■ KNEE

Evidence for the validity of a patient-based instrument for assessment of outcome after revision knee arthroplasty

EVALUATION OF THE OXFORD KNEE SCORE USING A UK NATIONAL PROMS DATASET

S. A. Sabah,
A. Alvand,
D. J. Beard,
A. J. Price

From University of Oxford, Oxford, UK

Aims

To estimate the measurement properties for the Oxford Knee Score (OKS) in patients undergoing revision knee arthroplasty (responsiveness, minimal detectable change (MDC-90), minimal important change (MIC), minimal important difference (MID), internal consistency, construct validity, and interpretability).

Methods

Secondary data analysis was performed for 10,727 patients undergoing revision knee arthroplasty between 2013 to 2019 using a UK national patient-reported outcome measure (PROM) dataset. Outcome data were collected before revision and at six months postoperatively, using the OKS and EuroQol five-dimension score (EQ-5D). Measurement properties were assessed according to COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) guidelines.

Results

A total of 9,219 patients had complete outcome data. Mean preoperative OKS was 16.7 points (SD 8.1), mean postoperative OKS 29.1 (SD 11.4), and mean change in OKS + 12.5 (SD 10.7). Median preoperative EQ-5D index was 0.260 (interquartile range (IQR) 0.055 to 0.691), median postoperative EQ-5D index 0.691 (IQR 0.516 to 0.796), and median change in EQ-5D index + 0.240 (IQR 0.000 to 0.567). Internal consistency was good with Cronbach's α 0.88 (baseline) and 0.94 (post-revision). Construct validity found a high correlation of OKS total score with EQ-5D index (r = 0.76 (baseline), r = 0.83 (post-revision), $p < 0.001$). The OKS was responsive with standardized effect size (SES) 1.54 (95% confidence interval (CI) 1.51 to 1.57), compared to SES 0.83 (0.81 to 0.86) for the EQ-5D index. The MIC for the OKS was 7.5 points (95% CI 5.5 to 8.5) based on the optimal cut-off with specificity 0.72, sensitivity 0.60, and area under the curve 0.66. The MID for the OKS was 5.2 points. The MDC-90 was 3.9 points. The OKS did not demonstrate significant floor or ceiling effects.

Conclusion

This study found that the OKS was a useful and valid instrument for assessment of outcome following revision knee arthroplasty. The OKS was responsive to change and demonstrated good measurement properties.

Cite this article: *Bone Joint J* 2021;103-B(4):627–634.

Correspondence should be sent to S. A. Sabah; email: shiraz.sabah@ndorms.ox.ac.uk

doi:10.1302/0301-620X.103B4.BJJ-2020-1560.R1 $2.00

Bone Joint J 2021;103-B(4):627–634.

Introduction

Since 2009, patient-reported outcome measures (PROMs) have been routinely collected in England for patients undergoing knee arthroplasty surgery.[1] PROMs are collected before surgery and at six months using a joint-specific score (Oxford Knee Score (OKS)) and a health-related quality of life measure (EuroQol five-dimension score (EQ-5D)).[2,3] Data are collected on both primary and revision procedures[4,5] and have been widely used to report the clinical outcome of knee arthroplasty,[6,7] as a way of comparing across healthcare providers[8] and more recently to feed back performance to individual surgeons.[9] In addition,

Table I. Hypotheses for construct validity and responsiveness of the Oxford Knee Score in revision knee arthroplasty.

Hypothesis		Preoperative	Postoperative	Change
1	The correlation between OKS and EQ-5D for the entire cohort is at least moderate ($r \geq 0.50$)	Accepted	Accepted	Accepted
2	The correlation between OKS and EQ-5D mobility is at least moderate ($r \geq 0.50$).	Rejected	Accepted	Rejected
3	The correlation between OKS and EQ-5D self-care was at least moderate ($r \geq 0.50$).	Accepted	Accepted	Rejected
4	The correlation between OKS and EQ-5D usual activities was at least moderate ($r \geq 0.50$).	Rejected	Accepted	Rejected
5	The correlation between OKS and EQ-5D pain/discomfort was at least moderate ($r \geq 0.50$).	Accepted	Accepted	Accepted
6	The correlation between OKS Pain Component Score and EQ-5D pain/discomfort is at least 0.05 greater than with EQ-5D mobility.	Accepted	Accepted	Accepted
7	The correlation between OKS Function Component Score and EQ-5D mobility is at least 0.05 greater than with EQ-5D pain/discomfort.	Rejected	Rejected	Rejected
8	The correlation between OKS and EQ-5D anxiety/depression was low ($r \leq 0.49$).	Accepted	Accepted	Accepted
Summary		63% accepted	88% accepted	50% accepted

EQ-5D, EuroQol five-dimension score; OKS, Oxford Knee Score.

PROMs data are now being used at the individual patient level to assist in shared decision-making.[10,11]

The OKS was developed in 1998 from interviews with patients undergoing primary knee arthroplasty.[2,12] A recent systematic review demonstrated it had good measurement properties and a robust evidence base for use in this population.[13] However, its measurement properties for revision arthroplasty are less well evaluated and the score was not developed using a revision knee arthroplasty cohort. As such, it is essential to understand the responsiveness of the OKS in the revision knee arthroplasty population. This has been performed for the Oxford Hip Score (OHS),[14] demonstrating that it is responsive to change in functional outcome following revision hip arthroplasty.[15] This information is important to support the expanded use of currently collected PROMs to deliver patient-centred care in the revision knee surgery pathway.

The main aim of this study was to examine the response to change of the OKS in patients undergoing revision knee arthroplasty. In addition, the study aimed to estimate meaningful changes and other measurement properties for the OKS in this patient group.

Methods

This was a retrospective cohort study using freely available, de-identified patient data where ethical approval is not required. This study followed guidelines from the COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) group for evaluation of internal consistency, construct validity, responsiveness, and interpretability for the OKS in revision knee arthroplasty.[16] Supporting Evaluation, Analysis and Reporting of routinely Collected Healthcare Data (SEARCHeD) guidelines were followed to report this study.[17]

Study dataset. Patient data were obtained from the UK NHS national PROMs programme from 1st April 2013 to 31st March 2019. The PROMs datasets include patient-, provider-, clinical commissioning group-, and national-level data for England.[1] Revision knee arthroplasties were identified from the 'Revision Flag' field in the PROM dataset.[18] This flag is raised if a PROM is matched to a revision procedure episode in Hospital Episode Statistics (HES).[18] A list of Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS-4) codes for eligible procedures is provided in Supplementary Table i.[18] This definition encompasses any subsequent procedure to add, remove, or modify a joint arthroplasty, and represents those procedures that contribute to publicly available provider reports from NHS Digital on PROMs performance in revision knee arthroplasty.[1] Patients who withheld informed consent, were unable to complete PROMs questionnaires (with or without assistance), and those who subsequently opted out were excluded. Patients undergoing emergency procedures are not included in the NHS PROMs programme.

Outcome measures. The OKS is a knee-specific PROM that asks patients about their pain and function over the past four weeks.[19] It has 12 Likert items, each with five possible responses (0 to 4), to give a total score from 0 (worst) to 48 (best).[12,20] Health-related quality of life is measured using the EuroQol five-dimension three-level score (EQ-5D-3L).[3] This consists of five domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each with three ordinal levels from 1 (no problems) to 3 (extreme problems). The scale is converted to an index ranging from -0.59 (worst) to 1 (best health) using utility values for the UK population. General health is also measured using the EQ "thermometer", a visual analogue scale (VAS) from 0 (worst) to 100 (best).

Patients complete a preoperative questionnaire (Q1), which includes these instruments together with additional questions on the need for assistance completing the questionnaire, comorbidities, and previous arthroplasty for the affected joint. The mode of questionnaire administration varies between centres, with paper-based and electronic systems in use. A follow-up questionnaire (Q2) is administered at six months following surgery. This contains additional questions on patient satisfaction ("How would you describe the results of your operation? Response: Excellent, Very good, Good, Fair, Poor") and perceived success of surgery ("Overall, how are the problems now in the knee on which you had surgery, compared to before your operation? Response: Much better, A little better, About the same, A little worse, Much worse"). Further information on the dataset and PROMs methodologies is available from NHS Digital.[1,21]

Statistical analysis. The study population was described using frequencies, means and standard deviations (SDs), or medians and interquartile ranges (IQRs), as appropriate. A p-value < 0.05 was taken to indicate statistical significance. Statistical analyses were performed using Stata statistical software release 16 (StataCorp, USA) and R version 3.6.2 (R Foundation for Statistical Computing, Austria).

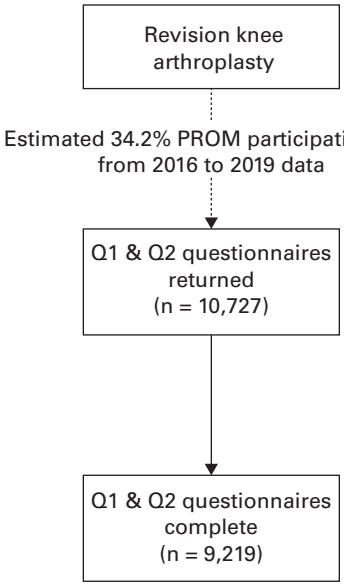

Fig. 1

Data flow diagram. PROM, patient-reported outcome measure.

Study size. A total of 10,727 revision knee arthroplasty procedures with returned pre- and postoperative questionnaires were available for analysis from the national PROMs programme.

PROM participation rates. Data on PROM participation for revision knee arthroplasty were available for 2016 to 2019 only. Over this period, 14,929 revision knee arthroplasty procedures were coded on HES, 7,432 patients (49.8%) completed a preoperative questionnaire, and 5,112 (34.2%) patients returned a postoperative questionnaire.

Procedures for handling missing data. Missing data were described for each variable in the dataset. The proportion of returned pre- and postoperative questionnaires with incomplete or erroneous outcome data was reported. Complete case analysis was used throughout, with no imputation for missing values, because the missingness mechanism was unknown. A complete outcome set was defined as a valid response to all items on the pre- and postoperative OKS and EQ-5D instruments, as well as the postoperative success transition question (which was required for calculation of the minimal important change (MIC).

Evaluation of internal structure. The internal structure of a PROM refers to how items within an instrument relate to one another. This information is important to determine how items should be combined into a scale. Internal consistency is a measure of the extent to which the items on a scale measure the same concept and was assessed using Cronbach's α. The effect of removing each item was also examined. We specified a Cronbach α of ≥ 0.7 as acceptable.[16]

Construct validity. Validity refers to the ability for an instrument to measure the construct it is intended to measure. Since there is no reference standard instrument for use in this context, a criterion approach could not be used. Instead we calculated construct validity, which is the extent to which a measure relates to other measures, consistent with a priori hypotheses. The construct validity of the OKS was estimated by comparison with the EQ-5D index through calculation of Spearman's correlation coefficients. We formulated hypotheses about the size and direction of relationships between OKS and the EQ-5D (Table I). Our main hypothesis concerned the total score for the OKS having at least moderate correlation with the EQ-5D at each timepoint. We formulated six further a priori hypotheses for convergent construct validity (Hypotheses 2 to 7) and one for discriminant construct validity (Hypothesis 8). Each hypothesis was tested against pre- and postoperative scores. The strength of r was interpreted according to thresholds from Munro:[22] little, if any (≤ 0.25), low (0.26 to 0.49), moderate (0.50 to 0.69), high (0.70 to 0.89), and very high (0.90 to 1.00). The proportion of accepted hypotheses was used to define quality of construct validity: $\geq 75\%$ accepted (good), 50% to 74% (moderate), and $< 50\%$ (low).

Responsiveness. Responsiveness is the ability for a PROM to detect change over time in the construct to be measured.[16] Pre- and postoperative scores were compared using paired t-tests. We also calculated and compared the standardized effect size (SES, which is the mean change score divided by the SD of the measure at baseline) and standardized response mean (SRM, which is the mean change score, divided by the SD of the change scores for the outcome measure used).[23] To evaluate construct responsiveness, we evaluated the same hypotheses as for construct validity, but this time applied to the change scores for the OKS and EQ-5D.

Meaningful changes in the OKS. The clinically relevant improvement for the average individual over a period of time was estimated by calculating the MIC using receiver operator curve (ROC) analysis.[24] The external anchor was the perceived success transition item from the Q2 Questionnaire.[25] Patients who responded to this question to indicate a minimal improvement ('a little better'), were compared with those who responded no improvement ('about the same'). The MIC was also calculated for a single group of patients as the mean change in OKS for patients who rated 'a little better'. The minimally important difference (MID) was calculated as the difference in mean change of OKS in patients 'a little better' versus 'about the same'. The minimal detectable change (MDC-90) was calculated to represent the smallest change beyond the measurement error of the instrument. We used the standard error of the mean ($SD \times \sqrt{(1 - \text{reliability})}$), multiplied by $\sqrt{2}$ (to account for measurement on two occasions) multiplied by a z score of 1.65 (to indicate a 90% confidence level).

Floor and ceiling effects. The effect of an intervention might be missed for patients who occupy the minimum or maximum scores. We defined floor and ceiling effects as being present if more than 15% of respondents achieved these scores.

Results

Study population. A total of 10,727 revision knee arthroplasty procedures (n = 5,320 female (49.6%), n = 4,760 male (44.3%), and n = 647 sex not specified (6.0%)) were available for analysis. The most frequent age category was 70 to 79 years (n = 3,997 (37.3%)).

Missing outcome data. For returned questionnaires, 9,219 patients had complete pre- and postoperative outcome data. This is illustrated in Figure 1. The baseline characteristics for

Table II. Summary of patient demographics and outcome measures.

Variable	Revision knee arthroplasty	
	Incomplete outcome measurements (n = 1,508)	**Complete outcome measurements (n = 9,219)**
Age band, yrs; n (%)		
40 to 49	2 (0.1)	17 (0.2)
50 to 59	151 (10.0)	1,037 (11.2)
60 to 69	431 (28.6)	3,115 (33.8)
70 to 79	602 (39.9)	3,395 (36.8)
80 to 89	243 (16.1)	1,087 (11.8)
Missing/suppressed	79 (5.2)	568 (6.2)
Sex, n (%)		
Male	618 (41.0)	4,142 (44.9)
Female	811 (53.8)	4,509 (48.9)
Not specified	79 (5.2)	568 (6.2)
Mean Q1 Oxford Score (SD)	15.6 (8.1)	16.7 (8.1)
n (%)	1,327 (88.0)	N/A
Mean Q2 Oxford Score (SD)	27.7 (11.5)	29.1 (11.4)
n (%)	1,276 (84.6)	N/A
Mean change in Oxford Score (SD)	12.0 (10.8)	12.5 (10.7)
n (%)	1,110 (73.6)	N/A
Median Q1 EQ-5D index (IQR)	0.195 (-0.003 to 0.620)	0.260 (0.055 to 0.691)
n (%)	820 (54.4)	N/A
Median Q2 EQ-5D index (IQR)	0.691 (0.516 to 0.796)	0.691 (0.516 to 0.796)
n (%)	949 (62.9)	N/A
Median change in EQ-5D index (IQR)	0.105 (0.000 to 0.367)	0.240 (0.000 to 0.567)
n (%)	330 (21.9)	N/A
Mean Q1 EQ-VAS (SD)	62.2 (22.9)	63.8 (22.0)
n (%)	1,042 (69.1)	8,569 (92.9)
Mean Q2 EQ-VAS (SD)	64.7 (22.5)	67.3 (21.0)
n (%)	1,251 (83.0)	8,903 (96.6)
Mean change in EQ-VAS (SD)	1.7 (25.9)	3.4 (23.4)
n (%)	837 (55.5)	8,306 (90.1)
Success transition, n (%)		
Much better	711 (47.1)	4,963 (53.8)
A little better	282 (18.7)	1,991 (21.6)
About the same	141 (9.4)	931 (10.1)
A little worse	97 (6.4)	704 (7.6)
Much worse	100 (6.6)	630 (6.8)
Missing	177 (11.7)	0 (0.0)
Patient satisfaction, n (%)		
Excellent	215 (14.3)	1,498 (16.2)
Very good	363 (24.1)	2,556 (27.7)
Good	369 (24.5)	2,476 (26.9)
Fair	283 (18.8)	1,746 (18.9)
Poor	142 (9.4)	896 (9.7)
Missing	136 (9.0)	47 (0.5)

Complete outcome data was defined as patients with fully complete pre- and postoperative Oxford Knee Score, EQ-5D index, and Postoperative Success Transition question. A complete EQ-VAS and Postoperative Satisfaction question were not required for our definition as these were not required to test our hypotheses or calculate the minimally important change. For the incomplete outcome data population, summary measures are provided for fully completed questionnaires within the group.
EQ-5D, EuroQol five-dimension score; SD, standard deviation; VAS, visual analogue scale.

populations with incomplete and complete sets of outcome measures are shown in Table II.

Outcome scores. For patients with complete outcome measurements, preoperative OKS approximated a normal distribution, with a subtle right-sided tail. Postoperative OKS had a left-sided tail (Figure 2). Preoperative OKS was mean 16.7 points (SD 8.1), postoperative OKS 29.1 (SD 11.4), and change in OKS + 12.5 (SD 10.7). The EQ-5D index had a bimodal distribution (Figure 3). Preoperative EQ-5D index was median 0.260 (IQR 0.055 to 0.691), postoperative EQ-5D index 0.691 (IQR 0.516 to 0.796), and change in EQ-5D index + 0.240 (IQR 0.000 to 0.567); EQ-VAS improvement was mean 3.4% (SD 23.4%). In all, 6,955 patients (75.4%) reported improvement on the postoperative success transition question, and 6,530 patients (70.8%) reported postoperative satisfaction of 'Good', 'Very good', or 'Excellent'.

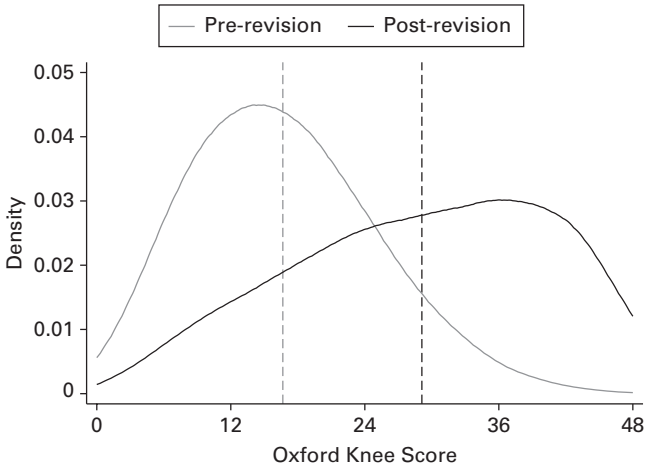

Fig. 2

Kernel Density Plot of Oxford Knee Score (OKS) prior to revision and at six months. Dashed lines represent mean OKS for each group.

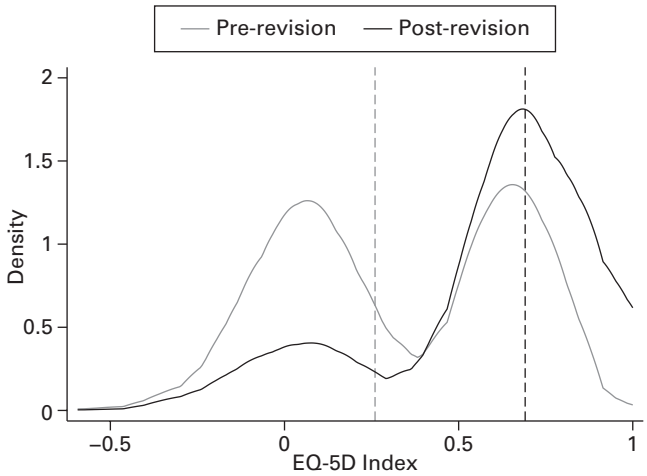

Fig. 3

Kernel Density Plot of EuroQol five-dimension (EQ-5D) index prior to revision and at six months. Dashed lines represent median EQ-5D index for each group.

Evaluation of internal structure. A high level of internal consistency was found for the OKS with Cronbach's α values of 0.88 and 0.94 pre- and post-revision. α values were all very slightly reduced with singular removal of items (observed at three or more decimal places), except for Item 7, 'Difficulty with kneeling', for the post-revision measurement only (Table III).

Construct validity. Pre- and post-revision OKS total score correlated highly (r 0.76, r 0.83, both $p < 0.001$, Spearman's rho) with EQ-5D index, consistent with our prespecified main hypothesis (Table I). For post-revision scores, seven out of eight hypotheses (88%) were accepted for the OKS and its subscales, which COSMIN rate as good.[16] For preoperative scores, five out of eight hypotheses (63%) were accepted, which is rated moderate. A correlation matrix is provided in Supplementary Table ii to support these judgements.

Responsiveness. Postoperative OKS and EQ-5D scores were each significantly improved compared to pre-operative scores (both $p < 0.001$, paired *t*-test). The OKS measured large mean effect sizes, all of which were greater than corresponding effect sizes for the EQ-5D index. The SES of the OKS was 1.54 (95% CI 1.51 to 1.57) and SRM 1.16 (1.14 to 1.19). This compared to SES 0.83 (0.81 to 0.86) and SRM 0.76 (0.74 to 0.79) for the EQ-5D index. The change in OKS had moderate correlation with the change in EQ-5D index (r 0.63, $p < 0.001$). Overall, exactly half of the hypotheses for responsiveness were accepted, which is rated moderate.

Meaningful changes in the OKS. The MIC for an individual patient was calculated to be a change in OKS of 7.5 points (95% CI 5.5 to 8.5) based on the optimal cut-off value with specificity 0.72 and sensitivity 0.60, with area under the ROC curve 0.66 (Figure 4). The MIC for a single group of patients (i.e. the mean change in Oxford Knee Score (COKS) for patients who were 'a little better') was calculated to be 9.5 points (SD 7.4). The mean COKS for patients who were 'about the same' was 4.3 (SD 6.7) points, which indicates a minimal important difference (MID) of 5.2 points. Patients who reported that they were 'a little worse' were actually a mean of 2.1 points better. The MDC-90 was 3.9 points.

Floor and ceiling effects. The OKS did not demonstrate significant floor or ceiling effects. The percentages of patients achieving the minimum score were 0.15% (n = 14/9,219) and 0.02% (n = 2/9,219) pre- and post-revision. The percentages achieving the maximum score were 0.01% (n = 1/9,219) and 1.59% (n = 147/9,219), respectively. Floor and ceiling effects are tabulated per item in Table III.

Discussion

This study demonstrated that the OKS was a responsive instrument for assessment of functional outcome in elective revision knee arthroplasty. At six months postoperatively, the OKS was demonstrably more responsive than the EQ-5D index. The anchor-based MIC was found to be 7.5 points, the MID 5.2 points, and the MDC-90 3.9 points. This study provides support for ongoing use of the OKS for revision knee arthroplasty in the national PROMs programme. Our work reinforces the paradigm that extension of a PROM beyond its development population should be accompanied by reassessment of its measurement properties to ensure that it is valid for use in the new population.

The OKS has been extensively studied in the setting of primary knee arthroplasty and has good evidence to support its use.[13] Outside joint arthroplasty, its measurement properties have been investigated for nonoperative management of osteoarthritis of the knee[26] and for soft-tissue knee pathology.[27] However, we believe this is the first study to investigate its measurement properties for revision knee arthroplasty. We corroborated findings from primary knee arthroplasty which showed that the OKS had good psychometric properties: high internal consistency, good convergent and divergent construct validity, and no significant floor or ceiling effects.[28–30] Our findings disagreed with a previous study which showed that divergent construct validity was unsatisfactory.[31] For the assessment of individual patients, we found a MIC using an anchor-based method of 7.5 points. This metric reflects what the patient

Table III. Internal consistency for Oxford Knee Score in revision knee arthroplasty.

Preoperative OKS (Overall Cronbach's α = 0.88)					
Item	**Mean item score**	**% at floor**	**% at ceiling**	**Item-total correlation**	**α if item removed**
1 Pain	0.58	54	1	0.57	0.88
2 Washing	2.65	2	28	0.65	0.88
3 Transport	1.88	2	5	0.69	0.87
4 Walking time	1.81	19	8	0.68	0.88
5 Standing	1.56	6	2	0.70	0.87
6 Limping	0.78	52	1	0.65	0.88
7 Kneeling	0.40	71	0	0.51	0.88
8 Night pain	1.10	39	6	0.62	0.88
9 Work interference	1.23	23	2	0.79	0.87
10 Instability	1.61	23	8	0.67	0.88
11 Shopping	1.48	31	8	0.76	0.87
12 Stairs	1.58	12	3	0.73	0.87
Postoperative OKS (Overall Cronbach's α = 0.94)					
Item	**Mean item score**	**% at floor**	**% at ceiling**	**Item-total correlation**	**α if item removed**
1 Pain	2.01	11	15	0.80	0.94
2 Washing	3.26	1	55	0.74	0.94
3 Transport	2.65	1	25	0.81	0.94
4 Walking time	2.78	6	37	0.78	0.94
5 Standing	2.57	1	21	0.83	0.94
6 Limping	2.31	16	23	0.83	0.94
7 Kneeling	0.94	54	4	0.62	0.95
8 Night pain	2.29	15	29	0.78	0.94
9 Work interference	2.46	6	23	0.88	0.93
10 Instability	2.96	5	43	0.77	0.94
11 Shopping	2.46	16	34	0.80	0.94
12 Stairs	2.41	7	22	0.79	0.94

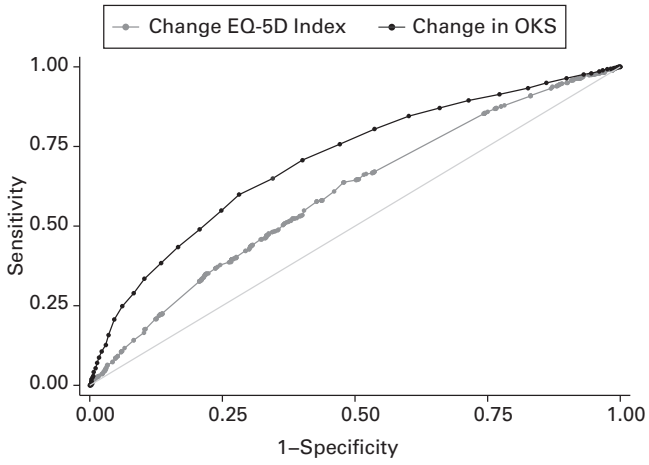

Fig. 4

Receiver operating characteristic (ROC) curve for revision knee arthroplasty comparing Oxford Knee Score (OKS) with EuroQol five-dimension (EQ-5D) Index. The minimally important change (MIC) for the OKS was 7.5 points (95% confidence interval 5.5 to 8.5) based on the optimal cut-off value with specificity 0.72 and sensitivity 0.60, with area under the ROC curve 0.66.

considers to be important – the difference between 'a little better' and 'about the same' – rather than statistical inference. In primary knee arthroplasty, the MIC for the OKS is similar at 6.5 points.[24] Future studies should investigate what constitutes a good outcome after revision knee arthroplasty.

This is the first study to evaluate the measurement properties of the OKS in revision knee arthroplasty. The study benefits from recruitment of a very large cohort of patients that had completed both the OKS and the EQ-5D, allowing assessment of construct validity. Our methodology is based on latest COSMIN guidance[16] and follows SEARCHeD reporting criteria.[17] The study does have some limitations, and it is important to interpret our findings in the context of the patient sample. Patients are typically enrolled in the NHS national PROMs programme from the outpatient or preoperative assessment clinic, while awaiting elective revision procedures. Patients attending with emergent, non-discretional reasons for revision (such as periprosthetic joint infection, or fracture) are less likely to be enrolled. As such, the measurement properties described are likely to be more applicable to patients with discretional reasons for revision (such as aseptic loosening). This study used a de-identified dataset where revision procedures were flagged by OPCS-4 codes through central linkage with HES.[18] Data are entered into HES by trained administrators, with NHS Trusts incentivized to code procedures accurately in order to receive the correct tariff. There is good evidence that HES is a reliable dataset for coding of comorbidities,[32,33] but more evidence is needed to determine to what extent this is true for procedural coding. The wider debate concerning which procedures should be considered to be a revision also needs to be addressed. For the purposes of this study, any procedure that contributes to NHS Digital PROMs performance reports for

revision knee arthroplasty was eligible for inclusion. This included procedures to replace a further compartment of the knee, which do not necessarily represent failure of the primary joint arthroplasty. There may be benefit in the future to examination of a PROM dataset linked to the National Joint Registry (NJR), both to investigate the quality of HES procedural coding and to examine the psychometric properties of the OKS for subgroups within the sample. However, data linkage is likely to be associated with heavy attrition of PROM data.[34,35] Future studies should examine the content validity of the OKS, particularly for non-discretional reasons for revision, where functional benefit may not be the principal objective of surgery.

This study has demonstrated that the OKS is a useful instrument to capture patients' perspectives on pain and function following revision knee arthroplasty. A valid score is important not only to interpret outcomes for individual patients, but also to compare across groups of patients and to allow meaningful participation of potential surgical candidates in shared decision-making. In the wider context, valid PROMs are necessary for health economic evaluations,[36] for managing healthcare providers,[8] and for implementing evidence-based service reconfigurations.[37] An outcomes framework to assess revision knee arthroplasty is needed and is likely to benefit from the inclusion of PROMs, but must recognize differences in the primary goal of surgery between different revision indications.

In conclusion, this study found that the OKS was a useful and valid instrument for assessment of outcome following revision knee arthroplasty. The OKS was responsive to change and demonstrated good measurement properties.

Take home message

- The Oxford Knee Score (OKS) was responsive to change with good measurement properties for the assessment of functional outcome following elective revision total knee arthroplasty (TKA).
- We have produced estimates for clinically meaningful changes in the OKS following revision TKA. These estimates will be useful for sample size calculations and clinical interpretation of studies using this instrument.
- The minimal important change (MIC) of the OKS was 7.5 points, the minimal important difference (MID) 5.2 points, and the minimal detectable change (MDC-90) 3.9 points.

Twitter

Follow S. A. Sabah @bonedoc247
Follow D. J. Beard @david_j_beard
Follow A. J. Price @oxfordkneegroup

Supplementary material

A table of Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures codes used to identify revision knee arthroplasty procedures from Hospital Episode Statistics, and a correlation matrix providing Spearman's rho for the Oxford Knee Score versus EQ-5D.

References

1. **No authors listed**. Patient reported outcome measures (PROMs). NHS digital. 2020. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms (date last accessed 2020 Apr 16).
2. **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-B(1):63–69.
3. **EuroQol Group**. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
4. **Baker PN**, **Deehan DJ**, **Lees D**, **et al**. The effect of surgical factors on early patient-reported outcome measures (PROMs) following total knee replacement. *J Bone Joint Surg Br*. 2012;94-B(8):1058–1066.
5. **Baker P**, **Cowling P**, **Kurtz S**, **Jameson S**, **Gregg P**, **Deehan D**. Reason for revision influences early patient outcomes after aseptic knee revision. *Clin Orthop Relat Res*. 2012;470(8):2244–2252.
6. **Wilson HA**, **Middleton R**, **Abram SGF**, **et al**. Patient relevant outcomes of unicompartmental versus total knee replacement: systematic review and meta-analysis. *BMJ*. 2019;28(286):l352–17.
7. **Johnson WB**, **Engh CA**, **Parks NL**, **Hamilton WG**, **Ho PH**, **Fricka KB**. A lower threshold for revision of aseptic unicompartmental vs total knee arthroplasty. *Bone Joint J*. 2020;102-B(6_Supple_A):91–95.
8. **Middleton R**, **Wilson HA**, **Alvand A**, **Abram SGF**, **Bottomley N**, **Jackson W**. Outcome-based commissioning of knee arthroplasty in the NHS: system error in a national monitoring programme and the unintended consequences on achieving the best practice tariff. *Bone Joint J*. 2018;100-B(12):1572–1578.
9. **No authors listed**. Clinician Feedback - Surgeon PROMs Reports. National Joint Registry. https://clinicianfeedback.njrcentre.org.uk/SurgeonPromsReports.aspx (date last accessed 28 April 2020).
10. **Price A**, **Smith J**, **Dakin H**, **et al**. The arthroplasty Candidacy help engine tool to select candidates for hip and knee replacement surgery: development and economic modelling. *Health Technol Assess*. 2019;23(32):1–216.
11. **No authors listed**. Patient decision support tool for joint replacement. National Joint Registry. https://jointcalc.shef.ac.uk/ (date last accessed 28 April 2020).
12. **Murray DW**, **Fitzpatrick R**, **Rogers K**, **et al**. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br*. 2007;89-B(8):1010–1014.
13. **Harris K**, **Dawson J**, **Gibbons E**, **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;7:101–108.
14. **Dawson J**, **Fitzpatrick R**, **Carr A**, **Murray D**. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78-B(2):185–190.
15. **Dawson J**, **Fitzpatrick R**, **Frost S**, **Gundle R**, **McLardy-Smith P**, **Murray D**. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br*. 2001;83-B(8):1125–1129.
16. **Prinsen CAC**, **Mokkink LB**, **Bouter LM**, **et al**. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
17. **Ollivere B**, **Metcalfe D**, **Perry DC**, **Haddad FS**. SEARCHeD: supporting evaluation, analysis and reporting of routinely collected healthcare data. *Bone Joint J*. 2020;102-B(2):145–147.
18. **NHS Digital**. Patient reported outcome measure (PROMs) in England: a guide to PROMs methodology. NHS Digital. 2017. https://digital.nhs.uk/binaries/content/assets/legacy/pdf/g/t/proms_guide_v12.pdf (date last accessed 31/01/2021).
19. **Harris K**, **Dawson J**, **Doll H**, **et al**. Can pain and function be distinguished in the Oxford knee score in a meaningful way? an exploratory and confirmatory factor analysis. *Qual Life Res*. 2013;22(9):2561–2568.
20. **Weale AE**, **Halabi OA**, **Jones PW**, **White SH**. Perceptions of outcomes after unicompartmental and total knee replacements. *Clin Orthop Relat Res*. 2001;382:143–153.
21. **NHS England Analytical Team**. Patient reported outcome measures: update to reporting and case-mix adjusting hip and knee procedure data. 2013.
22. **Munro B**. *Statistical methods for health care research*. 5th edition: Lippincott Williams and Wilkins, 2005.
23. **Price AJ**, **Ogollah R**, **Kang S**, **et al**. Determining responsiveness and meaningful changes for the musculoskeletal health questionnaire (MSK-HQ) for use across musculoskeletal care pathways. *BMJ Open*. 2019;9(10):e025357.
24. **Beard DJ**, **Harris K**, **Dawson J**, **et al**. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68(1):73–79.
25. **Gibbons E**, **Hewitson P**, **Morley D**, **Jenkinson C**, **Fitzpatrick R**. The outcomes and experiences questionnaire: development and validation. *Patient Relat Outcome Meas*. 2015;6:179–189.
26. **Harris KK**, **Dawson J**, **Jones LD**, **Beard DJ**, **Price AJ**. Extending the use of PROMs in the NHS--using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: a validation study. *BMJ Open*. 2013;3(8):e003365.
27. **Al-Arabi Y**, **Murray J**, **Wyatt M**, **Satish V**, **Deo S**. The Oxford knee score in soft tissue knee pathology. *Orthop Proc*. 2008;90-B(Supp_II).

28. **Martín-Fernández J**, **Gray-Laymón P**, **Molina-Siguero A**, **et al**. Cross-Cultural adaptation and validation of the Spanish version of the Oxford hip score in patients with hip osteoarthritis. *BMC Musculoskelet Disord*. 2017;18(1):205.
29. **Lim CR**, **Harris K**, **Dawson J**, **Beard DJ**, **Fitzpatrick R**, **Price AJ**. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open*. 2015;5(7):e007765.
30. **Harris K**, **Lim CR**, **Dawson J**, **Fitzpatrick R**, **Beard DJ**, **Price AJ**. The Oxford knee score and its subscales do not exhibit a ceiling or a floor effect in knee arthroplasty patients: an analysis of the National health service PROMs data set. *Knee Surg Sports Traumatol Arthrosc*. 2017;25(9):2736–2742.
31. **Xie F**, **Ye H**, **Zhang Y**, **Liu X**, **Lei T**, **Li S-C**. Extension from inpatients to outpatients: validity and reliability of the Oxford knee score in measuring health outcomes in patients with knee osteoarthritis. *Int J Rheum Dis*. 2011;14(2):206–210.
32. **Crooks CJ**, **West J**, **Card TR**. A comparison of the recording of comorbidity in primary and secondary care by using the Charlson index to predict short-term and long-term survival in a routine linked data cohort. *BMJ Open*. 2015;5(6):e007974.
33. **Wright FL**, **Green J**, **Canoy D**, **et al**. Vascular disease in women: comparison of diagnoses in hospital episode statistics and general practice records in England. *BMC Med Res Methodol*. 2012;12(161):161.
34. **Sabah SA**, **Henckel J**, **Koutsouris S**, **et al**. Are all metal-on-metal hip revision operations contributing to the National joint registry implant survival curves? *Bone Joint J*. 2016;98-B(1):33–39.
35. **Sabah SA**, **Henckel J**, **Cook E**, **et al**. Validation of primary metal-on-metal hip arthroplasties on the National joint Registry for England, Wales and Northern Ireland using data from the London implant retrieval centre: a study using the NJR dataset. *Bone Joint J*. 2015;97-B(1):10–18.
36. **Eibich P**, **Dakin HA**, **Price AJ**, **Beard D**, **Arden NK**, **Gray AM**. Associations between preoperative Oxford hip and knee scores and costs and quality of life of patients undergoing primary total joint replacement in the NHS England: an observational study. *BMJ Open*. 2018;8(4):e019477.
37. **Ahmed SS**, **Haddad FS**. Networks in orthopaedics. *Bone Joint J*. 2020;102-B(3):273–275.

Author information:
S. A. Sabah, FRCS(Orth), DPhil Student
D. J. Beard, GDPhys, MCSP, MSc, MA, DPhil, FRCS (hon), Professor of Musculoskeletal and Surgical Science
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, UK.

A. Alvand, DPhil, FRCS(Orth), Consultant Orthopaedic Surgeon
A. J. Price, DPhil, FRCS(Orth), Professor of Orthopaedics
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, UK; Nuffield Orthopaedic Centre, Oxford, UK.

Author contributions:
S. A. Sabah: Conceptualized the study and methodology, Conducted the analysis, Wrote and Edited the paper, Acted as the guarantor.
A. Alvand: Critically revised the paper.
D. J. Beard: Critically revised the paper.
A. Price: Conceptualized the study and methodology, Wrote and edited the paper.

Funding statement:
No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article

ICMJE COI statement:
S. Sabah has received funding from the Royal College of Surgeons One-Year Fellowship and Rosetrees Trust. Neither funding source had any role in the design or conduct of this study.

Acknowledgements:
We would like to acknowledge all patients who contributed to the NHS PROMs programme over the study period.

Ethical review statement:
Ethical approval was not required for this study as it used de-identified data freely available in the public domain.

This article was primary edited by G. Scott.