



## ■ KNEE

# Mapping analysis to predict the associated EuroQol five-dimension three-level utility values from the Oxford Knee Score

## A PREDICTION AND VALIDATION STUDY

**N. D. Clement,  
I. Afzal,  
C. J. H. Peacock,  
D. MacDonald,  
G. J. Macpherson,  
J. T. Patton,  
V. Asopa,  
D. H. Sochart,  
D. F. Kader**

From South West  
London Elective  
Orthopaedic Centre,  
Epsom, UK

### Aims

The aims of this study were to assess mapping models to predict the three-level version of EuroQoL five-dimension utility index (EQ-5D-3L) from the Oxford Knee Score (OKS) and validate these before and after total knee arthroplasty (TKA).

### Methods

A retrospective cohort of 5,857 patients was used to create the prediction models, and a second cohort of 721 patients from a different centre was used to validate the models, all of whom underwent TKA. Patient characteristics, BMI, OKS, and EQ-5D-3L were collected preoperatively and one year postoperatively. Generalized linear regression was used to formulate the prediction models.

### Results

There were significant correlations between the OKS and EQ-5D-3L preoperatively ( $r = 0.68$ ;  $p < 0.001$ ) and postoperatively ( $r = 0.77$ ;  $p < 0.001$ ) and for the change in the scores ( $r = 0.61$ ;  $p < 0.001$ ). Three different models (preoperative, postoperative, and change) were created. There were no significant differences between the actual and predicted mean EQ-5D-3L utilities at any timepoint or for change in the scores ( $p > 0.090$ ) in the validation cohort. There was a significant correlation between the actual and predicted EQ-5D-3L utilities preoperatively ( $r = 0.63$ ;  $p < 0.001$ ) and postoperatively ( $r = 0.77$ ;  $p < 0.001$ ) and for the change in the scores ( $r = 0.56$ ;  $p < 0.001$ ). Bland-Altman plots demonstrated that a lower utility was overestimated, and higher utility was underestimated. The individual predicted EQ-5D-3L that was within  $\pm 0.05$  and  $\pm 0.010$  (minimal clinically important difference (MCID)) of the actual EQ-5D-3L varied between 13% to 35% and 26% to 64%, respectively, according to timepoint assessed and change in the scores, but was not significantly different between the modelling and validation cohorts ( $p \geq 0.148$ ).

### Conclusion

The OKS can be used to estimate EQ-5D-3L. Predicted individual patient utility error beyond the MCID varied from one-third to two-thirds depending on timepoint assessed, but the mean for a cohort did not differ and could be employed for this purpose.

**Cite this article:** *Bone Jt Open* 2022;3-7:573–581.

**Keywords:** EQ-5D, Oxford Knee Score, Arthroplasty, Outcome, Predictors

Correspondence should be sent to  
Nick D Clement; email:  
nickclement@doctors.org.uk

doi: 10.1302/2633-1462.37.BJO-  
2022-0054.R1

*Bone Jt Open* 2022;3-7:573–581.

### Introduction

The Oxford Knee Score (OKS) is a validated joint-specific patient-reported outcome

measure (PROM) that is commonly employed to assess the outcome of total knee arthroplasty (TKA).<sup>1,2</sup> An alternative

measure of outcome is health-related quality of life (HRQoL) following TKA using a generic tool such as the EuroQol five-dimension health questionnaire (EQ-5D).<sup>3</sup> Such generic tools enable comparison of differing interventions on a patient's HRQoL across medicine.<sup>4</sup> The defined health state, or utility index, can then be assigned to each patient, and quality-adjusted life years (QALYs) gained or lost can then be calculated. The cost per QALYs can then be used to compare the cost-effectiveness of differing interventions across medicine, of which TKA is one of the most cost-effective.<sup>5</sup> However, the EQ-5D is not always collected and therefore makes cost economic studies difficult to undertake.

Mapping or cross walking methods can be employed to predict the EQ-5D index based on other available outcome measures, such as the joint-specific PROM OKS.<sup>6</sup> It is acknowledged that the development of such algorithms to predict the EQ-5D utility results in information loss and increased uncertainty, and is no substitute for direct measurement.<sup>7</sup> Two previous studies have used mapping analysis to predict EQ-5D utilities from the OKS and allow HRQoL to be estimated.<sup>7,8</sup> One of these studies was from Spain; their health utility index differs from that used in the UK, and is therefore not applicable to the UK population.<sup>8</sup> The other study by Dakin et al<sup>7</sup> used data from the UK to map the three-level EQ-5D (EQ-5D-3L) from the OKS; however, the mean absolute error was 0.129 and only 42% of predicted values were within  $\pm 0.1$  of the actual value on external validation of their mapping algorithm. They combined pre- and postoperative outcomes (OKS and EQ-5D-3L) into one model, but it is recognized that these two timepoints demonstrated different distributions of data.<sup>7,9</sup> Furthermore, the authors are not aware of previous published mapping analysis for change in the EQ-5D-3L utility, pre- to postoperatively, which is required to assess QALY gain as part of cost utility analysis.

The aims of this study were to assess mapping models to predict EQ-5D-3L from the OKS, and to validate these before and after TKA.

## Methods

Patients were identified retrospectively from two prospectively compiled established arthroplasty databases held at the South West London Elective Orthopaedic Centre (SWLEOC) and Royal Infirmary of Edinburgh (RIE). The data from SWLEOC were used to construct the prediction models. During a nine-year period (June 2007 to November 2016), 5,857 patients undergoing primary TKA at SWLEOC were asked to complete pre- and postoperative patient questionnaires. The second cohort of 721 patients, also undergoing a TKA during a 21-month period (January 2017 to

September 2018), were recruited from the RIE and used to validate the prediction models.

Patient characteristics and BMI were recorded preoperatively, and pre- and one-year postoperative PROMs were collected at both centres. The EQ-5D general health questionnaire evaluates five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression).<sup>3</sup> The three-level version of the questionnaire was used, with the responses to the five domains being recorded at three levels of severity. An individual patient's health state can be reported based on a three-digit code for each domain, resulting in 243 possible health states.<sup>10</sup> Each health state was converted to a single summary index by applying a weighting. These are specific to the UK population and are based on a time trade-off technique. This index is on a scale of -0.584 to 1, where 1 represents perfect health, zero represents death, and negative values represent a state perceived as worse than death.<sup>11</sup> A minimal clinically important difference (MCID) is the smallest change of a score to be of importance, but to the authors' knowledge this is to be defined following TKA. A review found the MCID to vary in orthopaedics from 0.03 to 0.54 depending on the intervention being assessed. For the current study 0.1 was defined as MCID.<sup>12</sup> The OKS consists of 12 questions that were assessed using a Likert scale with values from 0 to 4; a summative score is then calculated where 48 is the best possible score (least symptomatic) and 0 is the worst possible score (most symptomatic).<sup>1,2</sup>

**Statistical analysis.** Statistical analysis was performed using SPSS v. 17.0 (SPSS, USA). Simple descriptive analysis was undertaken. The ceiling EQ-5D-3L utility score was defined as achieving a maximal of 1.0 in the pre- or postoperative EQ-5D-3L assessment. Student's *t*-tests, paired and independent-samples, were used to compare linear variables between groups. Pearson or Spearman correlations were used to assess the association between linearly related variables. Dichotomous variables were assessed using a chi-squared test. Generalized linear regression analyses were used to map the EQ-5D-3L utility to the OKS pre- and postoperatively and for change in scores (without including of other factors). These models were then used to predict the EQ-5D-3L utility for both the modelling cohort (SWLEOC) and the validation cohort (RIE). The difference from the predicted and actual EQ-5D-3L was categorized as  $\pm 0.05$  and  $\pm 0.10$  (MCID) to allow for comparison with the mapping algorithm of Dakin et al.<sup>7</sup> Bland-Altman limits of agreement (predicted versus actual) were calculated and plotted. Bland and Altman recommend that the differences between each of the two scores be compared, plotting the differences against the means of the scores.<sup>13</sup> No linear relationship on the Bland-Altman plot indicates that the statistical variation was similar for

individuals with low clinical measurement scores and high clinical measurement scores. A p-value of < 0.05 was defined as statistically significant.

There was no additional patient contact and as such, this project was performed as a service evaluation without the need for formal ethical approval. The project was registered with the institution’s audit department, and was conducted in accordance with the Declaration of Helsinki<sup>14</sup> and the guidelines for Good Clinical Practice.<sup>15</sup>

**Results**

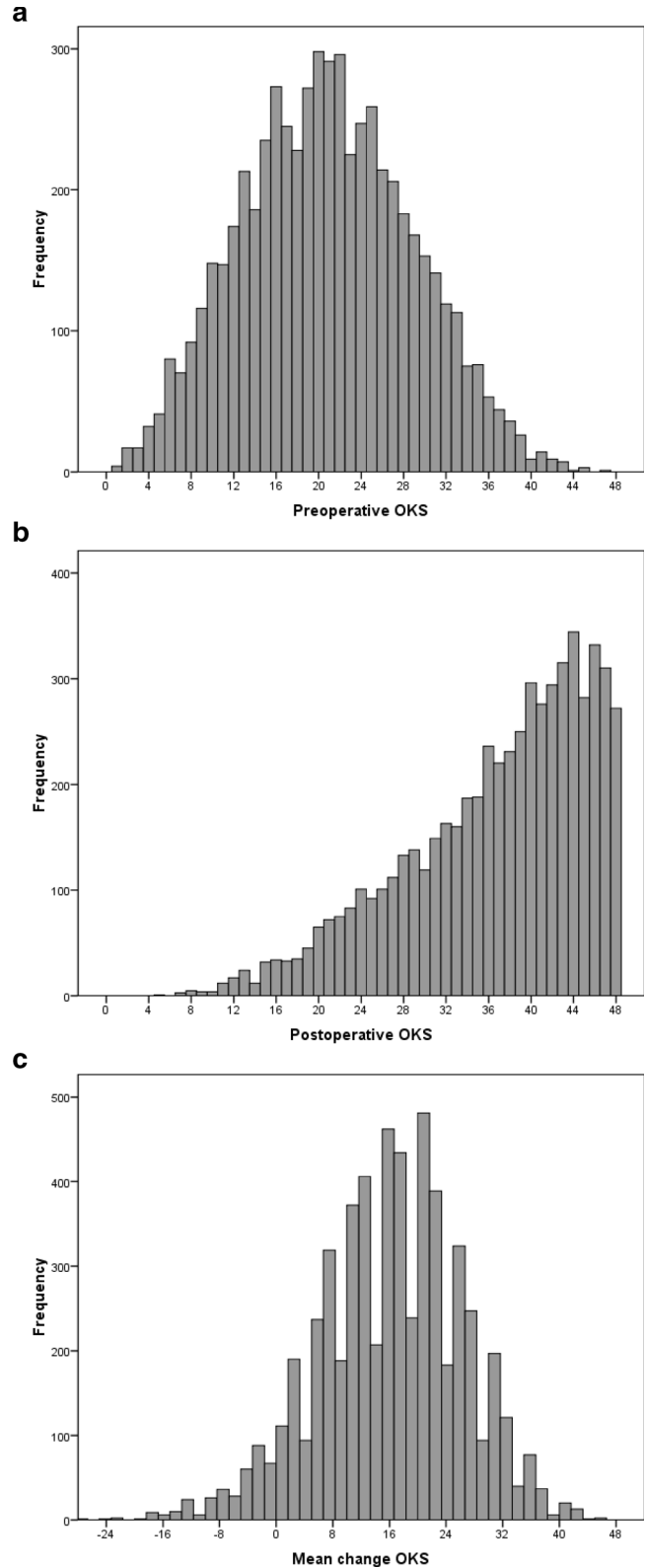
**Prediction models.** The mean preoperative OKS for the modelling cohort (SWLEOC) was 20.8 (standard deviation (SD) 8.0), which improved to 36.7 (SD 8.7) at one year (change 15.9 (95% confidence interval (CI) 15.7 to 16.2);  $p < 0.001$ , paired *t*-test). A normal distribution was observed for the preoperative and change in the OKS scores, but the postoperative scores were skewed to the left and demonstrated a ceiling effect (Figure 1). A bimodal distribution was observed in the preoperative EQ-5D-3L index, a normal distribution for the change, and a skewed-to-the-left distribution in the postoperative EQ-5D-3L, demonstrating a ceiling effect (Figure 2). There was a significant correlation between the OKS and EQ-5D-3L index preoperatively ( $r = 0.68$ ;  $p < 0.001$ , Spearman correlation) and postoperatively ( $r = 0.77$ ;  $p < 0.001$ , Spearman correlation), and for the change ( $r = 0.61$ ;  $p < 0.001$ , Pearson correlation) in the scores (Figure 3).

Due to the differing data distributions, three different models were created to predict the associated EQ-5D-3L. Generalized linear regression demonstrated a significant association between the OKS and EQ-5D-3L that explained between 37% and 59% of the variations in the data (Table I). These analyses were used to create equations that were predictive of the associated EQ-5D-3L index at the different timepoints and for change in the score:

$$\begin{aligned} \text{EQ-5D}_{\text{preoperative}} &= (\text{OKS}_{\text{preoperative}} \times 0.026) - 0.114 \\ \text{EQ-5D}_{\text{postoperative}} &= (\text{OKS}_{\text{postoperative}} \times 0.018) + 0.128 \\ \text{EQ-5D}_{\text{change}} &= (\text{OKS}_{\text{change}} \times 0.020) + 0.029 \end{aligned}$$

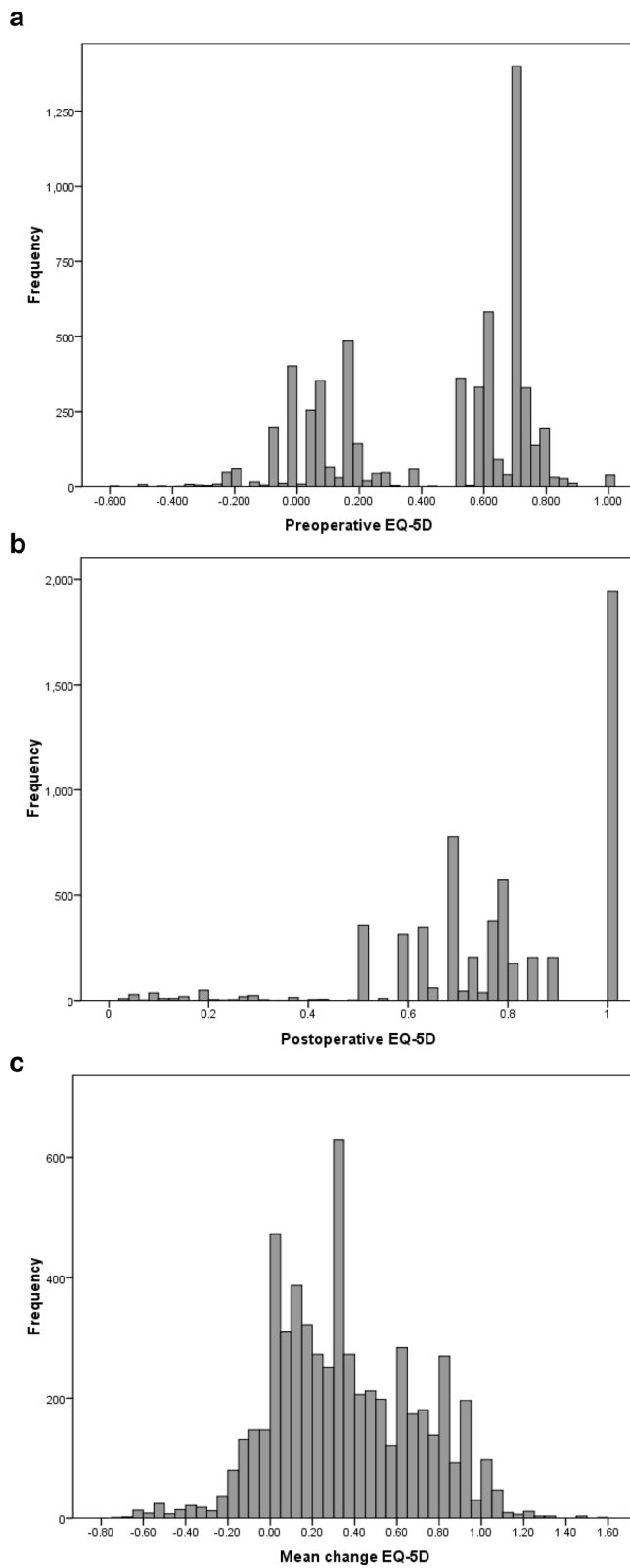
Using these equations, the predicted ED-5D-3L was calculated from the OKS at the differing timepoints and for change in the score using the SWLEOC data (Table II). There was a significantly lower preoperative and greater postoperative predicted EQ-5D-3L index, but no difference in the change in score (Table III). However, the mean differences of between 0.003 and 0.007 were less than the MCID (0.1). The predicted EQ-5D-3L that was within 0.05 and 0.10 of the actual EQ-5D-3L varied between 13% to 35% and 27% to 66%, respectively, according to timepoint assessed and change in the score (Table IV).

**Validation of prediction models.** There were significant differences between the RIE (validation) and SWLEOC



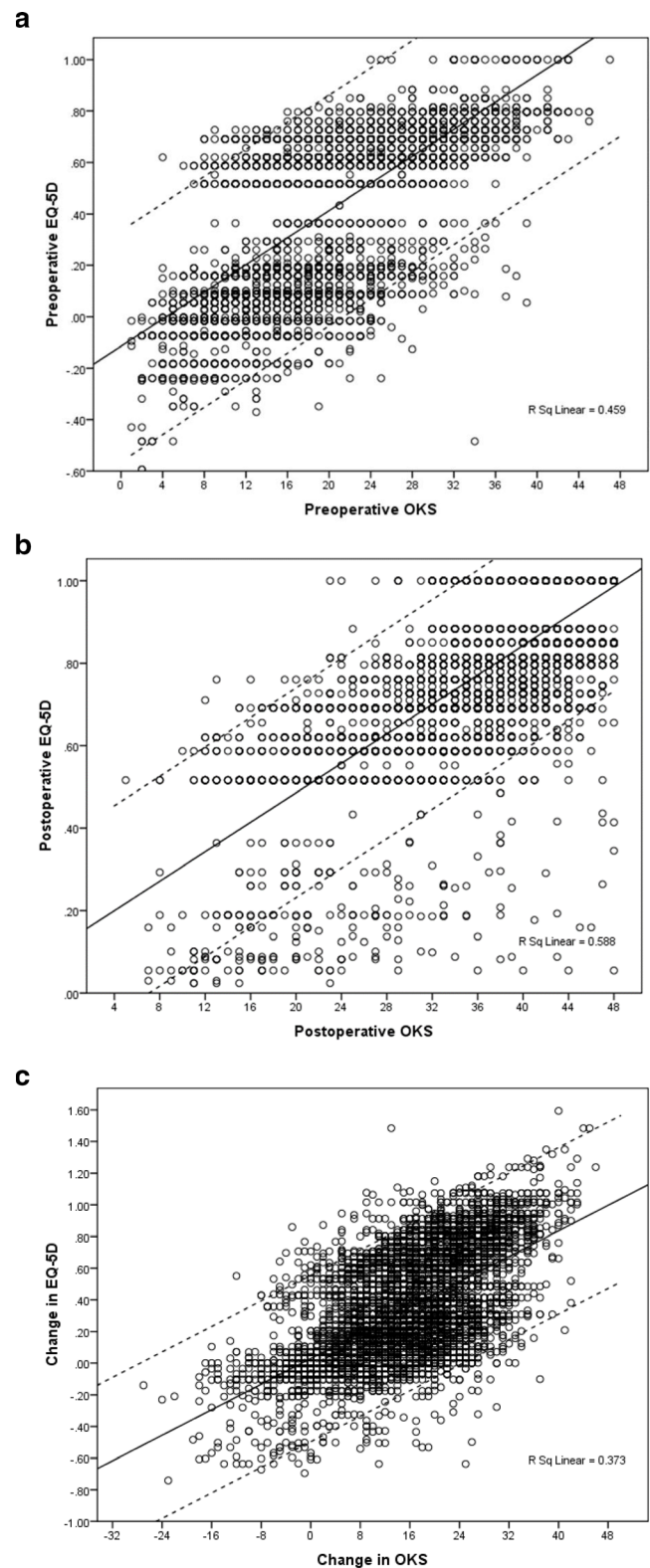
**Fig. 1**

Histograms illustrating the distribution of the a) preoperative, b) postoperative, and c) change in the Oxford Knee Score (OKS) after total knee arthroplasty.



**Fig. 2**

Histograms illustrating the distribution of the a) preoperative, b) postoperative, and c) change in the EuroQol five-dimension health questionnaire (EQ-5D) after total knee arthroplasty.



**Fig. 3**

Scatter plots for the relationship between the Oxford Knee Score (OKS) and EuroQol five-dimension (EQ-5D) utility a) preoperatively, b) postoperatively, and c) change after total knee arthroplasty. The solid black line is the linear line of best fit ( $R^2$ ) and the dashed lines are the 95% confidence intervals.

**Table I.** Simple linear regression analysis for the association of the preoperative, postoperative, and change in the Oxford Knee Score with the EuroQol five-dimension index.

Timepoint	Variable	B (95% CI)	p-value
Preoperative (R2 = 0.459)	Constant	-0.114 (-0.131 to -0.098)	< 0.001
	OKS	0.026 (0.026 to 0.027)	< 0.001
Postoperative (R2 = 0.588)	Constant	0.128 (0.114 to 0.143)	< 0.001
	OKS	0.018 (0.017 to 0.018)	< 0.001
Change (R2 = 0.373)	Constant	0.029 (0.016 to 0.042)	< 0.001
	OKS	0.020 (0.020 to 0.021)	< 0.001

CI, confidence interval; OKS, Oxford Knee Score.

**Table II.** Associated EuroQol five-dimension three-level utility according to preoperative, postoperative, or change in the Oxford Knee Score using the validated equations.

OKS	Associated EQ-5D-3L utility			OKS	Associated EQ-5D-3L utility		
	Preoperative	Postoperative	Change		Preoperative	Postoperative	Change
1	-0.088	0.146	0.049	25	0.536	0.578	0.529
2	-0.062	0.164	0.069	26	0.562	0.596	0.549
3	-0.036	0.182	0.089	27	0.588	0.614	0.569
4	-0.010	0.200	0.109	28	0.614	0.632	0.589
5	0.016	0.218	0.129	29	0.640	0.650	0.609
6	0.042	0.236	0.149	30	0.666	0.668	0.629
7	0.068	0.254	0.169	31	0.692	0.686	0.649
8	0.094	0.272	0.189	32	0.718	0.704	0.669
9	0.120	0.290	0.209	33	0.744	0.722	0.689
10	0.146	0.308	0.229	34	0.770	0.740	0.709
11	0.172	0.326	0.249	35	0.796	0.758	0.729
12	0.198	0.344	0.269	36	0.822	0.776	0.749
13	0.224	0.362	0.289	37	0.848	0.794	0.769
14	0.250	0.380	0.309	38	0.874	0.812	0.789
15	0.276	0.398	0.329	39	0.900	0.830	0.809
16	0.302	0.416	0.349	40	0.926	0.848	0.829
17	0.328	0.434	0.369	41	0.952	0.866	0.849
18	0.354	0.452	0.389	42	0.978	0.884	0.869
19	0.380	0.470	0.409	43	1.000	0.902	0.889
20	0.406	0.488	0.429	44	1.000	0.920	0.909
21	0.432	0.506	0.449	45	1.000	0.938	0.929
22	0.458	0.524	0.469	46	1.000	0.956	0.949
23	0.484	0.542	0.489	47	1.000	0.974	0.969
24	0.510	0.560	0.509	48	1.000	0.992	0.989

EQ-5D-3L, EuroQol five-dimension three-level health questionnaire; OKS, Oxford Knee Score.

cohorts but the differences between the groups in the OKS and EQ-5D-3L were less than the MCID (Table V). The mean preoperative OKS for the RIE cohort was 20.7 (SD 7.4), which improved to 35.6 (SD 9.5) at one year (change 14.8 (95% CI 14.2 to 15.5);  $p < 0.001$ , paired  $t$ -test). The mean preoperative EQ-5D-3L was 0.422 (SD 0.303), which improved to 0.758 (SD 0.235) at one year (change 0.336 (95% CI 0.314 to 0.360);  $p < 0.001$ , paired  $t$ -test).

There were no significant differences between the actual and predicted mean EQ-5D-3L utilities at any timepoint or for change in the scores (Table III). The 95% CI around the error of the predicted utility compared to the actual utility varied: preoperative was  $\pm 0.459$ ,

postoperative was  $\pm 0.299$ , and change was  $\pm 0.521$  (Figure 4). There was a significant correlation between the actual and predicted EQ-5D-3L index pre- ( $r = 0.63$ ;  $p < 0.001$ , Spearman correlation) and postoperatively ( $r = 0.77$ ;  $p < 0.001$ , Spearman correlation), and for the change ( $r = 0.56$ ;  $p < 0.001$ , Pearson correlation) in the scores. Bland-Altman plots demonstrated a liner trend in the error between the actual and predicted EQ-5D-3L utilities, with a lower utility being overestimated and higher utility being underestimated for the pre- and postoperative scores and change in score (Figure 4). The predicted EQ-5D-3L that was within 0.05 and 0.10 of the actual EQ-5D-3L varied between 13% to 35% and 26% to 64%, respectively, according to timepoint



**Table III.** The difference in the actual and predicted (from the Oxford Knee Score) preoperative, postoperative, and change in the EuroQol five-dimension utilities.

Variable	Actual	Predicted	Mean difference (95% CI)	p-value*
<b>SWLEOC (n = 5857)</b>				
Preoperative	0.434 (0.311)	0.427 (0.208)	0.007 (0.002 to 0.013)	0.014
Postoperative	0.784 (0.202)	0.789 (0.156)	0.005 (0.001 to 0.008)	0.007
Change	0.350 (0.338)	0.347 (0.204)	0.003 (-0.004 to 0.010)	0.425
<b>RIE (n = 721)</b>				
Preoperative	0.422 (0.303)	0.425 (0.193)	0.004 (-0.014 to 0.021)	0.680
Postoperative	0.758 (0.235)	0.768 (0.170)	0.009 (-0.001 to 0.020)	0.091
Change	0.336 (0.321)	0.325 (0.186)	0.012 (-0.009 to 0.030)	0.277

\*Independent-samples *t*-test.

CI, confidence interval; RIE, Royal Infirmary of Edinburgh; SWLEOC, South West London Elective Orthopaedic Centre.

**Table IV.** Predicted EuroQol five-dimension three-level utility within 0.05 and 0.10 of the actual utility according to timepoint and change in score.

Predicted within	SWLEOC (n = 5857)		RIE (n = 721)		OR (95% CI)	p-value*
	Yes, n (%)	No, n (%)	Yes, n (%)	No, n (%)		
<b>Preoperative</b>						
± 0.05	768 (13.1)	5,089 (86.9)	90 (12.5)	631 (87.5)	0.95 (0.75 to 1.19)	0.639
± 0.10	1678 (28.6)	4,179 (71.4)	188 (26.1)	533 (73.9)	0.88 (0.73 to 1.05)	0.148
<b>Postoperative</b>						
± 0.05	2,066 (35.3)	3,791 (64.7)	255 (35.4)	466 (64.6)	1.00 (0.85 to 1.18)	0.999
± 0.10	3,848 (65.7)	2,009 (34.3)	483 (63.7)	262 (36.3)	0.91 (0.78 to 1.07)	0.277
<b>Change</b>						
± 0.05	786 (13.4)	5,071 (86.6)	94 (13.0)	627 (87.0)	0.967 (0.77 to 1.21)	0.777
± 0.10	1,563 (26.7)	4,294 (73.3)	187 (25.9)	534 (74.1)	0.96 (0.81 to 1.15)	0.671

\*Chi-squared test.

CI, confidence interval; OR, odds ratio; RIE, Royal Infirmary of Edinburgh; SWLEOC, South West London Elective Orthopaedic Centre.

assessed and change in the score (Table IV), and was not significantly different from the SWLEOC group.

## Discussion

This study has shown that the EQ-5D-3L utility index can be predicted from the OKS. The pre- and postoperative scores and the change in score all demonstrated a different distribution, and prediction models were therefore created for each. There were no clinical or statistically significant differences between the actual or predicted mean EQ-5D-3L utilities in the validation cohort. However, the individual patient-predicted EQ-5D-3L utility to within ± 0.1, being defined as the MCID, did vary from 26% (preoperative) to 64% (postoperative). A linear trend in the difference between actual and predicted EQ-5D-3L utilities was demonstrated, with an actual lower index (worse) being overestimated (better) and conversely a higher index being underestimated.

There are several limitations to the current study. Although a significant correlation was demonstrated between the OKS and EQ-5D-3L, it was only strong ( $r > 0.7$ ) between postoperative scores. This might be expected, as these two scores do not measure the same aspect of a patient's health/function and therefore a strong correlation may not be expected for preoperative and change in scores, however the *r* value was greater

than 0.6. Previous authors have mapped each individual question in the OKS to each of the EQ-5D-3L dimensions.<sup>7,8</sup> The current study only used the composite OKS to map onto the EQ-5D-3L, and produced only one coefficient for each timepoint and change in the score. The main reason for doing this was to keep the algorithm simple and easy to use (Table II), and therefore improve the uptake and limit interpretation errors. Also, only one model (generalized linear) was used to predict the EQ-5D-3L utility. Again, this was to keep the interpretation of the mapping as clear as possible with the production of one set of modelling coefficients. The generalized linear model used has been shown to be optimal in a prior mapping study.<sup>8</sup> These limitations, however, do not seem to have resulted in diminished accuracy in prediction of the EQ-5D-3L when compared to other studies that have produced multiple coefficients for component scores and used multiple models.<sup>7,8</sup> Other patient-related factors may influence the mapping and prediction, such as age and sex, but again to have included these in the models would have resulted in complex algorithms. Including other factors such as sex and age may have helped the predictive value of the models and improved the  $R^2$ . However, the observed demographic differences between the modelling cohort and the validation cohort (Table V) did not seem to influence the prediction of the

**Table V.** Comparison of baseline demographics and patient-reported outcome measures between the two cohorts.

Variable	SWLEOC (n = 5,857)	RIE (n = 721)	Difference/OR (95% CI)	p-value
<b>Sex, n (%)</b>				< 0.001*
Male	2,204 (37.6)	321 (44.5)	0.75 (0.64 to 0.88)	
Female	3,653 (62.4)	400 (55.5)		
Mean age, yrs (SD)	71.1 (9.0)	69.7 (9.1)	1.5 (0.8 to 2.2)	< 0.001†
<b>BMI, n (%)</b>				< 0.001*
Underweight	24 (0.4)	4 (0.6)	N/A	
Normal	743 (12.7)	91 (12.6)		
Overweight	2,108 (36.0)	255 (35.4)		
Obese I	1,893 (32.3)	234 (32.5)		
Obese II	726 (12.4)	90 (12.5)		
Obese III	363 (6.2)	44 (6.1)		
Unknown	0	3 (0.4)		
<b>PROMs</b>				
<b>Preoperative mean (SD)</b>				
OKS	20.8 (8.0)	20.7 (7.4)	0.1 (-0.6 to 0.7)	0.850†
EQ-5D	0.434 (0.311)	0.422 (0.303)	0.012 (-0.012 to 0.036)	0.309†
<b>Postoperative mean (SD)</b>				
OKS	36.7 (8.7)	35.6 (9.5)	1.2 (0.5 to 1.8)	0.001†
EQ-5D	0.784 (0.202)	0.758 (0.235)	0.027 (0.011 to 0.042)	0.001†
<b>Mean change (SD)</b>				
OKS	15.9 (10.2)	14.8 (9.3)	1.1 (0.3 to 1.9)	0.006†
EQ-5D	0.350 (0.338)	0.336 (0.321)	0.014 (-0.012 to 0.040)	0.291†

\*Chi-squared test.

†Independent-samples t-test.

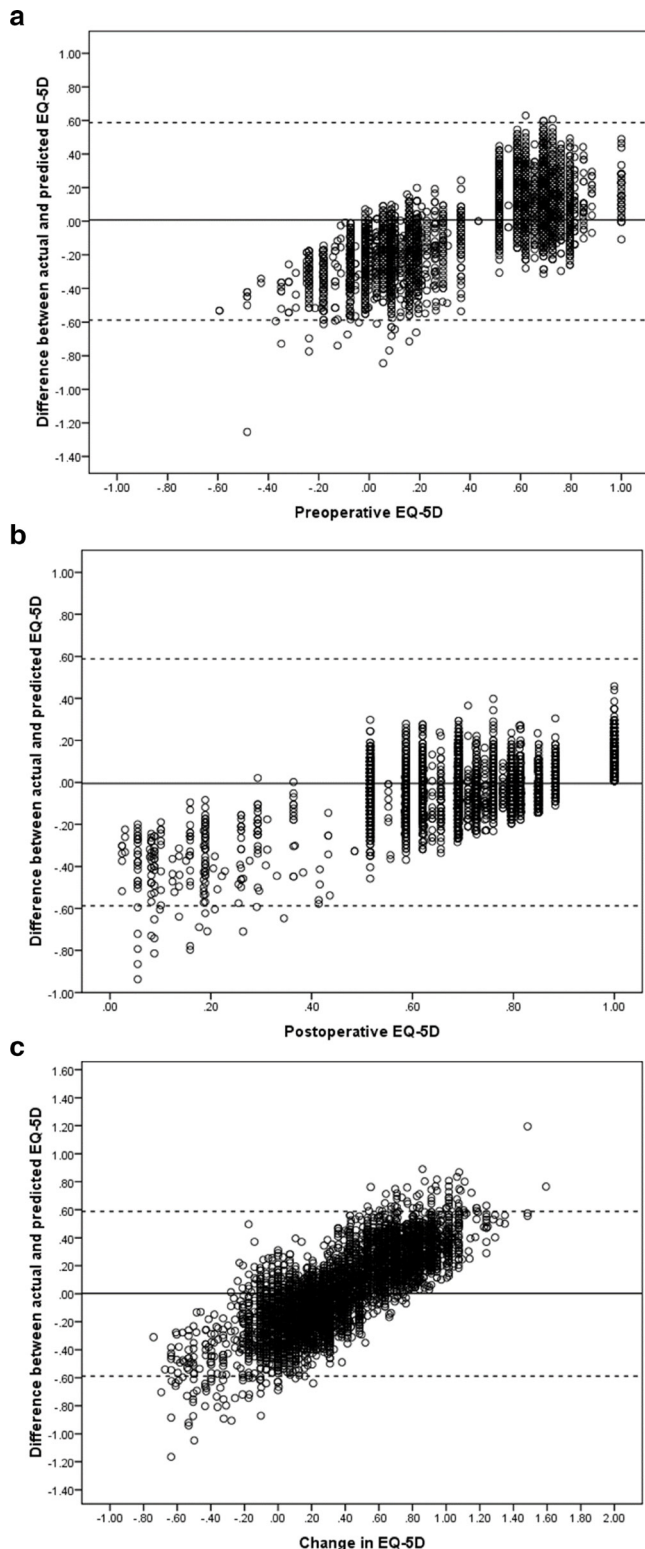
CI, confidence interval; EQ-5D, EuroQol five-dimension health questionnaire; OKS, Oxford Knee Score; OR, odds ratio; PROMs, patient-reported outcome measure; RIE, Royal Infirmary of Edinburgh; SD, standard deviation; SWLEOC, South West London Elective Orthopaedic Centre.

EQ-5D-3L utility, but this would need to be assessed in future studies.

The study by Dakin et al,<sup>7</sup> the first group to map the OKS onto the EQ-5D-3L, used data from two different datasets to construct their models, and validated these models using a further different dataset. They combined pre- and postoperative (following TKA) scores into one dataset for analysis. The current study has demonstrated differing data distributions according to timepoint assessed and for the change in the scores. These different datasets were also shown to have different modelling coefficients (Table I), ranging from 0.018 to 0.026 for variation in the EQ-5D-3L utility for each point change in the OKS post- and preoperatively, respectively. Although there were no clinically ( $\leq 0.012$ ) or statistically significant ( $p \geq 0.091$ ) mean differences between the actual and predicted EQ-5D-3L utilities for the validation cohort (RIE), the reliability to predict an individual patient's EQ-5D-3L to within  $\pm 0.1$  varied from 26% (preoperative and change in score) to 64% (postoperative). This is a similar observation to Dakin et al,<sup>7</sup> who found their model to offer better postoperative prediction accuracy. Their model offered a reliability of 42% to predict the EQ-5D-3L to within  $\pm 0.1$  of the actual score, which is similar to the current study of 44% ( $n = 671/1,516$ ) when combining pre- and

postoperative datasets from Table IV. This may suggest that the models presented in the current study and that of Dakin et al<sup>7</sup> are not sufficiently reliable to predict an individual patient's EQ-5D-3L utility. However, there is a recognized test-retest variation in response to each of the five dimensions of the EuroQol assessment in patients with knee arthritis,<sup>16</sup> with only moderate intra-class correlation.<sup>17</sup> Therefore, the prediction error associated with an individual patient's EQ-5D-3L utility may, in part, reflect patient variation in completion of the EQ-5D-3L questionnaire.

The most novel finding of the current study was the accuracy of the mean predicted and actual EQ-5D utilities, with a difference less than the MCID when defined as 0.1. Dakin et al,<sup>7</sup> in their mapping model, found a 0.129 mean absolute error between the predicted and actual EQ-5D utilities. A similar mean absolute error was also demonstrated in the mapping of the Spanish versions of OKS to the EQ-5D of 0.113, using their generalized linear model.<sup>8</sup> Mean absolute errors of between 0.001 to 0.190 are described in studies mapping non-preference-based measures, such as the OKS, to generic preference-based measures such as the EQ-5D.<sup>6</sup> However, if the upper limit of the MCID in the EQ-5D is taken to be 0.1, as it was in the current study, then an error greater than this may result in clinically significant changes being missed



**Fig. 4**

Bland-Altman plots for the relationship between the EuroQol five-dimension (EQ-5D) utility and error a) preoperatively, b) postoperatively, and c) change after total knee arthroplasty. The solid black line is the mean error and the dashed lines are the 95% confidence intervals.

or, conversely, a clinically significant change is assumed when there may not have been an actual change. Further validation of the presented model is required in other centres with different case mixes, but it would appear that the suggested mapping of the OKS onto the EQ-5D offers comparable values to that expected when pre- and postoperative and change data are assessed separately.

There is no MCID for the EQ-5D-3L utility following TKA of which the authors are aware. The suggested MCID in the current study of 0.1 was taken as an average of the available MCIDs in musculoskeletal medicine, but these range from 0.03 to 0.54 depending on the intervention being assessed.<sup>12</sup> The MCID following total hip arthroplasty (THA) is thought to be 0.08; however, that was based on patients receiving accelerated rehabilitation and not based on patient satisfaction.<sup>18</sup> The MCID for the OKS has been defined as five points;<sup>19</sup> using the algorithm from the current study, a five-point change (without the constant (0.029)) would be associated with a 0.1 change on the EQ-5D-3L utility postoperatively. This value is consistent with the literature,<sup>12</sup> and could potentially be used to power studies until a MCID is established. On a similar theme, the minimal important change (MIC), a clinically significant change in the pre- to postoperative score, in the EQ-5D-3L utility is, to the authors' knowledge, not known either following TKA or THA. Using the same assumptions as for the MCID, the known MIC in the OKS is seven points, and this would equate with a 0.140 change in the EQ-5D-3L without the constant.<sup>20</sup>

TKA is one of the most cost-effective interventions available in medicine, with one of the lowest associated cost per QALY.<sup>5,21</sup> However, these data are based on relatively short- to medium-term evidence at one and five years' follow-up.<sup>5,21</sup> Jenkins et al<sup>5</sup> used one-year outcome data, then modeled costs (revision) and benefits (diminishing with time) over the patients' predicted life span. Using the mapping available from the current study, this may allow other author groups to follow up patient cohorts into the longer term, and potentially produce cost-utility analysis to support the cost-effectiveness of TKA when the EQ-5D-3L has not been collected primarily. This may also be applicable to assessing the cost-effectiveness of interventions such as robotic surgery on longer-term outcomes.<sup>22</sup>

The OKS is a versatile, validated, joint-specific PROM that can be used over the telephone and for retrospective use.<sup>23,24</sup> To be able to convert the OKS into an EQ-5D-3L utility will facilitate future cost analyses that may not have been possible. The simple conversion chart in Table II can be used to calculate the associated EQ-5D-3L utility pre- and postoperatively and for change in scores, although there is no substitute for prospective collection of EQ-5D-3L data and this should still be considered the gold standard.



In conclusion, the OKS can be used to estimate the preoperative, postoperative, and change in the EQ-5D-3L utility index. Predicted individual patient utility error beyond the MCID varied from one-third to two-thirds depending on timepoint assessed, but the mean for a cohort did not differ and could be employed for this purpose.



### Take home message

- The Oxford Knee Score can be used to estimate the preoperative, postoperative, and change in the EuroQol five-dimension three-level utility index.

### Twitter

Follow N. D. Clement @EdinOrthopaedic  
Follow I. Afzal @iafzal1  
Follow G. J. Macpherson @gjmapherson  
Follow V. Asopa @vipin\_asopa  
Follow D. F. Kader @DeiaryKader

### References

- 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.
- 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.
- Brooks R. EuroQol: the current state of play. *Health Policy.* 1996;37(1):53–72.
- Paulden M. Recent amendments to NICE's value-based assessment of health technologies: implicitly inequitable? *Expert Rev Pharmacoecon Outcomes Res.* 2017;17(3):239–242.
- Jenkins PJ, Clement ND, Hamilton DF, Gaston P, Patton JT, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement: A health economic analysis. *Bone Joint J.* 2013;95-B(1):115–121.
- Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ.* 2010;11(2):215–225.
- Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res.* 2013;22(3):683–694.
- Martín-Fernández J, Morey-Montalvo M, Tomás-García N, et al. Mapping analysis to predict EQ-5D-5L utility values based on the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) questionnaires in the Spanish population suffering from lower limb osteoarthritis. *Health Qual Life Outcomes.* 2020;18(1):184.
- Scott CEH, MacDonald DJ, Howie CR. "Worse than death" and waiting for a joint arthroplasty. *Bone Joint J.* 2019;101-B(8):941–950.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35(11):1095–1108.
- Clement ND, Scott CEH, Murray JRD, Howie CR, Deehan DJ, IMPACT-Restart Collaboration. The number of patients "worse than death" while waiting for a hip or knee arthroplasty has nearly doubled during the COVID-19 pandemic. *Bone Joint J.* 2021;103-B(4):672–680.
- Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: A critical review. *Expert Rev Pharmacoecon Outcomes Res.* 2014;14(2):221–233.
- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician.* 1983;32(3):307.
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191–2194.
- Vijayanathan A, Nawawi O. The importance of Good Clinical Practice guidelines and its role in clinical trials. *Biomed Imaging Interv J.* 2008;4(1):e5.
- Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Qual Life Res.* 2015;24(7):1775–1784.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163.
- Larsen K, Hansen TB, Søballe K. Hip arthroplasty patients benefit from accelerated perioperative care and rehabilitation: a quasi-experimental study of 98 patients. *Acta Orthop.* 2008;79(5):624–630.
- Clement ND, MacDonald D, Simpson A. 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):1933–1939.
- 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.
- Dakin H, Gray A, Fitzpatrick R, MacLennan G, Murray D, KAT Trial Group. Rationing of total knee replacement: a cost-effectiveness analysis on a large trial data set. *BMJ Open.* 2012;2(1):e000332.
- Clement ND, Deehan DJ, Patton JT. Robot-assisted unicompartmental knee arthroplasty for patients with isolated medial compartment osteoarthritis is cost-effective: a markov decision analysis. *Bone Joint J.* 2019;101-B(9):1063–1070.
- Makaram N, Lee T, Macdonald D, Clement ND. The verbal Oxford Knee Score is not clinically different from the written score when assessed before or after total knee arthroplasty. *Knee.* 2020;27(5):1396–1405.
- Yeoman TFM, Clement ND, Macdonald D, Moran M. Recall of preoperative Oxford Hip and Knee Scores one year after arthroplasty is an alternative and reliable technique when used for a cohort of patients. *Bone Joint Res.* 2018;7(5):351–356.

#### Author information:

- N. D. Clement, MD, PhD, FRCS (Tr&Orth), Honorary Research Consultant, Orthopaedic Consultant, South West London Elective Orthopaedic Centre, Epsom, UK; Edinburgh Orthopaedics, Royal Infirmary of Edinburgh, Edinburgh, UK.
- I. Afzal, MSc, Research & Outcomes Manager
- C. J. H. Peacock, BMBS, BMedSci, Junior Clinical Research Fellow in Trauma & Orthopaedics
- V. Asopa, PhD, FRCS (Tr&Orth), Orthopaedic Consultant
- D. H. Sochart, FRCS (Tr&Orth), Orthopaedic Consultant
- D. F. Kader, FRCS (Tr&Orth), MFS EM (UK), Acting Director of Research and Orthopaedic Consultant  
South West London Elective Orthopaedic Centre, Epsom, UK.
- D. MacDonald, BA, Clinical Research and Database Manager, Department of Orthopaedics, University of Edinburgh, Edinburgh, UK.
- G. J. Macpherson, FRCS (Tr&Orth), Orthopaedic Consultant, Edinburgh Orthopaedics, Royal Infirmary of Edinburgh, Edinburgh, UK; Department of Orthopaedics, University of Edinburgh, Edinburgh, UK.
- J. T. Patton, FRCS (Tr&Orth), Orthopaedic Consultant, Edinburgh Orthopaedics, Royal Infirmary of Edinburgh, Edinburgh, UK.

#### Author contributions:

- N. D. Clement: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Writing – review & editing.
- I. Afzal: Data curation, Writing – review & editing.
- C. J. H. Peacock: Writing – review & editing.
- D. MacDonald: Data curation.
- G. Macpherson: Supervision, Writing – review & editing.
- J. T. Patton: Supervision, Writing – review & editing.
- V. Asopa: Supervision, Writing – review & editing.
- D. H. Sochart: Supervision, Writing – review & editing.
- D. Kader: Conceptualization, Resources, Supervision, Writing – review & editing.

#### Funding statement:

- The authors received no financial or material support for the research, authorship, and/or publication of this article.

#### Acknowledgements:

- South West London Elective Orthopaedic Centre, Epsom, UK.  
Edinburgh Orthopaedics, Royal Infirmary of Edinburgh Edinburgh, UK.

#### Ethical review statement:

- There was no additional patient contact and as such this project was performed as a service evaluation without the need for formal ethical approval.

#### Open access funding

- The open access fee was funded by the South West London Elective Orthopaedic Centre.

© 2022 Author(s) et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>