■ **KNEE**

Meaningful values in the Short Form Health Survey-36 after total knee arthroplasty – an alternative to the EuroQol five-dimension index as a measure for health-related quality of life

MINIMAL CLINICALLY IMPORTANT DIFFERENCE, MINIMAL IMPORTANT CHANGE, PATIENT-ACCEPTABLE SYMPTOM STATE THRESHOLDS, AND RESPONSIVENESS

N. D. Clement,
D. Weir,
D. Deehan

From Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

Aims

To identify the responsiveness, minimal clinically important difference (MCID), minimal clinical important change (MIC), and patient-acceptable symptom state (PASS) thresholds in the 36-item Short Form Health Survey questionnaire (SF-36) (v2) for each of the eight dimensions and the total score following total knee arthroplasty (TKA).

Methods

There were 3,321 patients undergoing primary TKA with preoperative and one-year postoperative SF-36 scores. At one-year patients were asked how satisfied they were and "How much did the knee arthroplasty surgery improve the quality of your life?", which was graded as: great, moderate, little (n = 277), none (n = 98), or worse.

Results

Physical function, role limitations due to physical problems ('role physical'), bodily pain, and the total score SF-36 scores demonstrated the greatest effect sizes (> 0.9). The MCID for each of SF-36 dimensions ranged from 1.7 for role emotional to 6.4 for bodily pain. The MICs for a cohort of patients ranged from -1.0 for general health to 11.1 for bodily pain. The MICs for an individual patient were marginally greater (one to two points) compared to those for a cohort, and ranging from 0.0 for general and mental health to 13.5 for physical function. The lowest PASS score threshold was associated with physical function (> 34 points) whereas the greatest threshold (> 69 points) was associated with mental health.

Conclusion

The SF-36 is a responsive tool, and the estimates for MCID, MIC, and PASS thresholds that can be used to power studies, assess whether there has been a meaningful change in patients' health-related quality of life, and can be used as a marker of achieving patient satisfaction following TKA.

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

Article focus

■ To identify the responsiveness and meaningful changes/thresholds in the SF-36 (v2) for each of the eight dimensions and the total score following total knee arthroplasty.

- Meaningful values identified were minimal clinically important difference, minimal clinical important change, and patient-acceptable symptom state thresholds.

Key messages
- The meaningful changes demonstrated for the eight dimensions and total SF-36 scores can be used to define clinically significant differences between groups, a clinically significant change for a cohort or an individual, and a postoperative score associated with patient satisfaction.
- The physical function, role physical, bodily pain dimensions, and total scores were the most responsive scores.
- No score demonstrated a floor effect, and a ceiling effect was only observed with the role emotional and social function dimensions.
- The findings of this study suggest that the SF-36 may be a better tool to measure health related quality of life following TKA compared to the EuroQol five-dimension index, however it is recognized to be more burdensome and takes longer to complete.

Strengths and limitations
- The main strength of the study is the size of the arthroplasty registry used, which allowed smaller comparisons between sub-groups of patients.
- The retrospective design and short follow-up (12 months) are weaknesses of the study.

Introduction
Patient reported outcomes measures (PROMs) are now routinely accepted as a measure of outcome following knee arthroplasty.[1] These can be divided into limb- or joint-specific measures, such as the Oxford Knee Score,[2,3] or generic measures of the patient's health-related quality of life (HRQoL).[4] HRQoL PROMs may be overshadowed by the joint-specific outcome and form a secondary aim for most studies assessing the outcome of total knee arthroplasty (TKA).[5–7] Assessment of the HRQoL is, however, of importance when assessing the cost-effectiveness of knee arthroplasty, and more recently has been used to assess the deterioration in patients preoperatively due to prolonged waiting list times.[8] HRQoL scores enable comparisons to be made across medical and surgical specialities in relation to the patient's preintervention status and their ultimate outcome.[9] The EuroQol five-dimension index (EQ-5D) is a useful tool when measuring HRQoL preoperatively, but after TKA there is a reported 84% ceiling effect at one year and it therefore may not be the ideal measure to assess improvement in HRQoL.[10] An alternative measure of HRQoL is the 36-item Short Form Health Survey questionnaire (SF-36), which was designed to be applicable to a wide range of types and severities of conditions and assesses: behavioural functioning, perceived wellbeing, social and role disability, and personal evaluations of health in general.[9] The SF-36 score has been shown to be a responsive tool to assess the outcome of TKA, when using the bodily pain and physical function dimensions, more so than the joint-specific Knee Society function score.[11]

Essential to any PROM are the associated meaningful values, such as the minimal clinically important difference (MCID), minimal clinical important change (MIC), and threshold score associated with a patient-acceptable symptom state (PASS), which are needed to aid interpretation of results and power studies.[12–14] These definitions are often confused and used interchangeably. Escobar et al[15] defined the MCID for the Spanish version of the SF-36 scores following TKA, however it may be argued that methods used to calculate these are more consistent with the MIC.[12,16] Keurentjes et al[17] used a distribution method (80% of the standard deviation) to define the clinically important difference (CID) using the Dutch version of the SF-36, but this may not reflect 'minimal' important difference. More recently, Goh et al[18] defined the PASS in the physical and mental components of the SF-36 following unicompartmental knee arthroplasty (UKA), but the PASS for the eight dimensions of the SF-36 following TKA have not been reported to the authors' knowledge. Furthermore, some authors have used a "total" score SF-36 as a single measure of HRQoL to assess patient outcome,[19] which to the authors' knowledge has not been used to assess the outcome of TKA. Although Lingard et al[11] have previously shown the SF-36 to be responsive to change following TKA, this was only for the physical function and bodily pain dimensions of the score. In addition, the depth of health measured by the SF-36 is an important property, with scores demonstrating a floor or ceiling effect not being optimal measures to assess the outcome of TKA.[20,21]

The primary aim of this study was to identify MCID, MIC, and PASS thresholds in the English version of the SF-36 for each of the dimensions and the total score at one year following TKA. The secondary aims were: 1) to assess responsiveness (effect size); 2) to define the floor and ceiling effects for each of the SF-36 dimensions and the total score one year following TKA; and 3) to assess the association of the pre- and postoperative SF-36 dimension and total scores with level of patient satisfaction.

Methods
Patients for this study were identified retrospectively from an established arthroplasty database held at the study centre. During a 14-year period, 3,791 patients undergoing primary TKA at the study centre completed a preoperative patient questionnaire. Patients were excluded if they had bilateral TKA, a postoperative deep infection, or underwent revision during the first postoperative year. There were 3,321 TKAs performed during the study period with complete pre- and postoperative (one year) data that met the inclusion criteria. There were 1,517 male patients and 1,804 female patients, with a combined mean age of 69.1 years (standard deviation

Table I. Pre- and postoperative 36-item Short Form Health Survey questionnaire scores for the study cohort and the change in the scores.

SF-36 Component	Preoperative, mean (SD)	Postoperative, mean (SD)	Change (95% CI)	p-value*
Physical Function	25.6 (19.0)	52.2 (27.8)	26.6 (25.7 to 27.5)	< 0.001
Role Physical	31.3 (24.2)	56.6 (30.4)	25.3 (24.2 to 26.4)	< 0.001
Bodily Pain	29.2 (18.0)	56.1 (27.2)	26.9 (26.0 to 27.8)	< 0.001
General Health	58.4 (22.0)	61.3 (23.4)	2.9 (2.3 to 3.6)	< 0.001
Vitality	42.1 (21.0)	52.9 (22.5)	10.7 (10.1 to 11.5)	< 0.001
Social Function	52.9 (29.0)	74.0 (30.0)	21.1 (20.1 to 22.2)	< 0.001
Role Emotional	62.5 (34.3)	74.0 (30.8)	11.5 (10.3 to 12.7)	< 0.001
Mental Health	70.0 (19.4)	74.3 (19.6)	4.3 (3.7 to 4.9)	< 0.001
Total	46.9 (16.8)	63.2 (21.4)	16.3 (15.7 to 17.0)	< 0.001

*Paired *t*-test.
CI, confidence interval; SD, standard deviation.

Table II. Mean improvement in the 36-item Short Form Health Survey questionnaire dimension and total scores for those patients declaring no improvement (n = 98) and those with little improvement (n = 277) in their quality of life. The minimal clinically important difference and minimal important change for a cohort were derived from these groups.

SF-36 component	No improvement, mean (SD)	Little improvement, mean (SD)	MCID	MIC (cohort)
Physical Function	6.5 (22.1)	10.4 (24.1)	3.9	10.4
Role Physical	6.2 (23.4)	10.3 (26.2)	4.0	10.3
Bodily Pain	4.7 (18.3)	11.1 (19.4)	6.4	11.1
General Health	-6.3 (17.8)	-1.0 (17.1)	5.3	-1.0
Vitality	-0.5 (22.1)	2.4 (18.7)	2.9	2.4
Social Function	1.6 (27.8)	9.0 (29.2)	7.3	9.0
Role Emotional	0.9 (36.7)	2.7 (34.2)	1.7	2.7
Mental Health	-3.6 (18.9)	0.8 (16.9)	4.4	0.8
Total	1.1 (13.9)	5.7 (15.0)	4.6	5.7

MCID, minimal clinically important difference; MIC, minimal clinical important change; SD, standard deviation; SF-36, 36-item Short Form Health Survey questionnaire.

(SD) 9.7). The mean BMI preoperatively was 30.0 kg/m^2 (SD 6.5).

The SF-36 v2 (QualityMetric Incorporated, Canada) was assessed preoperatively and at one year postoperatively using a written questionnaire.[9,22] The SF-36 assesses eight dimensions that include: physical function, role limitations due to physical health problems, bodily pain, general health, vitality, social functioning, role limitation due to emotional problems, and mental health. Scores for each of the dimensions range from 0 (worst level of functioning) to 100 (best level of functioning). The total SF-36 was calculated as a mean of all eight dimensions assessed, by combining all eight dimension scores and dividing by eight.[19]

Two anchor questions assessing patient quality of life and satisfaction at one year following TKA were used to define the MCID, MIC, and PASS, respectively. Patients were asked "How much did the knee arthroplasty surgery improve the quality of your life?" A five-point Likert scale was used to record the response to the question: "a great improvement", "moderate improvement", "little improvement", "no improvement at all", or "the quality of my life is worse".[16] Patients were also asked "Overall how satisfied are you with the results of your knee arthroplasty surgery?" The response to the

question was recorded using a four-point Likert scale: very satisfied, somewhat satisfied, somewhat dissatisfied, and very dissatisfied. This question with the four-point Likert assessment has been validated and demonstrated to be reliable to measure satisfaction after TKA.[23] Patients stating they were 'very satisfied' and 'satisfied' were categorized as satisfied, and those who defined their outcome as 'dissatisfied' or 'very dissatisfied' were categorized as dissatisfied of assessment on the PASS.

MCID. The MCID was defined using an anchor-based method according to the patients' assessment of improvement in their quality of life. The MCID was defined as the difference in the mean change for each of the eight dimensions of the SF-36 and the total SF-36 scores between patients responding with "no improvement" compared to those responding with "little improvement" in quality of life.

MIC. The MIC for a cohort was defined for each of the dimensions of the SF-36 and the total SF-36 scores as the change, relative to preoperative scores, for those patients declaring their improvement in quality of life as "little improvement". Receiver operating characteristic (ROC) curve analysis was used to determine the MIC for an individual patient, and was defined as the threshold value

Table III. Minimal important change for individual patients using receiver operating characteristic curve analysis to identify a threshold in the postoperative 36-item Short Form Health Survey questionnaire component scores that were predictive of patient satisfaction one year following total knee arthroplasty.

SF-36 component	AUC	95% CI	p-value	MIC (individual)
Physical Function	0.804	0.780 to 0.829	< 0.001	13.5
Role Physical	0.744	0.717 to 0.772	< 0.001	11.5
Bodily Pain	0.791	0.768 to 0.814	< 0.001	11.0
General Health	0.636	0.603 to 0.669	< 0.001	0.0
Vitality	0.673	0.640 to 0.705	< 0.001	5.2
Social Function	0.697	0.665 to 0.728	< 0.001	12.8
Role Emotional	0.636	0.601 to 0.672	< 0.001	4.2
Mental Health	0.632	0.597 to 0.667	< 0.001	0.0
Total	0.803	0.778 to 0.828	< 0.001	7.7

AUC, area under the curve; CI, confidence interval; MIC, minimal important change; SF-36, 36-item Short Form Health Survey questionnaire.

in the dimensions of the SF-36 and the total SF-36 scores that were predictive of patient satisfaction.

PASS. The PASS threshold was defined as the postoperative score in each dimension of the SF-36 and the total SF-36 scores that were predictive of a patient declaring their outcome as satisfactory. ROC curve analysis was used to determine the PASS threshold value in the postoperative SF-36 score that was predictive of patient satisfaction.

Responsiveness, and floor and ceiling effects. Responsiveness was assessed using the effect size, which was defined as the mean change in the SF-36 scores divided by the SD.[24] The greater the effect size, the greater the difference between the groups (pre- and postoperative score), and therefore responsiveness. Cohen suggested that an effect size of 0.2 was small, 0.5 was medium, and 0.8 was large.[24] A floor or ceiling effect was defined as present when the percentage of patients scoring the minimum or maximum score pre- or postoperatively was more than 15% of the study cohort.[20]

Statistical analysis. Statistical Package for Social Sciences version 17.0 (SPSS, USA) was used for all data analysis. Simple descriptive statistics were undertaken. Data were assessed for normality and parametric tests where appropriate. Scalar variables were assessed for change using paired (within group) t-tests. Spearman correlation was used to assess association of the SF-36 with postoperative satisfaction. ROC curve analysis was used to identify a threshold (point of maximal sensitivity and specificity) in the mean change in scores that were predictive of a "little improvement" in quality of life and patient satisfaction in the change (MIC) and the postoperative (PASS) scores, respectively. The area under the ROC curve (AUC) ranges from 0.5, indicating a test with no accuracy in distinguishing whether a patient is satisfied, to 1.0, where the test is perfectly accurate identifying all satisfied patients. An AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.[25] Significance was set as a p-value of < 0.05.

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

(Newcastle Hospitals NHS Foundation Trust, Project Record Number 3290) and was conducted in accordance with the Declaration of Helsinki[26] and the guidelines for good clinical practice.

Results

The mean preoperative SF-36 scores ranged from 25.6 to 70.0, which significantly increased postoperatively for all dimensions of the SF-36 and the total score (Table I). The majority of patients felt their quality of life had improved postoperatively at one year: great improvement n = 2,066 (62.2%), moderate improvement n = 742 (22.3%), little improvement n = 277 (8.3%), no improvement n = 138 (4.2%), and worse n = 98 (3.0%). There were 2,979 patients (89.7%) who were satisfied at one year and 342 (10.3%) who were not.

The MCID for each of SF-36 dimensions ranged from 1.7 for role emotional to 7.3 for social function, and the MCID for the total score was 4.6 (Table II). The MICs for a cohort of patients ranged from -1.0 for general health to 11.1 for bodily pain (Table II). The MICs for an individual patient were marginally greater by one to two points when compared to those for a cohort, ranging from 0.0 for general and mental health to 13.5 for physical function (Table III). All dimensions of the SF-36 and the total score changes and postoperative scores were significant predictors of postoperative satisfaction following TKA (Tables III and IV, Supplementary Figures a and b). However, only physical function, role physical, bodily pain dimensions, and the total SF-36 were acceptable or excellent predictors of satisfaction when the MIC was achieved according to the AUC (Table III). On the other hand, all of the postoperative dimensions of the SF-36 and the total scores were acceptable or excellent predictors of satisfaction when the PASS score was achieved according to the AUC (Table IV). The lowest PASS score was associated with physical function (> 34 points) whereas the greatest score (> 69 points) was associated with mental health (Table IV).

Physical function, role physical, bodily pain dimensions, and the total SF-36 scores demonstrated a large effect size (> 0.8), whereas vitality and social function

Table IV. Patient-acceptable symptom state for the postoperative 36-item Short Form Health Survey questionnaire component scores that were predictive of patient satisfaction one year following total knee arthroplasty.

SF-36 component	AUC	95% CI	p-value	PASS
Physical Function	0.800	0.775 to 0.824	< 0.001	> 34
Role Physical	0.786	0.763 to 0.809	< 0.001	> 43
Bodily Pain	0.818	0.797 to 0.838	< 0.001	> 54
General Health	0.730	0.702 to 0.758	< 0.001	> 55
Vitality	0.750	0.724 to 0.777	< 0.001	> 47
Social Function	0.781	0.754 to 0.807	< 0.001	> 56
Role Emotional	0.717	0.686 to 0.748	< 0.001	> 65
Mental Health	0.703	0.673 to 0.733	< 0.001	> 69
Total	0.823	0.802 to 0.843	< 0.001	> 51

AUC, area under the curve; CI, confidence interval; PASS, patient-acceptable symptom state; SF-36, 36-item Short Form Health Survey questionnaire.

Table V. The effect size, and the floor and ceiling effects (%) for the pre- and postoperative dimensions of the 36-item Short Form Health Survey questionnaire and the total score.

SF-36 component	Mean change	SD	Effect size	Preoperative (%)		Postoperative (%)	
				Floor	Ceiling	Floor	Ceiling
Physical Function	26.6	27.0	0.99	6.0	0.2	2.1	1.7
Role Physical	25.3	30.6	0.83	12.0	1.3	4.1	11.4
Bodily Pain	26.9	26.3	1.02	6.0	0.4	1.7	10.1
General Health	2.9	18.1	0.16	0.2	1.0	0.3	2.4
Vitality	10.8	20.7	0.52	2.7	0.2	1.7	1.3
Social Function	21.1	30.6	0.69	3.7	9.8	1.7	31.0
Role Emotional	11.5	33.6	0.34	5.4	24.1	2.9	32.7
Mental Health	4.3	17.4	0.25	0.2	2.5	0.2	5.7
Total	16.3	17.5	0.93	0.0	0.0	0.0	0.2

SD, standard deviation; SF-36, 36-item Short Form Health Survey questionnaire.

dimensions had a moderate effect size, and general health, role emotional, and mental health dimensions had a small effect size (Table V). No dimension of the SF-36 or the total score demonstrated a floor effect (Table V). A ceiling effect was only observed with role emotional pre- and postoperatively and social function postoperatively, but was not present for any other score (Table V).

Level of patient satisfaction was associated with each of the pre- and postoperative SF-36 dimension and the total score, with a higher (better) pre- or postoperative score being associated with higher levels of satisfaction (Supplementary Table i, Supplementary Figures c and d). The highest correlation with satisfaction was observed in the postoperative scores, and more specifically highest with the postoperative total score (Supplementary Table i).

Discussion

This study has identified meaningful values in the SF-36 dimensions and total scores that can be used to define clinically significant differences between groups, a clinically significant change for a cohort or an individual, and postoperative scores associated with patient satisfaction. However, the responsiveness to change following TKA varied according to the dimension of the SF-36 assessed, with physical function, role physical, bodily pain dimensions, and the total scores being the most responsive. No score demonstrated a floor effect, and a ceiling effect was only observed with role emotional and social function dimensions. Both the pre- and postoperative SF-36 dimension scores and the total scores were associated with level of patient satisfaction.

Assessment of HRQoL has become a predominant topic when assessing preoperative health status, with longer waiting times for hip and knee arthroplasty resulting in a greater proportion of patients living in a state worse than death (WTD).[8] A state WTD was defined as an EQ-5D score of less than zero;[27] such a threshold does not exist for the SF-36 scores. The major limitation of the EQ-5D is the postoperative ceiling effect of up to 84%, making it a potentially poor measure to assess the outcome of TKA.[10] In contrast, the SF-36 dimension score, with the exception of social functioning and emotional role, did not demonstrate floor or ceiling effects. This suggests that the SF-36 may be a better tool to measure HRQoL following TKA compared to the EQ-5D, however it is recognized to be more burdensome and take longer to complete.[28] Furthermore, the SF-36 has established population age- and sex-standardized scores with which outcome could be compared, and would also allow for the comparison of outcomes in relation to HRQoL improvement relative to other non-specific orthopaedic interventions.[22,29]

An alternative to using the eight dimensions of the SF-36 to assess HRQoL is to use the two summary scores which can be calculated from the dimensions: the physical component summary (PCS) and the mental component summary (MCS).[22] These are aggregates of the eight scale scores that account for more than 80% of the variance in the dimensions.[9] The PCS is calculated by positive weighting of four of the physical subscales and negative weighting of the psychological subscales, whereas the MCS is calculated using the opposite weighting.[22] However, due to the extent of physical disability endured by patients with knee arthritis, the method for calculating the MCS may be artificially increased, causing mental health distress to be under-represented in the summary score.[22] Therefore, retaining all eight dimensions of the SF-36 may be preferable to using the summary scores alone when trying to assess the influence of arthritis on a patient's mental health.

The MCID and MIC are essential definitions required for assessing the clinically relevant difference between two groups of patients, and whether a cohort or an individual patient has had a clinically relevant improvement in their score, respectively. Escobar et al[15] have previously defined the MCID in the SF-36 dimensions, but the methodology used may be more in keeping with the MIC. Their quoted scores were, however, generally higher than the MIC demonstrated in the current study for a cohort of patients. This may be due to the fact that Escobar et al[15] use the Spanish version of the SF-36, or to their limited number of patients in the "somewhat better" (n = 76) and equal groups (n = 32). Keurentjes et al,[17] using the Dutch version of the SF-36, defined the CID using a 'rule of thumb' method by using 80% of the SD, however their suggested scores are far greater than the current study and those suggested by Escobar et al.[15] The authors are not aware of any published studies defining the MCID and/or MIC in the English version of the SF-36 after TKA. Clement et al[13,30] have previously defined the MCID following TKA for the SF-12, which varies from 2 to 5 for the PCS score and 2 for the MCS score. Despite the SF-12 score being calculated differently, the MCID for the PCS is similar to the physical dimensions (PF, RP, BP, and GH) of the SF-36, but the MCID for the MCS was lower than the 4.4 points for the mental health component demonstrated in the current study. This may be due to the way the MCS is calculated and the effect of physical dysfunction resulting in an underestimate of the mental effect of arthritis, as discussed above.[22] The PASS threshold scores identified in the current study for physical function and mental health were lower than those defined by Goh et al[18] of 50 for the PCS and 55 for the MCS scores following UKA. This may reflect intrinsic patient differences between TKA and UKA,[5] but may also be related to the differences associated with the component scores used.[22] However, it would appear that the values presented in the current study were more predictive of a PASS than those suggested by Goh et al.[18] with a greater AUC.

The SF-36 has previously been shown to be responsive to change following TKA, but this was only assessed for the physical function and bodily pain components with effect sizes similar to the current study of one.[11] The current study also demonstrated role physical to have a large effect size, whereas the mental health and genral health had small effect sizes. These differences would need to be acknowledged when powering a study, with a small effect size requiring more patients to be recruited.[24]

An original aspect of the current study was assessment of total SF-36 following TKA. Although the total SF-36 is not acknowledged by the designers, and it has been suggested to be pointless to combine the two summary measures into one overall measure of HRQoL.[19] Despite this numerous studies have done so and it is increasingly being used in the scientific literature.[19] This study has shown that the total SF-36 score may be an ideal measure to assess HRQoL pre- and post-TKA, being responsive to change (large effect size) and no floor or ceiling effect. In addition, it is predictive of patient satisfaction. The defined MCID, MIC, and PASS threshold would allow this new score to be used in future studies assessing the outcome of TKA.

The SF-36 score was assessed at an early timepoint of only one year, which constitutes a limitation. However, it would seem that the one-year postoperative measure is similar to that observed at six and 24 months.[15,31] The identified meaningful values for the SF-36 may be specific to TKA, as a previous study has shown differences in the MCID for total hip arthroplasty and TKA when using a joint-specific PROM.[12] The style of anchor question used to calculate the MCID has been shown to yield variations in scores, which has been shown previously.[14] In the present study, improvement in the patient's 'quality of life' was assessed as the anchor question. Cultural and demographic variations in a patient population are known to influence PROM scoring, and the results from the present study are likely to be specific to a UK population and may not be translatable to other countries for this reason.[32] A further limitation was using patient satisfaction to define the individual MIC and the PASS, as the preoperative score was shown to significantly influence the level of satisfaction, and therefore without adjusting for this confounding factor, these may vary according to the patient's preoperative score.

In conclusion, the SF-36 is a responsive tool, and the estimates for MCID, MIC, and PASS thresholds can be used to power studies and assess whether there has been a meaningful change in patients' HRQoL, as well as serve as a marker of patient satisfaction following TKA.

Supplementary material

ë   Table and figures showing associations between patient satisfaction and the 36-item Short Form Health Survey questionnaire.

References

1. **Wang Y**, **Yin M**, **Zhu S**, **Chen X**, **Zhou H**, **Qian W**. Patient-reported outcome measures used in patients undergoing total knee arthroplasty. *Bone Joint Res.* 2021;10(3):203–217.

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. **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.

4. **Rolfson O**, **Eresian Chenok K**, **Bohm E**, **Lübbeke A**, **Denissen G**, **Dunn J**. Patient-reported outcome measures in arthroplasty registries: report of the patient-reported outcome measures working group of the international society of arthroplasty registries: part I. overview and rationale for patient-reported outcome measures. *Acta Orthop.* 2016;87(362):3–8.

5. **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.

6. **Holland G**, **Keenan OJ**, **Gillespie M**, **et al**. Patellar cartilage loss does not affect early outcomes of total knee arthroplasty performed without patella resurfacing. *Knee.* 2021;28:194–201.

7. **MacDonald DJ**, **Clement ND**, **Howie CR**, **Scott CEH**. The effect of COVID-19 restrictions on rehabilitation and functional outcome following total hip and knee arthroplasty during the first wave of the pandemic. *Bone Jt Open.* 2021;2(6):380–387.

8. **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.

9. **McDowell I**. General health status and quality of life. *Meas Heal.* 2009;520–703.

10. **Giesinger K**, **Hamilton DF**, **Jost B**, **Holzner B**, **Giesinger JM**. Comparative responsiveness of outcome measures for total knee arthroplasty. *Osteoarthr Cartil.* 2014;22(2):184–189.

11. **Lingard EA**, **Katz JN**, **Wright RJ**, **Wright EA**, **Sledge CB**, **Kinemax Outcomes Group**. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am.* 2001;83-A(12):1856–1864.

12. **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.

13. **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.

14. **Clement ND**, **Scott CEH**, **Hamilton DF**, **MacDonald D**, **Howie CR**. Meaningful values in the Forgotten Joint Score after total knee arthroplasty. *Bone Joint J.* 2021;103-B(5):846–854.

15. **Escobar A**, **Quintana JM**, **Bilbao A**, **Aróstegui I**, **Lafuente I**, **Vidaurreta I**. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthr Cartil.* 2007;15(3):273–280.

16. **Clement ND**, **Bardgett M**, **Weir D**, **Holland J**, **Gerrand C**, **Deehan DJ**. What is the minimum clinically important difference for the WOMAC Index after TKA? *Clin Orthop Relat Res.* 2018;476(10):2005–2014.

17. **Keurentjes JC**, **Fiocco M**, **Nelissen RG**. Willingness to undergo surgery again validated clinically important differences in health-related quality of life after total hip replacement or total knee replacement surgery. *J Clin Epidemiol.* 2014;67(1):114–120.

18. **Goh GS**, **Liow MHL**, **Chen JY**, **Tay DKJ**, **Lo NN**, **Yeo SJ**. The patient acceptable symptom state for the knee society score, oxford knee score and short form-36 following unicompartmental knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc.* 2021; Epub ahead of print.

19. **Lins L**, **Carvalho FM**. SF-36 total score as a single measure of health-related quality of life: Scoping review. *SAGE Open Med.* 2016;4:2050312116671725.

20. **McHorney CA**, **Tarlov AR**. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995;4(4):293–307.

21. **Clement ND**, **Afzal I**, **Demetriou C**, **Deehan DJ**, **Field RE**, **Kader DF**. The preoperative Oxford Knee Score is an independent predictor of achieving a postoperative ceiling score after total knee arthroplasty. *Bone Joint J.* 2020;102-B(11):1519–1526.

22. **Laucis NC**, **Hays RD**, **Bhattacharyya T**. Scoring the SF-36 in Orthopaedics: A Brief Guide. *J Bone Joint Surg Am.* 2015;97-A(19):1628–1634.

23. **Mahomed N**, **Gandhi R**, **Daltroy L**, **Katz JN**. The self-administered patient satisfaction scale for primary hip and knee arthroplasty. *Arthritis.* 2011;2011:591253.

24. **Sullivan GM**, **Feinn R**. Using effect size-or why the p-value is not enough. *J Grad Med Educ.* 2012;4(3):279–282.

25. **Mandrekar JN**. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9):1315–1316.

26. **World Medical Association**. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191–2194.

27. **Scott CEH**, **MacDonald DJ**, **Howie CR**. "Worse than death" and waiting for a joint arthroplasty. *Bone Joint J.* 2019;101-B(8):941–950.

28. **Dunbar MJ**, **Robertsson O**, **Ryd L**, **Lidgren L**. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Joint Surg Br.* 2001;83-B(3):339–344.

29. **Burholt V**, **Nash P**. Short Form 36 (SF-36) Health Survey Questionnaire: normative data for Wales. *J Public Health (Oxf).* 2011;33(4):587–603.

30. **Clement ND**, **Weir D**, **Holland J**, **Gerrand C**, **Deehan DJ**. Meaningful changes in the Short Form 12 physical and mental summary scores after total knee arthroplasty. *Knee.* 2019;26(4):861–868.

31. **Tew M**, **Dalziel K**, **Clarke P**, **Smith A**, **Choong PF**, **Dowsey M**. Patient-reported outcome measures (PROMs): can they be used to guide patient-centered care and optimize outcomes in total knee replacement? *Qual Life Res.* 2020;29(12):3273–3283.

32. **Giesinger JM**, **Giesinger K**, **Federico B**, **Howie CD**, **Hamilton DF**. Differences in case mix and outcomes between Swiss and Scottish total knee arthroplasty patients. *Knee Surg Sports Traumatol Arthrosc.* 2020;28(6):1797–1804.

Author information:
- N. D. Clement, MD, PhD, FRCS Ed (Tr&Orth), Orthopaedic Consultant, Department of Orthopaedics, Royal Infirmary of Edinburgh, Edinburgh, UK.
- D. Weir, FRCS (Tr&Orth), Orthopaedic Consultant
- D. Deehan, MD, MSc, FRCS (Tr&Orth), DSc, Professor of Orthopaedics Department of Orthopaedics, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.

Author contributions:
- N. D. Clement: Conceptualization, Formal analysis, Writing – original draft.
- D. Weir: Investigation, Writing – review & editing.
- D. Deehan: Conceptualization, Investigation, Writing – review & editing.