Substantial clinical benefit and patient acceptable symptom states of the Forgotten Joint Score 12 after primary knee arthroplasty

Siri HEIJBEL [1], Annette W-DAHL [2,3], Kjell G NILSSON [4], and Margareta HEDSTRÖM [1,5]

[1] Department of Clinical Science Intervention and Technology, Karolinska Institutet, Stockholm; [2] Orthopedics, Department of Clinical Sciences Lund, Lund University, Skane University Hospital, Department of Orthopedics, Lund; [3] The Swedish Knee Arthroplasty Register; [4] Department of Surgical and Perioperative Sciences, Orthopedics, Umeå University, Umeå; [5] Trauma & Reparative Medicine Theme, Karolinska University Hospital Huddinge, Stockholm, Sweden
Correspondence: siri.heijbel@regionstockholm.se

Background and purpose — Knowing how to interpret values obtained with patient reported outcome measures (PROMs) is essential. We estimated the substantial clinical benefit (SCB) and patient acceptable symptom state (PASS) for Forgotten Joint Score 12 (FJS) and explored differences depending on methods used for the estimates.

Patients and methods — The study was based on 195 knee arthroplasties (KA) performed at a university hospital. We used 1 item from the Knee injury and Osteoarthritis Outcome Score domain quality of life and satisfaction with surgery, obtained 1-year postoperatively, to assess SCB and PASS thresholds of the FJS with anchor-based methods. We used different combinations of anchor questions for SCB and PASS (satisfied, satisfied with no or mild knee difficulties, and satisfied with no knee difficulties). A novel predictive approach and receiver-operating characteristics curve were applied for the estimates.

Results — 70 and 113 KAs were available for the SCB and PASS estimates, respectively. Depending on method, SCB of the FJS (range 0–100) was 28 (95% CI 21–35) and 22 (12–45) respectively. PASS was 31 (2–39) and 20 (10–29) for satisfied patients, 40 (31–47) and 38 (32–43) for satisfied patients with no/mild difficulties, and 76 (39–80) and 64 (55–74) for satisfied patients with no difficulties. The areas under the curve ranged from 0.82 to 0.88.

Interpretation — Both the SCB and PASS thresholds varied depending on methodology. This may indicate a problem using meaningful values from other studies defining outcomes after KA. This study supports the premise of the FJS as a PROM with good discriminatory ability in patients undergoing KA.

Patient reported outcome measures (PROMs) in orthopedics enable the assessment of patient perspective after interventions such as knee arthroplasty (KA). There are various hypothesis-driven approaches on how to interpret changes and interventional outcomes assessed by PROMs (1). Determination of meaningful values such as the substantial clinical benefit (SCB) and patient acceptable symptom state (PASS) is used to understand the clinical relevance of obtained scores. The SCB is an estimate of an individual change value in PROMs corresponding to a considerate clinical benefit (2). Postoperative threshold values for treatment success can be defined with PASS, reflecting a state when patients perceive themselves as well.

The Forgotten Joint Score 12 (FJS) has repeatedly shown good validity and reliability since it was introduced by Behrend et al. (3). We recently validated the Swedish translation in patients undergoing KA (4). There is no previous assessment of the SCB for the FJS. A few recent studies have estimated PASS thresholds (5–7). However, due to cultural differences and variations in study cohorts as well as practices in health care it is important to multiply assessments of meaningful values in PROMs.

We estimated the SCB, PASS values, and the discriminative ability for the FJS in patients undergoing KA. In addition, we explored differences depending on assumptions and methods related to meaningful values.

Patients and methods

Study setting, data collection, and inclusion

This is a retrospective study of prospectively collected data. All patients (n = 195) who underwent KA between June 1, 2016 and March 21, 2018 at the Department of Reconstrutive Ortho-
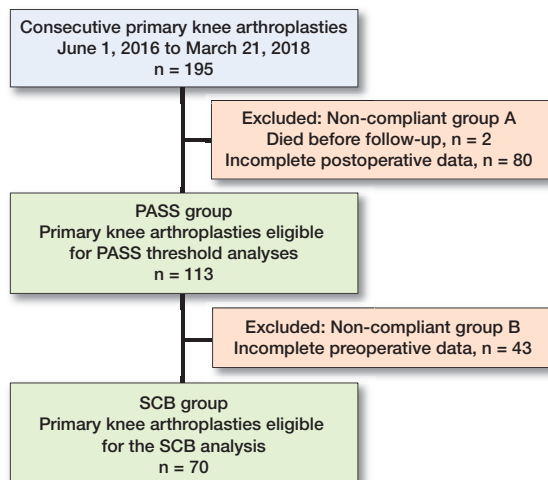
Inclusion processes for the patient acceptable symptom state (PASS) group and the substantial clinical benefit (SCB) group.

Table 1. Anchor question chosen for assessment of the substantial clinical benefit and patient acceptable symptom state thresholds of the Forgotten Joint Score 12, obtained from the Knee injury and Osteoarthritis Outcome Score domain quality of life (Q4) 1 year postoperatively

In general, how much difficulty do you have with your knee?

Optimal clinical outcome
    None
    Mild
Not optimal clinical outcome
    Moderate
    Severe
    Extreme

pedics at Karolinska University Hospital Huddinge in Stockholm, Sweden were included in the study. The clinics' routine before and 1 year after KA is to provide a set of questionnaires to the patients, which subsequently is reported to the Swedish Knee Arthroplasty Register (SKAR). In this study, the FJS was added to the regular set of questionnaires and extracted manually while SKAR provided data from the Knee injury and Osteoarthritis Outcome Score (KOOS), satisfaction with the surgery, sex, age, BMI, and ASA classification.

Patients were retrospectively included in the study if the prospectively assessed FJS and predetermined anchor questions for each hypothesis were sufficiently completed and returned (Figure). The characteristics of patients scheduled for a KA with (compliant group) and without (non-compliant group) sufficient data were compared to identify any differences between groups and potential non-response bias. Depending on the analysis performed, the compliant group was divided into the PASS group and the SCB group. The non-compliant group was divided into non-compliant group A (insufficient postoperative data) and non-compliant group B (insufficient pre- and postoperative data).

Forgotten Joint Score 12

The FJS consists of 12 items related to joint awareness in daily life (3). Each item is answered within a 5-point Likert scale with the following response options: never (0 p); almost never (1 p); seldom (2 p); sometimes (3 p); and mostly (4 p). The initial raw data was converted to a scale ranging from 0 to 100 (worst to best), by dividing the summarized score by the number of completed items, which subsequently is multiplied by 25. Thereafter the score was subtracted from 100.

Anchor questions

Anchor questions are used to hypothesize relevant outcomes related to the specific research question. The anchor ques-

tions were assessed 1 year postoperatively. We selected 1 item from the KOOS domain quality of life (Q4) as a determinant of optimal clinical outcome (Table 1). Furthermore, we selected patient satisfaction with the surgery measured with a VAS scale ranging from 0 to 100 (highest to worst imaginable satisfaction). Patients were classified as very satisfied (0–20), satisfied (21–40), moderately satisfied (41–60), unsatisfied (61–80), and very unsatisfied (81–100). We defined satisfaction as those who reported very satisfied and satisfied (8).

Substantial clinical benefit

Our hypothesis was that optimal clinical outcome was reached if the patients were satisfied with the surgery and reported their knee difficulties as none or mild 1 year postoperatively. We used that classification as an anchor to estimate the change scores in the FJS that were predictive of treatment success.

Patient acceptable symptom state

We explored differences in PASS threshold estimates depending on the choice of anchor in: 1) satisfied patients, 2) satisfied patients with no or mild knee difficulties, and 3) satisfied patients with no knee difficulties. Anchor 2 – satisfied patients with no or mild knee difficulties – was our primary choice of anchor, used in both the SCB analysis and the PASS analysis.

Statistics

For test of normality, the Shapiro–Wilk test was used. Mean values (SD) were given for normally distributed data and median (range) was given for non-normally distributed data. The results were presented with 95% confidence interval (CI) obtained by bootstrapping (2,000 samples). We used the R syntax provided by Terwee et al. (9) for the analyses of meaningful values. The statistical analyses were performed with SPSS Statistics 27.0 (IBM Corp, Armonk, NY, USA) and R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/).

Table 2. Patient characteristics of the groups

Factor	All KA n = 195	PASS group n = 113	SCB group n = 70	N-C group A n = 82	N-C group B n = 43
Mean age (SD)	68 (10)	69 (9)	69 (9)	67 (10)	69 (10)
Females, n (%)	120 (62)	69 (61)	41 (59)	51 (62)	28 (65)
ASA 3/4, n (%)	86/3 (46)	47/2 (43)	27/1 (40)	39/1 (49)	20/1 (49)
BMI (SD)	30 (5)	30 (5)	29 (5)	30 (5)	30 (4)

Non-compliant (N-C) group A had no postoperative PROM data, and non-compliant (N-C) group B had no pre- or postoperative PROM data.

Table 3. Estimated thresholds for patient acceptable symptom state (PASS) using 2 different methods, predictive modeling, and receiver operating characteristics (ROC) curve analyses, and 3 different anchor questions. Values in parentheses are 95% CI

Factor	Satisfied patients	Satisfied patients with no or mild knee difficulties	Satisfied patients with no knee difficulties
PASS predictive	20 (10–29)	38 (32–43)	64 (55–74)
PASS ROC	31 (2–39)	40 (31–47)	76 (39–80)
AUC	0.82 (0.71–0.93)	0.87 (0.80–0.94)	0.88 (0.80–0.96)
SNS/SPC (%)	63/89	78/89	74/92

Area under the curve (AUC), sensitivity (SNS), and specificity (SPC) are given as measurements of separability.

Predictive approach

We used a novel adjusted predictive anchor-based approach (10,11). The model is based on a logistic regression analysis, which is used to predict a correlation between 1 dependent dichotomous variable (positive outcome, i.e., the anchor), and 1 independent interval scale (the score of the FJS). The prediction separates patients with a large probability of belonging to the group of patients reaching a positive outcome from those with a small probability (10).

It was calculated by first extracting the logistic regression equation parameters C (the constant) and B (the regression coefficient of the score X in the FJS). Afterwards, the postoperative odds of belonging to the group reaching a positive outcome (postodds) were calculated by dividing the probability (P) of belonging to that group by 1–P, in which P is determined by the anchor. The logistic regression equation for a certain change score X reads:

$$\ln(postodds) = C + B \times X$$

The likelihood ratio (LR) can be calculated by dividing the postodds by the preoperative odds of positive outcome (preodds). Therefore, the preodds are equal to the postodds if the LR = 1. Given this, the ln(postodds) in the logistic regression equation was exchanged for ln(preodds) and worked around to determine what change score X yields the SCB:

$$X = (\ln(preodds) – C)/B$$
$$SCB = X$$

If the group of patients reaching a positive outcome is the same size as the group of patients who do not, the score X is the SCB/PASS. However, if the groups differ in size the formula must be adjusted (11). To assess the adjusted value the following formula was used:

$$adjusted\ value =$$
$$SCB – (0.090 + 0.103 \times Cor) \times SD_{FJS} \times \ln(postodds)$$

Cor is the Pearson's correlation coefficient between the FJS score and the dichotomous variable. The SD_{FJS} is the standard deviation of the relevant FJS score.

Receiver-operating characteristics (ROC) curve analyses

Furthermore, ROC curve analyses were performed to estimate the SCB and PASS thresholds. The area under the curve (AUC) was assessed as a measurement of separability. The meaningful values were determined by Youden index, which is the score in which the sum of sensitivity and specificity reaches its maximum.

Ethics, funding, and potential conflicts of interest

The study was conducted in accordance with the Declaration of Helsinki and approved by the Regional Ethical Review Board in Stockholm (2017/280-32). Grants for funding were provided by Region Stockholm (NSV project). No competing interests were declared.

Results

There were 113 patients available for the PASS analyses and 70 patients for the SCB analyses (Figure and Table 2).

PASS group

The median postoperative FJS score was 35 (0–100). There were 95 (84%) satisfied patients, 58 (51%) satisfied patients with no or mild knee difficulties, and 19 (17%) satisfied patients with no knee difficulties. PASS thresholds varied between 20 and 76 depending on anchor and methodology (Table 3).

SCB group

The median preoperative and postoperative FJS score was 8 (0–67) and 45 (0–100), respectively, and the median change score was 33 (–33 to 94). There were 41 patients who reached an optimal outcome. The group of patients who reached an optimal outcome had a median change score from before to 1 year after surgery of 52 (–4 to 94), while the group who did not reach an optimal outcome had a median change score of 4 (–33 to 81). The SCB estimated with the predictive approach was 28 (CI 20–35) and the estimate with ROC curve analysis was 22 (CI 12–45).

Discussion

We aimed to estimate meaningful values for the FJS and explore differences dependent on methodology. We found the SCB corresponding to a change of 28 (CI 20–35) or 22 (CI 12–45) from pre- to 1 year postoperatively. PASS thresholds differed between 20 and 76. The assessed values varied depending on methodology and choice of anchor question.

Choice of anchor question

Meaningful values derived from anchor-based methodological approaches are highly dependent on the choice of anchor question. Although the subject has been thoroughly investigated in the past decades there is still no gold standard. A credible anchor is well defined and relevant to the driven hypothesis such as the presence of specific symptoms, disease activity, or ratings of change (12). Obviously, it should correlate well with the PROM of interest and be comprehensible (9,12). It is important to emphasize that, depending on hypothesis, the anchor questions can be either patient- or clinician-rated, although the former is preferred to assure patient perspective (12).

The anchor questions used in our study were assessed 1 year after KA. Regarding the SCB analysis, all classified as reaching the treatment goal were overall satisfied with the result after surgery and had experienced no or mild knee difficulties during the present week at the time of assessment. A common critique of transitional items as anchor questions, such as ratings of change, is the risk of recall bias and that the rating therefore complies more with the postoperative than the preoperative status (12). Our approach with a combined anchor may reduce that risk. However, a limitation was that patients were not specifically asked about change, improvement, or health status compared with before surgery, which is otherwise commonly done when meaningful change values are estimated (1,12).

Substantial clinical benefit

The SCB, which has been investigated mostly in shoulder and spine surgery, was introduced by Glassman et al. (2) to define treatment success in health-related quality-of-life measures after lumbar spine arthrodesis. Glassman argued that although minimal important change (MIC) values are essential for interpretation and understanding of change scores in PROMs, they represent a floor value rather than a treatment goal. There is no previous assessment of the SCB for the FJS, but there are some suggestions on MIC values that vary between 11 and 18. (6,7,13,14). As expected, the estimations of 28 (CI 20–35) and 22 (CI 12–45) for the SCB are higher in comparison.

To our knowledge, there is no previous assessment of the SCB with the predictive approach but instead most commonly ROC curve based approaches. The last few years the predictive approach has been used more frequently for assessment of MIC values and is now one of the recommended statistical strategies in that area (9). SCB and MIC values are fundamentally different as they represent distinctly different markers on the outcome span. However, since this difference is based solely on the hypothesis and outcome definition, we believe that the predictive approach with advantage can be applied to future SCB values as well.

Patient acceptable symptom state

As with the SCB there are no previous assessments of PASS using the predictive approach; however, the theory can be applied to PASS thresholds as well. A strength in our study is that we estimated PASS values based on 3 different anchors and 2 different methods. These values represent thresholds beyond which patients have reached a satisfactory outcome, depending on predetermined assumptions. The postoperative FJS threshold for being satisfied with the surgery was lower compared with if patients were satisfied and had no postoperative knee difficulties (Table 3). This is not surprising, as measurements of satisfaction after KA are multifaceted and can be affected by several factors (15)

The original idea behind FJS is that it would inherently have the discriminatory ability to separate the best-performing patients and distinguish an excellent outcome from a very good outcome (3). This premise is supported by our findings of the estimated difference between satisfied patients with no knee difficulties and satisfied patients with no or mild knee difficulties. Moreover, AUC values obtained from the ROC analyses were between 0.82 and 0.88, which further strengthen the discriminatory power of the FJS.

Our primary choice of anchor, satisfaction combined with no or mild knee difficulties, generated a PASS ROC threshold of 40 (CI 31–47) and a PASS predictive threshold of 38 (CI 32–43). This specific anchor estimated similar values, which is probably due to the fact that the groups (positive or negative outcome) were almost equal in size (10,11). The confidence interval is slightly wider in the ROC estimate and this supports previous suggestions that the predictive approach is more precise (10).

Previous PASS thresholds for the FJS have been suggested. Wang et al. (5) and Longo et al. (6) suggested thresholds of 41 and 73, respectively, in patients who underwent UKA. In patients who underwent TKA a threshold of 22 have been suggested (7). Wang et al. (5) and Clement et al. (7) used satisfaction as anchor. In the study performed by Longo et al. (6) patients were asked to rate their general health as good or not. The preceding results and the PASS thresholds in our study illustrate the anchor-dependent impact on meaningful values, how values can vary between studies on the same PROM, and differences depending on methodology.

The implication of meaningful values

Estimations such as SCB, MIC, and PASS are used to interpret data in research and, at least in theory, to aid shared decision-

making prior to treatments or interventions (**9**). Previously assessed meaningful values of the FJS vary broadly between studies due to differences in the study cohorts and perhaps due to variations in healthcare systems and cultural contexts. As noted in our study as well as in others, the values obtained also vary depending on methodology and assumptions made (**5,7,13**). This indicates that although the estimates can be used as guidance when arbitrary thresholds are decided in clinical trials, perhaps the most applicable method would be to make estimates for that specific cohort if PROMs are used as a primary outcome measurement.

Response rate

The main limitation in our study was the low response rate, which is a known factor in outcome assessment with PROMs (**16**). There are several possible explanations to this. First, a logistical disadvantage with the PROM assessment was that patients on the postoperative occasion received a set of questionnaires at home that required to be returned to the clinic by postal mail. At the preoperative assessment patients were asked to complete the PROMs at the clinic and in comparison, as seen in the Figure, there were almost twice as many missing PROMs at the postoperative assessment. Second, some of the patients living in the area served by the present hospital do not have Swedish as their primary language, which has been shown to decrease the likelihood of PROM completion (**16**). Finally, the PROMs were distributed non-electronically, which could lower the completion rates (**16**). However, we assessed the patient characteristics in the non-compliant groups and compared these with the compliant groups (Table 2). The proportion of patients with ASA classification 3 and 4 was somewhat larger while age, proportion of females, and BMI were similar.

Generalizability of the results

Characteristics in the compliant groups such as age, sex, and BMI are similar according to national data from the SKAR. However, the percentage of patients with ASA classification 3 or 4 was higher (**8**). Comorbidity does not only have an influence on the PROM ratings but also on rehabilitation and completion rate and may therefore affect the generalizability of our results (**16**). Furthermore, the generalizability is also affected by the single hospital format as well as the previously mentioned low response rate.

Conclusion

Both the SCB and the PASS thresholds varied depending on choice of anchor and methods used. This may indicate a problem using such values for defining successful outcome after knee arthroplasty, although they may be used as guidance for determining arbitrary values in research, clinical practice, and in registry settings. This study supports the premise of the FJS as a PROM with good discriminatory ability in patients undergoing KA.

1. **Engel L, Beaton D E, Touma Z**. Minimal Clinically important difference: a review of outcome measure score interpretation. Rheum Dis Clin North Am 2018; 44(2): 177-88. doi: 10.1016/j.rdc.2018.01.011

2. **Glassman S D, Copay A G, Berven S H, Polly D W, Subach B R, Carreon L Y.** Defining substantial clinical benefit following lumbar spine arthrodesis. J Bone Joint Surg Am 2008; 90(9): 1839-47. doi: 10.2106/jbjs.g.01095

3. **Behrend H, Giesinger K, Giesinger J M, Kuster M S.** The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. J Arthroplasty 2012; 27(3): 430-6 e1. doi: 10.1016/j.arth.2011.06.035

4. **Heijbel S, Naili JE, Hedin A,W-Dahl A, Nilsson KG, Hedstrom M.** The Forgotten Joint Score-12 in Swedish patients undergoing knee arthroplasty: a validation study with the Knee Injury and Osteoarthritis Outcome Score (KOOS) as comparator. Acta Orthop 2020; 91(1): 88-93. doi: 10.1080/17453674.2019.1689327

5. **Wang Z, Deng W, Shao H, Zhou Y, Li** H. Forgotten Joint Score thresholds for forgotten joint status and patient satisfaction after unicompartmental knee arthroplasty in Chinese patients. J Arthroplasty 2020; 35(10): 2825-9. doi: 10.1016/j.arth.2020.05.010

6. **Longo U G, De Salvatore S, Candela V, Berton A, Casciaro C, Sciotti G, et al.** Unicompartmental knee arthroplasty: minimal important difference and patient acceptable symptom state for the Forgotten Joint Score. Medicina (Kaunas) 2021; 57(4). doi: 10.3390/medicina57040324

7. **Clement N D, Scott C E H, Hamilton D F, MacDonald D, Howie C R.** Meaningful values in the Forgotten Joint Score after total knee arthroplasty. Bone Joint J 2021; 103-B(5): 846-54. doi: 10.1302/0301-620x.103b5.bjj-2020-0396.r1

8. **SKAR.** Swedish Knee Arthroplasty Register Annual Report 2020; 2020. ISBN 978-91-88017-31-4.

9. **Terwee C B, Peipert J D, Chapman R, Lai J S, Terluin B, Cella D, et al.** Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021; 30(10): 2729-54. doi: 10.1007/s11136-021-02925-y

10. **Terluin B, Eekhout I, Terwee C B, de Vet H C.** Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol 2015; 68(12): 1388-96. doi: 10.1016/j.jclinepi.2015.03.015.

11. **Terluin B, Eekhout I, Terwee C B.** The anchor-based minimal important change, based on receiver operating characteristic analysis or predictive modeling, may need to be adjusted for the proportion of improved patients. J Clin Epidemiol 2017; 83: 90-100. doi: 10.1016/j.jclinepi.2016.12.015. doi: 10.1016/j.jclinepi.2015.03.015

12. **Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston B C, Devasenapathy N, 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: m1714. doi: 10.1136/bmj.m1714

13. **Ingelsrud L H, Roos E M, Terluin B, Gromov K, Husted H, Troelsen A.** Minimal important change values for the Oxford Knee Score and the Forgotten Joint Score at 1 year after total knee replacement. Acta Orthop 2018; 89(5): 541-7. doi: 10.1080/17453674.2018.1480739

14. **Holtz N, Hamilton D F, Giesinger JM, Jost B, Giesinger K.** Minimal important differences for the WOMAC osteoarthritis index and the Forgotten Joint Score-12 in total knee arthroplasty patients. BMC Musculoskelet Disord 2020; 21(1): 401. doi: 10.1186/s12891-020-03415-x

15. **Kahlenberg C A, Nwachukwu B U, McLawhorn A S, Cross M B,** **Cornell C N, Padgett D E.** Patient satisfaction after total knee replacement: a systematic review. HSS J 2018; 14(2): 192-201. doi: 10.1007/s11420-018-9614-8. doi: 10.1007/s11420-018-9614-8

16. **Schamber E M, Takemoto S K, Chenok K E, Bozic K J.** Barriers to completion of patient reported outcome measures. J Arthroplasty 2013; 28(9): 1449-53. doi: 10.1016/j.arth.2013.06.025