Meaningful values of the EQ-5D-3L in patients undergoing primary knee arthroplasty

L. Z. Yapp, C. E. H. Scott, C. R. Howie, D. J. MacDonald, A. H. R. W. Simpson, N. D. Clement
From Royal Infirmary of Edinburgh, Edinburgh, UK

Aims
The aim of this study was to report the meaningful values of the EuroQol five-dimension three-level questionnaire (EQ-5D-3L) and EuroQol visual analogue scale (EQ-VAS) in patients undergoing primary knee arthroplasty (KA).

Methods
This is a retrospective study of patients undergoing primary KA for osteoarthritis in a university teaching hospital (Royal Infirmary of Edinburgh) (1 January 2013 to 31 December 2019). Pre- and postoperative (one-year) data were prospectively collected for 3,181 patients (median age 69.9 years (interquartile range (IQR) 64.2 to 76.1); females, n = 1,745 (54.9%); median BMI 30.1 kg/m² (IQR 26.6 to 34.2)). The reliability of the EQ-5D-3L was measured using Cronbach’s alpha. Responsiveness was determined by calculating the anchor-based minimal clinically important difference (MCID), the minimal important change (MIC) (cohort and individual), the patient-acceptable symptom state (PASS) predictive of satisfaction, and the minimal detectable change at 90% confidence intervals (MDC-90).

Results
The EQ-5D-3L demonstrated good internal consistency with an overall Cronbach alpha of 0.75 (preoperative) and 0.88 (postoperative), respectively. The MCID for the Index score was 0.085 (95% confidence interval (CI) 0.042 to 0.127) and EQ-VAS was 6.41 (95% CI 3.497 to 9.323). The MIC COHORT was 0.289 for the EQ-5D and 5.27 for the EQ-VAS. However, the MIC INDIVIDUAL for both the EQ-5D-3L Index (0.105) and EQ-VAS (-1) demonstrated poor-to-acceptable reliability. The MDC-90 was 0.023 for the EQ-5D-3L Index and 1.0 for the EQ-VAS. The PASS for the postoperative EQ-5D-3L Index and EQ-VAS scores predictive of patient satisfaction were 0.708 and 77.0, respectively.

Conclusion
The meaningful values of the EQ-5D-3L Index and EQ-VAS scores can be used to measure clinically relevant changes in health-related quality of life in patients undergoing primary KA.

Cite this article: Bone Joint Res 2022;11(9):619–628.

Keywords: EQ-5D, Quality of life, Knee, Osteoarthritis, Arthroplasty

Article focus
To report the meaningful values of the EuroQol five-dimension three-level questionnaire (EQ-5D-3L) Index and EuroQol visual analogue scale (EQ-VAS) scores in patients undergoing primary knee arthroplasty.

Key messages
The EQ-5D-3L was found to be a reliable measure of health-related quality of life in patients undergoing primary knee arthroplasty (KA), with high levels of internal consistency observed.

The minimal clinically important difference for the EQ-5D-3L Index score was 0.085 (95% confidence interval (CI) 0.042 to 0.127) and for the EQ-VAS score was 6.41 (95% CI 3.497 to 9.323).

The patient-acceptable symptom state predictive of postoperative satisfaction
was 0.708 for the EQ-5D-3L Index score and 77.0 for the EQ-VAS.

**Strengths and limitations**

- Comprehensive summary of the meaningful values of the EQ-5D-3L in patients undergoing primary KA for osteoarthritis.
- Focusing on the summary values of the EQ-5D-3L such as the Index and EQ-VAS scores may obscure useful information derived from the health profile data.

**Introduction**

Healthcare systems have finite resources, which necessitates prioritization. It is therefore expected that competing interventions demonstrate effectiveness in terms of patient outcomes and costs to the health service. The quality-adjusted life year (QALY) is a measure designed to capture the impact of a treatment on a patient’s duration of life and their associated health-related quality of life (HRQoL).1 QALYs are generated using health utilities, often termed HRQoL weights, which can equate to desirability or value.

The National Institute for Health and Care Excellence (NICE) currently recommends using the EuroQol five-dimension questionnaire (EQ-SD) to measure HRQoL when assessing an intervention’s cost-effectiveness (cost-utility analysis).2 The EQ-SD – developed by the EuroQol Group – is a widely used measure of HRQoL.3 It is weighted according to the relative importance that specific populations place on different types of health problems. The EQ-SD is often described as ‘generic’ because it is used to measure HRQoL in a way that can be compared across different patient populations, diseases, and treatments. The UK version of the EQ-SD generates a numerical score (index) and follows the HRQoL convention whereby 1 is considered full health and 0 death, although negative scores are possible with a scoring range of -0.594 to 1.

Two versions of the EQ-SD score are available: with three (3 L) or five (5 L) levels of severity. Critics of the EuroQol five-dimension three-level questionnaire (EQ-SD-3L) argue that it is prone to ceiling effects and is therefore less likely to identify clinically meaningful changes in specific conditions.4 A newer version of the EQ-SD, the EQ-SD-5L, was developed to tackle these issues and is thought to be more sensitive than the EQ-SD-3L.5 However, concerns regarding the 3L-value sets for the UK have limited its use.6,7 Current NICE guidance allows the EQ-SD-5L to be used in reference cases, but states that utilities should be mapped to the original EQ-SD-3L.8

There is increasing scrutiny regarding the evidence base for common orthopaedic procedures, such as primary knee arthroplasty (KA).10 It is essential that clinically relevant differences between a cohort, or occurring in an individual patient, can be accurately measured. A substantial proportion of patients awaiting total knee arthroplasty (TKA) have been shown to be living in a health state “worse than death” (WTD) with negative EQ-SD indices.11 Longer waiting times, due to healthcare service disruption following the global COVID-19 pandemic,12,13 have led to further measurable deterioration in these patients’ HRQoL.14 Measures of the EQ-SD-3L’s responsiveness, such as minimal clinically important difference (MCID), minimal important change (MIC), minimal detectable change (MDC), and the patient-acceptable symptom state (PASS), are poorly defined in patients undergoing primary KA. The EQ-SD-3L is commonly used to measure HRQoL and to perform cost-utility analyses in orthopaedic surgery. Therefore, it is fundamentally important that clinically meaningful values of the EQ-SD are identified in this patient population. The aim of this study was to report the meaningful values of the EQ-5D-3L in patients undergoing primary KA.

**Methods**

This is a retrospective study of patients in a large university teaching hospital (Royal Infirmary of Edinburgh) during the period 1 January 2013 to 31 December 2019. Ethical approval was obtained from the Scotland (A) Research Ethics Committee (16/SS/0026). This study is reported in accordance with the CONsensus-based Standards for the selection of health status Measurement Instruments (COSMIN) guidelines for studies on measurement properties of patient-reported outcome measures (PROMs).15

**Data collection and questionnaire administration.** Data were collected prospectively and stored in an electronic research database. Preoperatively, patients completed standardized questionnaires upon attendance at the preadmission clinic. Postoperatively, data were collected at one-year following surgery via a postal questionnaire.

**Participants.** Patients were included in the study if they were undergoing primary KA (either unicondylar or TKA) for osteoarthritis (OA) during the study period. Patients who did not respond to the questionnaire, had incomplete data, or who underwent surgery for non-OA indications were excluded for the purposes of this study.

During the period of follow-up there were 4,485 patients undergoing primary KA, of whom 3,181 (70.9%) satisfied the inclusion and exclusion criteria of the study (median age 69.9 years (interquartile range (IQR) 64.2 to 76.1); females, 1,745 (54.9%); median BMI 30.1 kg/m2 (IQR 26.6 to 34.2)).

**EQ-SD-3L.** The EQ-SD-3L is broadly separated into two elements: the first is the EQ-SD health profile in which respondents answer five questions related to the dimensions: Mobility (MO), Self-Care (SC), Usual Activities (UA), Pain and Discomfort (PD), and Anxiety and Depression (AD).

In this version of the EQ-SD, there are three possible levels indicating the degree of impairment (“None” = 1, “Some” = 2, “Extreme” = 3) leading to 243 (3^5) potential health states. The EQ-SD health state data can be summarized in a variety of ways,16,17 however the most common is the use of the Index score. The EQ-SD-3L Index score summarizes each possible health state on a numerical scale ranging from -0.594 to 1, where a score of 1 indicates full health, and scores of 0 and less than 0 indicate a state equivalent to being dead and a state WTD, respectively. The Index score is created by applying value sets to the health state data, typically based on the health utility weights.
upon the country where respondents are based. This is based upon the relative importance a defined population places on different health problems. The use of preference-based weights enables the general population’s views on health problems to be accounted for and accepts that variation exists between countries. As this study was performed in Scotland, we used the ‘Time Trade Off’ (TTO) value set for the UK. The TTO is a direct method used for generating HRQoL weights, based upon the value that individuals place on different health states.1 The TTO method presents patients with two alternative health states and asks which they would prefer. This method assesses how much time members of the general population would be willing to sacrifice from their life in order to avoid an impaired health state.

The second element of the EQ-SD is a visual analogue scale (VAS), also called the EQ-VAS, on which respondents are asked to rate their overall health from 0 (worst health imaginable) to 100 (best health imaginable). The EQ-VAS is designed to encompass an individual’s health beyond the five dimensions covered in the health profile.

**Reliability: internal consistency.** Internal consistency measures how items within a questionnaire are correlated. This is particularly relevant for outcome scores measuring a single underlying construct, such as HRQoL.18 Internal consistency was measured using Cronbach’s alpha. Cronbach’s alpha is scored between 0 and 1, with higher numbers reflecting increasing correlation. Good internal consistency was considered to be a Cronbach’s alpha between 0.70 to 0.95.18

**Responsiveness.** Responsiveness reflects the sensitivity or ability of a PROM to detect clinically important changes over time.18,20 A clinically important, or meaningful, change can be described according to the individual patient or the cohort as a whole.

**Effect size: standardized response mean and standardized effect size.** The standardized response mean (SRM) was used to determine the effect size (ES) in independent and paired data. For independent data, the SRM is the ratio of the mean change score and the standard deviation (SD) of the change score.21 For the paired data, SRM was adjusted using the below formula:22,23

\[
\text{SRM}_{\text{Paired}} = \left( \frac{\text{Mean change score}}{\text{SD change score}} \right) / \left( \sqrt{2} \times \sqrt{1-r} \right)
\]

where \( r \) is the correlation coefficient between the pre- and postoperative scores, and the \( \sqrt{2} \) is used to account for the number of measurements.

The standardized effect size (SES) was calculated for each transition level of the anchor question by dividing the change in mean score by the SD of the preoperative score. The greater the SES, the greater the difference between groups and therefore responsiveness. The ES can be interpreted using Cohen’s thresholds:23,24 < 0.20 (Trivial); 0.20 to 0.49 (Small); 0.50 to 0.79 (Medium); and 0.80 to 1.00 (Large).

**Floor and ceiling effects.** Floor and ceiling effects occur when more than 15% of patients achieve the minimum or maximum score, respectively.18 When present, this results in an inability to discriminate between subjects at extremes of the scale. Floor and ceiling effects suggest limited content validity, reduced overall reliability, and impaired responsiveness.18

**Minimal clinically important difference.** The MCID is the smallest change in an outcome measure which patients perceive to be clinically relevant.25 The MCID can be defined using anchor-based or distribution-based methodology. We

### Table I. Descriptive summary of EuroQol five-dimension three-level questionnaire components.

| Component          | Level | Preoperative | Postoperative | Change | p-value *   |
|--------------------|-------|--------------|---------------|--------|-------------|
|                    | N     | %            | N             | %      | N           | %     |
| Mobility           | 1     | 296          | 9.3           | 1,774  | 55.8        | 1,478 | 46.5 | < 0.001 |
|                    | 2     | 2,881        | 90.6          | 1,403  | 44.1        | -1,478 | -46.5 |
|                    | 3     | 4            | 0.1           | 4      | 0.1         | 0     | 0.0 |
| Problems           |       |              |               | 1,407  | 44.2        | -1,478 | -46.5 |
| Self-care          | 1     | 2,517        | 79.1          | 2,632  | 82.7        | 113   | 3.6  | < 0.001 |
|                    | 2     | 653          | 20.5          | 528    | 16.6        | -125  | -3.9  |
|                    | 3     | 11           | 0.4           | 21     | 0.7         | 10    | 0.3  |
| Problems           |       |              |               | 549    | 17.3        | -115  | -3.6  |
| Usual activities   | 1     | 576          | 18.1          | 1,630  | 51.2        | 1,054 | 33.1 | < 0.001 |
|                    | 2     | 2,285        | 71.8          | 1,441  | 45.3        | -844  | -26.5 |
|                    | 3     | 320          | 10.0          | 110    | 3.5         | -210  | -6.6  |
| Problems           |       |              |               | 1,551  | 48.8        | -1,054 | -33.1 |
| Pain/discomfort    | 1     | 27           | 0.8           | 1,267  | 39.8        | 1,240 | 39.0 | < 0.001 |
|                    | 2     | 1,915        | 60.2          | 1,750  | 55.0        | -165  | -5.2  |
|                    | 3     | 1,239        | 39.0          | 164    | 5.2         | -1,075 | -33.8 |
| Problems           |       |              |               | 1,914  | 60.2        | -1,240 | -39.0 |
| Anxiety/depression | 1     | 2,137        | 67.2          | 2,535  | 79.7        | 398   | 12.5 | < 0.001 |
|                    | 2     | 956          | 30.0          | 593    | 18.6        | -363  | -11.4 |
|                    | 3     | 88           | 2.8           | 53     | 1.7         | -35   | -1.1  |
| Problems           |       |              |               | 646    | 20.3        | -398  | -12.5 |

*Chi-squared test.
chose to define the EQ-5D-3L MCID using anchor-based methods, as it is not affected by the statistical characteristics of the sample and directly incorporates the patient’s perspective.

The questionnaire assessed the patient’s global rating of change by using an anchor question assessing the patient’s satisfaction following surgery (“How satisfied are you with your operated knee?”), and responses were recorded using a five-point Likert scale: “Very satisfied”, “Satisfied”, “Neither satisfied nor dissatisfied”, “Dissatisfied”, and “Very dissatisfied”. We assessed the credibility of the anchor question for calculating the MCID using the guidelines described by Devji et al. The responses to the five-point Satisfaction anchor question were also grouped into a secondary binary classification of ‘Satisfied’ (Very Satisfied and Satisfied) and ‘Dissatisfied’ (Neither satisfied nor dissatisfied, Dissatisfied, and Very dissatisfied) patients.

The MCID was defined as the difference in the mean change within one year for the EQ-5D-3L Index and the EQ-VAS scores in patients responding “Neither satisfied nor dissatisfied” compared to those responding “Satisfied”. The MCID is reported with 95% confidence intervals (CIs) of the difference.

Minimal important change. The MIC for the cohort (MIC\textsubscript{COHORT}) was defined as the change, relative to preoperative scores, for those patients declaring themselves as ‘Satisfied’. Clinically relevant improvement for an individual patient (MIC\textsubscript{INDIVIDUAL}) predictive of satisfaction (‘Satisfied’ vs ‘Dissatisfied’) was estimated using receiver operating characteristic (ROC) curve analysis. Patient-acceptable symptom state. The PASS is the postoperative score which can predict a patient declaring their outcome as ‘Satisfied’. ROC curve analysis was also used to determine the optimal PASS threshold value.

Minimal detectable change. The MDC, also known as the smallest real difference, is a distribution-based method which represents the smallest change beyond the measurement error of the EQ-5D-3L. The MDC-90 was calculated by multiplying the standard error of measurement (SEM) by the square root of two (to account for measurement on two occasions) and by a z score giving a 90% confidence level (1.65).

Statistical analysis. All data handling, cleaning, and statistical analysis was undertaken using RStudio version 1.3.959 (RStudio, USA). The distribution of continuous variables was plotted to assess appropriateness of parametric or non-parametric tests of differences. Differences between pre- and postoperative EQ-5D-3L Index scores were compared using two-sided paired Wilcoxon signed-rank test. Differences between the change in EQ-5D-3L Index scores were measured using independent-samples two-sided \( t \)-test. Differences between categorical
Table II. Summary of the EuroQol five-dimension three-level questionnaire Index scores for the cohort.

| Cohort                  | n      | Preoperative | Postoperative | Change | p-value | SRMP | SRMI | SES | Preoperative (%) | Postoperative (%) |
|-------------------------|--------|--------------|---------------|--------|---------|------|------|-----|------------------|-------------------|
|                         | Mean (SD) | Median (IQR) | Range        | Mean (SD) | Median (IQR) | Range        | Mean (SD) | Range | Floor | Ceiling | Floor | Ceiling |
| Total                   | 3,181  | 0.426 (0.31) | 0.59 (0.69)   | -0.594 (0.016)  | 1.0 | 0.748 (0.26) | 0.796 (0.69)  | -0.429 (0.10)  | 1.0 | 0.322 (0.53) | -1.049 (1.349) |
| Very satisfied          | 1,763  | 0.447 (0.30) | 0.620 (0.69)  | -0.394 (0.016)  | 1.0 | 0.849 (0.19) | 0.883 (0.76)  | -0.239 (0.10)  | 1.0 | 0.402 (0.31) | -0.707 (1.349) |
| Satisfied               | 904    | 0.413 (0.30) | 0.587 (0.69)  | -0.240 (0.010)  | 1.0 | 0.702 (0.21) | 0.727 (0.62)  | -0.349 (0.01)  | 1.0 | 0.289 (0.31) | -0.642 (1.181) |
| Neither satisfied nor   | 285    | 0.364 (0.32) | 0.36 (0.69)   | -0.594 (0.016)  | 1.0 | 0.569 (0.25) | 0.620 (0.52)  | -0.429 (0.10)  | 1.0 | 0.205 (0.33) | -1.049 (1.079) |
| Dissatisfied            | 176    | 0.379 (0.31) | 0.516 (0.69)  | -0.349 (0.009)  | 1.0 | 0.441 (0.32) | 0.620 (0.16)  | -0.349 (0.0)   | 1.0 | 0.061 (0.34) | -0.820 (0.636) |
| Very dissatisfied       | 53     | 0.384 (0.31) | 0.440 (0.69)  | -0.077 (0.009)  | 1.0 | 0.263 (0.356) | 0.225 (0.02)  | -0.429 (0.0)   | 1.0 | -0.121 (0.38) | -1.035 (0.690) |

*Paired Wilcoxon signed-rank test.
EQ-SD-3L, EuroQol five-dimension three-level questionnaire; IQR, interquartile range; SD, standard deviation; SES, standardized effect size; SRM, standardized response mean.

Table III. Summary of the EuroQol visual analogue scale scores for the cohort.

| Cohort                  | n      | Preoperative | Postoperative | Change | p-value | SRMP | SRMI | SES | Preoperative (%) | Postoperative (%) |
|-------------------------|--------|--------------|---------------|--------|---------|------|------|-----|------------------|-------------------|
|                         | Mean (SD) | Median (IQR) | Range        | Mean (SD) | Median (IQR) | Range        | Mean (SD) | Range | Floor | Ceiling | Floor | Ceiling |
| Total                   | 3,181  | 70.4 (19.7)  | 72.0 (85.0)   | 0.0 to 100 | 80.2 (90.1) | 0.0 to 100 | 6.7 (22.9)  | -100.0 | 1.92 | 0.32 | 4.83 |
| Very satisfied          | 1,763  | 72.1 (19.3)  | 76.9 (89.0)   | 0.0 to 100 | 89.6 (92.0) | 0.0 to 100 | 10.6 (22.0) | -100.0 | 0.00 | 0.00 | 0.00 |
| Satisfied               | 904    | 69.3 (19.8)  | 70.3 (82.0)   | 0.0 to 100 | 80.0 (90.0) | 0.0 to 100 | 5.3 (21.5)  | -99.0 | 0.00 | 0.00 | 0.00 |
| Neither satisfied nor   | 285    | 66.9 (20.3)  | 70.0 (80.0)   | 0.0 to 100 | 70.0 (90.0) | 0.0 to 100 | -1.1 (22.0) | -60.7 | 0.389 | 0.00 | 0.00 |
| Dissatisfied            | 176    | 67.7 (19.9)  | 70.0 (80.0)   | 0.0 to 100 | 61.0 (79.9) | 0.0 to 100 | -1.6 (25.2) | -80.7 | 0.00 | 0.00 | 0.00 |
| Very dissatisfied       | 53     | 65.8 (22.4)  | 70.9 (85.2)   | 9.0 to 98.0 | 60.0 (39.5) | 0.0 to 99.0 | -8.8 (30.6) | -78.0 | 0.043 | 0.00 | 0.00 |

*Paired Wilcoxon signed-rank test.
EQ-SD-3L, EuroQol five-dimension three-level questionnaire; EQ-VAS, EuroQol visual analogue scale; IQR, interquartile range; SD, standard deviation; SES, standardized effect size; SRM, standardized response mean.
Of these patients, the proportion of “Extreme” responses was highest in the PD domain (“I have extreme pain or discomfort”, n = 1,239 (39.0%)), with few respondents reporting an inability to perform their usual activities (UA level 3 (n = 320 (10.0%))) or confinement to bed (MO level 3 (n = 4 (0.1%))). Significant overall improvements in all domains were seen at one-year postoperatively, with the greatest levels of improvement seen in the PD, MO, and UA domains (Table I).

The EQ-5D-3L Index and EQ-VAS scores demonstrated a non-normal distribution pre- and postoperative (Figures 1 and 2). The overall median EQ-5D-3L Index (median preoperative 0.590 (IQR 0.16 to 0.69) vs 0.796 (IQR 0.69 to 1.0); p < 0.001, Wilcoxon signed-rank test) and EQ-VAS (median preoperative 72.0 (IQR 60.0 to 85.0) vs 80.2 (70 to 90.1); p < 0.001, Wilcoxon signed-rank test) scores significantly increased (improved) at one-year postoperatively.

The majority of patients considered themselves to be either ‘Very Satisfied’ or ‘Satisfied’ (n = 2,667 (83.8%)). The five-level Satisfaction anchor was significantly associated with mean postoperative change in the Index and EQ-VAS scores (Tables II and III). The direction and magnitude of change was associated with outcome; the highest mean change was seen in patients who considered themselves ‘Very Satisfied’ and ‘Satisfied’ while the lowest mean change was seen in patients who considered themselves ‘Neither satisfied nor dissatisfied’, ‘Dissatisfied’, and ‘Very dissatisfied’. Internal consistency. The EQ-5D-3L demonstrated good internal consistency with an overall Cronbach alpha of

**Table IV.** Cronbach alpha values for the EuroQol five-dimensional three-level questionnaire.

| EQ-5D-3L            | Preoperative | Postoperative |
|---------------------|--------------|---------------|
| EQ-5D-3L Index      | 0.750        | 0.884         |
| Mobility            | 0.760        | 0.867         |
| Self-care           | 0.727        | 0.878         |
| Usual activities    | 0.724        | 0.862         |
| Pain/discomfort     | 0.701        | 0.867         |
| Anxiety/depression  | 0.735        | 0.881         |
| EQ-VAS              | 0.733        | 0.877         |

EQ-5D-3L, EuroQol five-dimensional three-level questionnaire; EQ-VAS, EuroQol visual analogue scale.
0.75 (preoperative) and 0.88 (postoperative), respectively (Table IV).

Responsiveness. The EQ-5D-3L Index score showed good overall responsiveness, with ‘Large’ ESs seen across the SRM_{Paired}, SRM_{Independent}, and SES (Table II). In comparison, the EQ-VAS demonstrated only ‘Small’ ESs, suggesting it is less sensitive to change. However, the ESs differed greatly across the 5-level transition question (Table III).

There was no evidence of a floor or ceiling effect from the preoperative EQ-5D-3L Index and EQ-VAS scores. The postoperative EQ-5D-3L Index score demonstrated evidence of a ceiling effect of 29.8%, but had no evidence of a floor effect (0.06%) (Table II). The postoperative EQ-VAS demonstrated no evidence of a floor (0.32%) or ceiling effect (4.83%) (Table III).

Mean improvement in the EQ-5D-3L Index and EQ-VAS scores for those patients declaring themselves as ‘Neither satisfied nor dissatisfied’ (n = 285) and those who reported themselves as ‘Satisfied’ (n = 904) was used to determine the MCID and MIC for a cohort. The MCID for the EQ-5D-3L Index was 0.085 (95% CI 0.042 to 0.127) and EQ-VAS was 6.41 (3.497 to 9.323), respectively. The MIC for a cohort was 0.289 for the EQ-5D and 5.27 for the EQ-VAS. However, the MIC for both the Index and EQ-VAS scores demonstrated poor-to-acceptable levels of reliability (Figure 3) (Table V). The MDC-90 was 0.023 for the EQ-5D-3L Index and 1.0 for the EQ-VAS.

The PASS for the postoperative EQ-5D-3L Index and EQ-VAS that were predictive of patient satisfaction were 0.708 and 77.0, respectively (Figure 4). The Index score demonstrated ‘Good’ levels of prediction, whereas the EQ-VAS was only ‘Acceptable’.

Discussion
This study reports meaningful values for the EQ-5D-3L Index and EQ-VAS scores for patients undergoing primary KA for OA. This can be used to measure MCID between groups, the MIC for a cohort or an individual patient, and a postoperative score that is predictive of patient satisfaction. The relative strengths of the EQ-5D-3L are that it is easily understood by respondents, simple to complete, and incorporates the patient’s values through preference-based weighting. In the current study, it was found to be a reliable measure of HRQoL in patients undergoing primary KA with high levels of internal consistency observed.

The EQ-5D-3L Index score demonstrated a ‘Large’ ES, suggesting that it is a responsive measure in this patient population. In comparison, the EQ-VAS demonstrated ‘Small’ ESs and appeared to be less sensitive to small changes. These findings are consistent with those reported by Shim and Hamilton,\(^34\) and may simply reflect that the EQ-VAS is a broader measure of the patient’s health. The EQ-VAS declines with increasing age and is worsened by the presence of ‘problems’ in the self-reported health state.\(^16\,\,35\) Previous studies have reported that patients reporting perfect health via the EQ-5D-3L Index score frequently do not report perfect health in the EQ-VAS.\(^36\) Furthermore, it has been shown that the EQ-5D-3L domain Anxiety & Depression has a stronger correlation with EQ-VAS scores, whereas Pain & Disability had the smallest.\(^16\) These findings may limit the utility of the EQ-VAS in the patients undergoing primary KA.

The Index score demonstrated a large postoperative ceiling effect, affecting approximately one in three respondents. Although previous studies of patients undergoing TKA have also demonstrated large ceiling effects,\(^37\) our findings are far smaller than the 84% reported by Giesinger et al.\(^4\) These differences may reflect the fact that both studies were performed in different countries with their own preference weights and values.

In the current study, ceiling effects were observed primarily in patients who considered themselves ‘Very Satisfied’. This is problematic because this subgroup comprised more than half of the cohort. Large ceiling effects limit both the responsiveness and reliability of a PROM, and could potentially lead to underestimation of

| Responsiveness measure | EQ-5D-3L | EQ-VAS |
|-------------------------|---------|--------|
| Index MCID (95% CI)     | 0.085 (0.042 to 0.127) | 6.41 (3.497 to 9.323) |
| MIC_Cohort              | 0.289   | 5.27   |
| MIC_Individual          | 0.105   | -1     |
| AUC (95% CI)            | 0.702 (0.675 to 0.729) | 0.662 (0.634 to 0.689) |
| Sensitivity, %          | 78.0    | 76.1   |
| Specificity, %          | 58.8    | 50.8   |
| PASS                    | 0.708   | 77     |
| AUC (95% CI)            | 0.839 (0.824 to 0.856) | 0.750 (0.727 to 0.773) |
| Sensitivity, %          | 71.7    | 76.1   |
| Specificity, %          | 86.1    | 50.8   |
| MDC-90                  | 0.023   | 0.96   |

AUC, area under the curve; CI, confidence interval; EQ-5D-3L, EuroQol five-dimension three-level questionnaire; EQ-VAS, EuroQol visual analogue scale; MCID, minimal clinically important difference; MDC-90, minimal detectable change at 90% confidence intervals; MIC, minimal important change; PASS, patient-acceptable symptom state.
benefit in this patient group. Changes to the EQ-SD, incorporated in the five-level version, have led to considerably lower ceiling effects and may be more sensitive to changes in these patients undergoing primary KA.

There are few estimates for the MCID of the EQ-SD-3L Index, and none for the EQ-VAS, following KA. Walters and Brazier used anchor-based methods using a five-point global rating of change question (“Compared to one year ago, how would you rate your health in general now?”) to calculate a minimal important difference (MID) of 0.121 in 149 patients undergoing TKA for OA. However, this MID may be overestimated due to small sample size, and their calculations were based on differences in the mean change between patients who responded “Somewhat better” and “Somewhat worse”, rather than those who responded “About the same”. Kang used NHS PROMs data to examine the responsiveness of the EQ-SD-3L in 191,379 patients undergoing primary KA at six months post-surgery. Although NHS PROMs data have a follow-up rate of 51.3%, potentially resulting in a response bias, the anchor-based MCID of 0.09 is consistent with the current study’s estimated MCID of 0.085 (95% CI 0.042 to 0.127).

A limitation of the current study is related to the calculation of ‘meaningful’ values and the EQ-SD-3L. The EQ-SD-3L Index scores already encompass a measure of ‘importance’ as defined by a population of patients, based on their preferences for various health states. Therefore, it could be argued that any difference, no matter how small, in the underlying Index value is clinically meaningful, as this reflects the values of the person affected.

A second limitation relates to the methodology of MCID estimation. Anchor-based MCIDs rely on a patient’s self-rating of change. This requires the patient to retrospectively assess the change in their health between two timepoints and may be subject to recall bias. In addition, the wording of the transition question can influence the way patients respond. Clement et al demonstrated that the rate of patient satisfaction following TKA could be influenced by the focus of the question.

Third, focusing on the Index and EQ-VAS scores may obscure useful information derived from the EQ-SD-3L. Examining the health profile data, the current study observed that patients were primarily affected in the domains Pain and Discomfort.
Activities, and Mobility. However, while 90.7% (n = 2,885) of patients recorded problems with Mobility, less than 1% (n = 4) recorded the most extreme level of problems with this domain (“I am confined to bed”). Similar limitations in the response categories for Mobility have previously been identified in patients undergoing total hip arthroplasty. This has implications for how patients undergoing primary KA can describe their health using the EQ-SD-3L, and suggests that only small improvements in the Mobility domain are likely to be detected.

Finally, the pre- and postoperative EQ-SD-3L Index scores demonstrated a non-normal distribution (Figures 1 and 2). This appearance has been noted previously in different health conditions and patient populations summarized by the EQ-5D-3L. This has led to concerns that this appearance is a construct issue of the EQ-5D-3L classification system. The presence, and non, concluding that these patterns arose as a result of creating the appearance of two clusters of index scores.42 This appearance has been noted previously in patients undergoing total hip arthroplasty.43 This has implications for how patients undergoing primary KA can describe their health using the EQ-5D-3L, and suggests that only small improvements in the Mobility domain are likely to be detected.

In conclusion, this study reports the clinically meaningful values of the EQ-SD-3L Index and EQ-VAS scores in patients undergoing primary KA for OA. This can be used to measure clinically relevant differences between groups, in individual patients, and may be used to target postoperative scores strongly associated with patient satisfaction. The EQ-SD-3L was found to be reliable and the Index score demonstrated large ESs. However, clinicians should be aware of the presence of postoperative ceiling effects and limitations in the sensitivity of measuring change in specific domains, such as mobility, which could lead to underestimation of benefit in patients undergoing primary KA.

**References**

1. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med J. 2010;341:c6119–21.
2. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: A systematic review, statistical modelling and survey. Health Technol Assess. 2014;18(9).1–224.
3. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):198–208.
4. Giesinger K, Hamilton DF, Jost B, Holzner B, Giesinger JM. Comparative responsiveness of outcome measures for total knee arthroplasty. Osteoarthritis Cartilage. 2014;22(2):184–189.
5. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-SD (EQ-SD-5L). Qual Life Res. 2011;20(10):1727–1736.
6. Hernandez Alava M, Wailoo A, Grim C, et al. EQ-SD-5L versus EQ-SD-3L: the impact on cost effectiveness in the United Kingdom. Value Health. 2018;21(1):49–56.
7. Pennington B, Hernandez-Alava M, Pudney S, Wailoo A. The impact of moving from EQ-SD-3L to -5L in NICE technology appraisals. Pharmacoeconomics. 2019;37(1):75–84.
8. No authors listed. Guide to the methods of technology appraisal 2013. National Institute for Health and Care (NICE). 2013. https://www.nice.org.uk/process/pmg9/
9. No authors listed. Position statement on use of the EQ-5D-5L value set for England (updated October 2019). National Institute for Health and Care Excellence (NICE). 2019. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/
10. Blom AW, Donovan RL, Beswick AD, Whitehouse MR, Kunutsor SK. Common elective orthopaedic procedures and their clinical effectiveness: umbrella review of level 1 evidence. BMJ. 2021;374:n1511.
11. Scott CEH, MacDonald DJ, Howie CR. “Worse than death” and waiting for a joint arthroplasty. Bone Joint J. 2019;101-B(9):941–950.
12. Yapp LZ, Clarke JV, Moran M, Simpson A, Scott CEH. National operating volume for primary hip and knee arthroplasty in the COVID-19 era: A study utilising the Scottish arthroplasty project dataset. Bone Joint Open. 2021;3(2):203–210.
13. Oussedik S, Maclntyre S, Gray J, McMeekin P, Clement ND, Deehan DJ. Elective orthopaedic cancellations due to the COVID-19 pandemic: where are we now, and where are we heading? Bone Joint Open. 2021;3(2):103–110.
14. 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.
15. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021;30(8):2197–2218.
16. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data [Internet]. Cham: Springer, 2020.
17. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. Health Econ. 2012;19(8):886–905.
18. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
19. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–745.
20. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: A clarification. J Clin Epidemiol. 1998;51(25):403–408.
21. Liang MH, Fosse AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care. 1990;28(7):632–642.
22. Kang S. Assessing responsiveness of the EQ-5D-3L, the Oxford Hip Score, and the Oxford Knee Score in the NHS patient-reported outcome measures. J Orthop Surg Res. 2021;16(1):1–12.
23. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care. 2002;2(1):16.
24. Cohen J. Statistical Power Analysis for the Behavioural Sciences. Second ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1988.
25. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407–415.
26. Rolfe S, Bohl E, Franklin P, et al. Patient-reported outcome measures in arthroplasty registries Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II: Recommendations for selection, administration, and analysis. Acta Orthop. 2018;87 Suppl 1(362):9–23.
27. Devji T, Carrasco-Labra A, Gass M, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ. 2020;369:m3714.
28. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005;64(1):29–33.
29. 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.
The focus of the question: A standard satisfaction question is required. and predictors of patient satisfaction after total knee arthroplasty are influenced by specific outcome measures for people with osteoarthritis of the knee. Rheumatology, Value Health and Oxford Hip Score: implications for mapping. Med Decis Making, april-2019-to-march-2020 (date last accessed 1 August 2022).

Author information:

The authors thank all orthopaedic surgeons whose patients were included in this study.

Acknowledgements:

The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: financial support from NHS Research Scotland (NRS), through C. E. H. Scott of NHS Lothian.

Funding statement:

The authors thank all orthopaedic surgeons whose patients were included in this study.

Ethical review statement:

Ethical approval was obtained from the Scotland (A) Research Ethics Committee (16/SS/0026).

© 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/