRS correlated with both the OHS (Pearson correlation coefficient 0.52 (95% CI 0.46 to 0.58)) and the OKS (Pearson correlation coefficient 0.64 (95% CI 0.57 to 0.69)), indicating a strong correlation.

 ROC curves of OHS thresholds and OKS thresholds are shown in Figures 8.1 (p. 113) and 8.2 (p. 114). The OHS ROC curve revealed a PASS threshold of 42, with a sensitivity of 67.0% and a specificity of 65.5%. The area under the ROC curve (AUC) was 0.72 (95% CI 0.60 to 0.84). The OKS ROC curve revealed a PASS threshold of 37, with a sensitivity of 76.3% and a specificity of 76.5%. The AUC was 0.83 (95% CI 0.74 to 0.93). ROC curves of subgroups showed variation in the thresholds found (Tables II and

	Primary THR: n = 550	Primary TKR: n = 367
Mean Follow-up Years (SD); Range	3.2 (1.1); 1.5 – 6.0	3.2 (1.1); 1.3 – 6.0
Mean Age at Joint Replacement (SD)	65.9 (10.5)	68.7 (9.64)
% Men:	34,2	33,3
BMI*		
% $<$25:	34,3	17,9
% 25-30:	42,6	44,5
% 30-35:	17,7	23,4
% $>$35:	5,82	14,2
Indication for Joint Replacement		
% Osteoarthritis:	86,3	89
Patient-reported Charnley Classification*		
% A:	23,1	13,9
% B:	14,3	10,2
% C:	62,6	75,9
Median Preoperative SF36 Summary Scores (IQR)		
Physical Component Summary	39.8 (34.1 – 45.3)	41.3 (35.0 – 47.3)
Mental Component Summary	54.8 (45.6 – 60.0)	54.1 (45.4 – 59.1)

Table 8.1: **Patient Characteristics.** *Measured at follow-up.

III). The variation appears larger in OHS thresholds than in OKS thresholds.

The mean NRS for satisfaction was significantly higher in responders than in non-responders, both for THR and TKR (Table 8.4 (p. 116)). Both models showed a mean difference between responders and non-responders of approximately two points for THR patients and three points for TKR patients. Responders were more likely to be willing to undergo surgery again (Table 8.5 (p. 117)). All models showed odds ratios of approximately 7, indicating a seven-times higher odds of willingness to undergo surgery again in responders versus non-responders, while controlling for confounding (Table 8.5 (p. 117)).
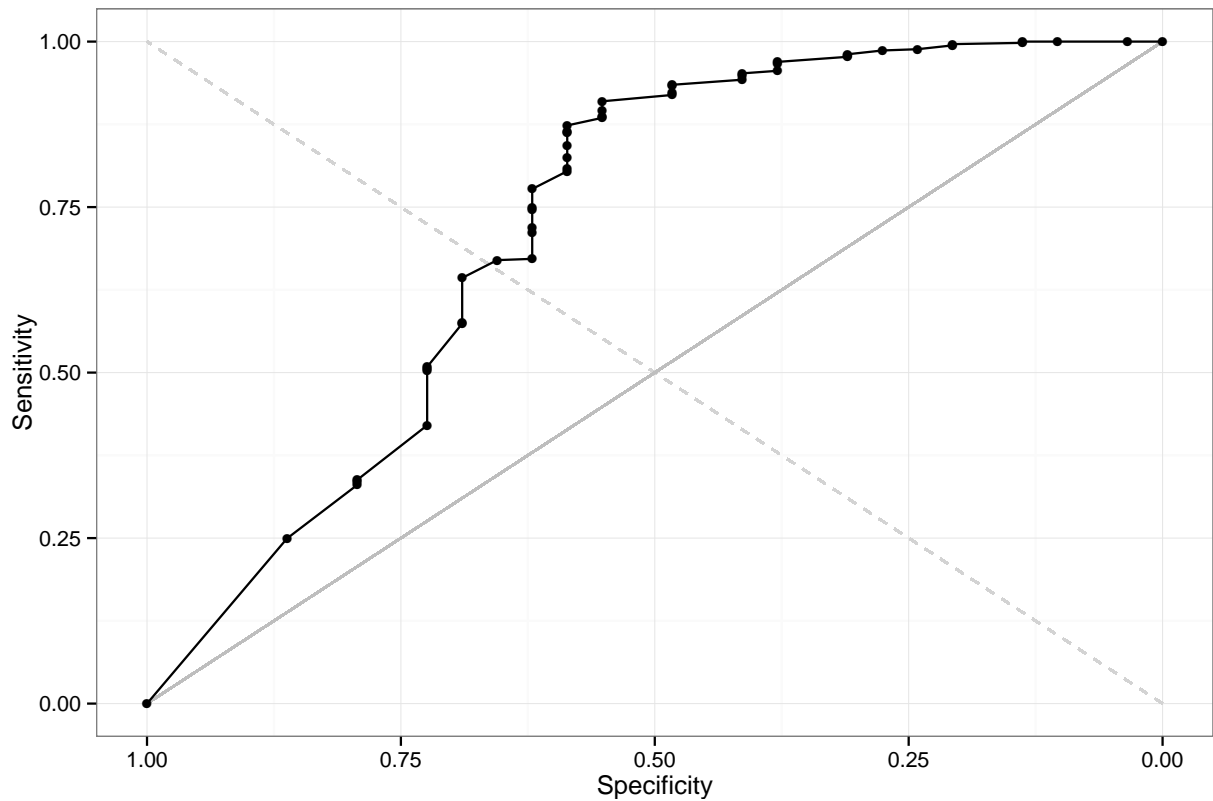
Figure 8.1: **Receiver Operating Characteristic (ROC) curves to identify thresholds for the mid-term follow-up Oxford Hip Score (OHS) associated with mid-term satisfaction with surgery.** Area under ROC curve (AUC): 0.72 (95%-CI: 0.60–0.84). Sensitivity: 67.0%, Specificity: 65.5%.

Discussion

PASS thresholds for the OHS and OKS are considerably higher at mid-term follow-up than those at six months post-operatively. The multiple approaches in validating the PASS thresholds and the rigorous efforts to minimise confounding are the main strengths of this study. All approaches show that the thresholds of 42 points for the OHS and 37 points for the OKS can discriminate between successful and less successful patient outcomes after THR or TKR in this study population, according to the overall satisfaction assessment and the willingness to undergo surgery again.

A limitation of our study is that we did not measure the OHS or OKS pre-operatively. Consequently, we could not investigate whether the PASS thresholds are valid across strata of baseline OHS or OKS scores. As a surrogate measurement of pre-operative joint functioning, we investigated differences in PASS thresholds in strata of the pre-
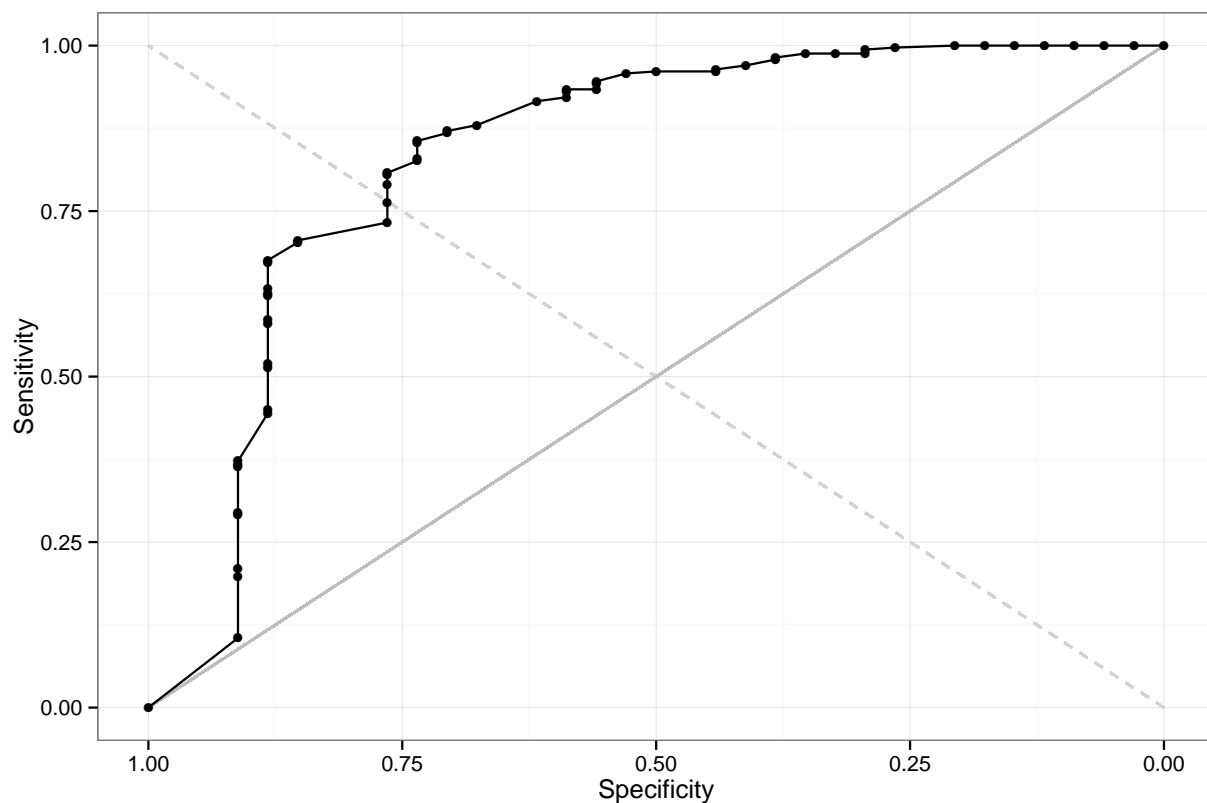
Figure 8.2: **Receiver Operating Characteristic (ROC) curves to identify thresholds for the mid-term follow-up Oxford Knee Score (OKS) associated with mid-term satisfaction with surgery.** Area under ROC curve (AUC): 0.83 (95%-CI: 0.74–0.93). Sensitivity: 76.3%, Specificity: 76.5%.

operative physical and mental component summaries of the SF-36, which only had a small effect. Furthermore, other evidence from the same research group suggests that baseline OHS or OKS values are poor predictors of overall patient satisfaction with the outcome of the joint replacement.[210, 214]

Another limitation of our study is the broad range in follow-up length. In order to account for this range, we stratified our analysis per quartile of follow-up period. Although a residual effect of follow-up length cannot be excluded, we do not think this is very plausible, as recent evidence suggests that after full recovery has taken place, the improvement in joint function is sustained throughout mid-term follow-up.[148]

Demographically, our study population is similar to that of Judge et al.[210] Cultural differences cannot be excluded from explaining the differences found in PASS thresholds, although this is unlikely, given the resemblance of English and Dutch urban joint

	Thresholds (95% CI)	Satisfactory Symptom State: n (%)
Entire Population	42 (38.5 – 44.2)	359 (65.3)
Follow-up < 3 yrs	45 (39.6 – 46.5)	134 (47.2)
Follow-up ≥ 3 yrs	39 (30.5 – 42.5)	184 (69.2)
Males	47 (44.2 – 47.5)	85 (45.2)
Females	38 (30.5 – 41.2)	258 (71.3)
Age < 70 yrs	44 (39.1 – 45.5)	204 (60.0)
Age ≥ 70 yrs	40 (30.5 – 42.5)	127 (60.5)
BMI < 30	43 (39.6 – 45.5)	280 (68.6)
BMI ≥ 30	36 (25.5 – 43.5)	55 (45.1)
Charnley Class A / B	45 (26.0 – 47.5)	150 (77.7)
Charnley Class C	41 (37.0 – 43.5)	177 (54.8)
PCS < 50	43 (37.5 – 44.7)	287 (64.8)
PCS ≥ 50	47 (34.0 – 48.0)	94 (90.0)
MCS < 50	38 (28.7 – 41.5)	94 (54.0)
MCS ≥ 50	45 (40.7 – 46.5)	156 (48.3)

Table 8.2: **Patient acceptable symptom state (PASS) score thresholds (95% Confidence Intervals) for Mid-term follow-up Oxford Hip Score and percentage of patients classified as reaching a satisfactory symptom state, per relevant subgroups.**

replacement patients. A more plausible explanation could be the difference in physical recovery. Patients who are fully recovered at mid-term follow-up could be less prone to be satisfied with lower OHS or OKS scores, as the probability of further improvement in physical functioning is small. Six months after joint replacement, patients might be more readily satisfied with suboptimal OHS or OKS scores, as the speed of recovery is quite high.

Conceptually, MCIDs and PASSs are complementary. Both approach an individual patient's health state, but from a slightly different angle. In MCIDs, the emphasis is on whether or not an individual has improved after a certain therapy.[215] In PASSs, the emphasis is on whether or not the achieved outcome is acceptable from the patients perspective.[215] Both MCIDs and PASS have gained in interest recently.[141, 216–218] PASS might be more important than the MCID, as Dougados[219] phrased eloquently: "It's good to feel better but it's better to feel good." Similar methodological difficulties are encountered both in MCIDs and PASSs: both approaches lead to a loss of power

	Thresholds (95% CI)	Satisfactory Symptom State: n (%)
Entire Population	37 (31.8 – 38.5)	271 (73.8)
Follow-up < 3 yrs	34 (29.7 – 37.5)	144 (78.7)
Follow-up ≥ 3 yrs	38 (32.5 – 41.5)	131 (74.0)
Males	39 (33.5 – 44.4)	89 (73.6)
Females	33 (30.5 – 37.5)	171 (70.7)
Age < 70 yrs	34 (29.7 – 38.1)	139 (77.7)
Age ≥ 70 yrs	38 (31.3 – 41.7)	140 (77.3)
BMI < 30	38 (32.0 – 39.5)	170 (78.7)
BMI ≥ 30	35 (29.7 – 40.5)	84 (64.6)
Charnley Class A / B	39 (30.8 – 45.5)	53 (62.4)
Charnley Class C	36 (30.5 – 38.5)	180 (67.4)
PCS < 50	35 (31.3 – 38.5)	172 (63.2)
PCS ≥ 50	39 (25.0 – 47.0)	23 (39.0)
MCS < 50	33 (21.5 – 38.5)	108 (86.4)
MCS ≥ 50	38 (33.5 – 42.3)	147 (71.4)

Table 8.3: **Patient acceptable symptom state (PASS) score thresholds (95% Confidence Intervals) for Mid-term follow-up Oxford Knee Score and percentage of patients classified as reaching a satisfactory symptom state, per relevant subgroups.**

	Responders:	Non-responders:	Mean Difference Adjusted for Age and Sex*	Mean Difference Adjusted for All Potential Confounders**
THR	9.28 (9.26 – 9.29)	7.26 (7.22 – 7.31)	1.89 (1.53 – 2.25)	1.87 (1.47 – 2.26)
TKR	9.04 (9.01 – 9.06)	6.78 (6.71 – 6.84)	2.89 (2.40 – 3.38)	2.96 (2.44 – 3.48)

Table 8.4: **Comparison of Mean Numerical Rating Scale of Satisfaction between responders and non-responders according to the Patient acceptable symptom state (PASS) score thresholds for THR and TKR.** * Stratified by quartiles of follow-up length. ** Adjusted for Age, Sex, BMI, Charnley Classification, OA vs other indications for joint replacement, preoperative Physical Component Summary Scale and preoperative Mental Component Summary Scale and stratified by quartiles of follow-up length.

compared with the population-level mean difference, both approaches depend on population and contextual characteristics and there is no clear consensus on the optimal statistical approach.[167, 193, 206] Despite these difficulties, MCIDs and PASSs are the best tools available to analyse PROM data at the individual level.

In this study, we estimated PASS in OHS/OKS at mid-term follow-up in a Dutch population. We found evidence suggesting that PASSs are time-dependent. Besides

	Crude Odds Ratio	Odds Ratio Adjusted for Age and Sex*	Odds Ratio Adjusted for All Potential Confounders**
THR	7.64 (4.09 – 15.3)	6.97 (3.51 – 13.8)	8.53 (3.80 – 19.1)
TKR	7.28 (3.85 – 14.3)	7.92 (3.79 – 16.6)	7.73 (2.84 – 21.0)

Table 8.5: **Comparison of Willingness to Undergo Surgery Again between responders and non-responders according to the Patient acceptable symptom state (PASS) score thresholds for THR and TKR.** An Odds Ratio >1 indicates a larger odds of Willingness to Undergo Surgery Again in Responders than in Non-responders. * Stratified by quartiles of follow-up length. ** Adjusted for Age, Sex, BMI, Charnley Classification, OA vs other indications for joint replacement, preoperative Physical Component Summary Scale and preoperative Mental Component Summary Scale and stratified by quartiles of follow-up length.

being time-dependent, PASS might also be population-dependent, as different sub-groups had different PASS thresholds. Without any form of external validation at a similar follow-up period, we would advise against using these PASS thresholds as absolute thresholds in defining whether or not a patient has attained an acceptable symptom state after THR/TKR.