

**DEVELOPMENT OF A PATIENT-REPORTED OUTCOME MEASURE OF ACTIVITY &
PARTICIPATION (THE OKS-APQ) TO SUPPLEMENT THE OXFORD KNEE SCORE**

J Dawson*, DJ Beard*, H McKibbin, K Harris, C Jenkinson, AJ Price

Jill Dawson MA MSc DPhil SRN SCM (Corresponding author), Nuffield Department of Population Health (HSRU), University of Oxford, Rosemary Rue Building, Old Road Campus, Oxford OX3 7LF, Tel: 01865 289423, Fax: 01865 289260 jill.dawson@dph.ox.ac.uk.

Professor David J Beard GDPhys MCSP MA MSc DPhil, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, Tel: 01865 227695, Fax: 01865 227966, david.beard@ndorms.ox.ac.uk

Heather McKibbin MSc MCSP MMACP, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, Tel:01865 737539, heathermckibbin@aol.com

Kristina Harris BSc MSc DPhil MCSP, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, Tel:01865 227617, Fax: 01865 227966, kristina.harris@ndorms.ox.ac.uk

Professor Crispin Jenkinson MA MSc DPhil, Nuffield Department of Population Health (HSRU), University of Oxford, Oxford OX3 7LF, Tel: 01865 289441, Fax: 01865 289260, crispin.jenkinson@dph.ox.ac.uk

Professor Andrew J Price MA DPhil FRCS (Orth), Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Oxford OX3 7LD, Tel: 01865 737539, Fax: 01865 227966, andrew.price@ndorms.ox.ac.

*Jill Dawson and David Beard are co-first authors on this paper.

DEVELOPMENT OF A PATIENT-REPORTED OUTCOME MEASURE OF ACTIVITY & PARTICIPATION (THE OKS-APQ) TO SUPPLEMENT THE OXFORD KNEE SCORE

ABSTRACT

The primary aim of this study was to develop a patient-reported Activity & Participation Questionnaire (the “OKS-APQ”) to supplement the Oxford Knee Score, to assess higher levels of activity and participation. Item generation involved interviews with 26 patients. Psychometric analysis (exploratory & confirmatory factor analysis; Rasch) guided item reduction and scale generation within a prospective study of 122 relatively young patients prior to knee replacement, mean age 61.5 (range 42 - 71) years; n=99 completed pre- and six month post-operative assessments (new items, OKS, SF-36 & American Knee Society Score). The 8-item OKS-APQ scale is unidimensional, reliable (Cronbach’s alpha .85; ICC .79; or .92 when one outlier was excluded), valid ($r>0.5$ with related scales) and responsive (effect size 4.16 – highest of all assessments).

We recommend that the OKS-APQ is used with the OKS, with adults of all ages, where further detail regarding patients’ activity and participation levels is required.

DEVELOPMENT OF A PATIENT-REPORTED OUTCOME MEASURE OF ACTIVITY & PARTICIPATION (THE OKS-APQ) TO SUPPLEMENT THE OXFORD KNEE SCORE

BACKGROUND

The Oxford Knee Score (OKS), patient-reported outcome measure (PROM) was developed in 1998, based on interviews with patients undergoing total knee replacement (TKR)¹. Designed to be used as a composite scale, reflecting patients' perception of knee pain and functional impairment, the OKS has demonstrably good measurement properties for patients having TKR¹⁻³. Since April 2009, OKS data have been collected routinely, throughout the NHS, to monitor clinical quality for patients undergoing TKR⁴.

When the OKS was developed, evidence of implant longevity (approximately 10 years), plus perceptions that younger patients placed relatively high activity demands on an implant, meant that TKR was rarely recommended for people under 65 years old. However, technological change, (including new surgical options for more localized joint damage e.g. patello-femoral and unicompartmental replacement (UKR)), supported by contemporary outcomes data, greatly increased confidence in using arthroplasty with younger patients⁵. Also, many patients with knee problems, of all ages, participate in social and recreational activities (a minority involving high-level activities such as playing sports⁶) that they expect to resume (and improve) following knee replacement⁷. Overall, while alleviation of pain remains the most important driver of both need for, and patient satisfaction with, surgery⁸, patients now have greater expectations and functional requirements compared with twenty years ago.

In response to these societal changes, we conducted a study to develop and validate a new questionnaire: the Oxford Knee Score – Activity & Participation questionnaire (OKS-APQ) to complement the OKS as an additional scale for extending and enriching outcomes assessment in patients with severe knee OA, but especially following knee replacement. The questionnaire was

conceptualized as a concise measure, with a single domain. It was further envisaged that where the OKS is essentially 'disability based'; the new questionnaire would incorporate important additional concerns raised by relatively younger or more active patients regarding higher levels of activity and participation.

MATERIALS AND METHODS

Research Ethics Committee approval was obtained (reference 10/H0606/10) and all subjects consented to participate in the study.

The study employed a three-stage approach consistent with recommendations for best practices:

Stage 1 - Item generation

Participants

Inclusion criteria for interviews (and piloting) were: consecutive patients, age 30 to 70 years, attending orthopaedic or pre-admission outpatient clinics at a regional centre, who were: (i) under consideration for knee replacement, (ii) already listed for TKR or UKR (within the next 4 weeks), or (iii) post-operative (within 2 to 24 months previously) primary unilateral TKR or UKR, with an underlying diagnosis of primary or secondary OA. The sample size for this stage of the study was guided by the point at which no new substantive themes emerged from the interviews. Patients unable or unwilling to give informed consent or lacking fluency in English were excluded.

Interviews and candidate item development

Following recruitment by the study researcher (an Orthopaedic Physiotherapy Specialist in Knees), patients self-completed the OKS. An exploratory, semi-structured interview then followed,

lasting approximately 30 minutes, and employed prompts concerning patients' perceptions of *the impact of limitations imposed by their knee condition on their work/domestic role, family and social interactions, including sports activities and interests*. For pre-operative subjects this included an exploration of roles and activities that they hoped (or expected) to return to, following knee replacement. For post-operative subjects, the interviewer explored patients' feelings about the outcome of their surgery (to date), particularly in relation to physical activity and social participation. Interview prompts were informed by the everyday experience of the clinical members of the research team in addition to relevant research literature.

Interviews were recorded and transcribed; recruitment continued until no new substantive themes emerged (n=26) [supplementary tables S1 & S2 show details of sample characteristics and thematic saturation]. Relevant themes were independently extracted from the transcribed interviews by two members of the research team, who drafted questionnaire items to reflect these themes. A larger group of researchers (co-authors) then scrutinised them for ambiguity and repetition.

This process produced a questionnaire initially containing 29 candidate items which was piloted (self-completed) with new patients in clinics, in 3 different formats associated with 3 alternative response sets (all 5 category Likert). Additional space was provided, in which extra comments/criticisms were encouraged. A brief cognitive interview then followed with the researcher, in order to ascertain the meaning that patients attributed to the items and reasons for their responses. Items were amended as necessary. Piloting continued, involving 50 patients in total, until patients' preferred format was identified. An independent formal assessment of readability and translatability was also conducted¹.

Seventeen candidate questionnaire items were retained following this process (see table 1 for item content). Items rejected included those that many respondents did not find relevant to their situation (such as an item about patients' concern that inactivity, due to their knee problem, was

¹ Conducted by Pharmaquest. A medical translation company specialising in PROMs. Bloxham Mill Business Centre, Oxfordshire, OX15 4FF, UK

making their general health suffer); any items that contained the words 'compared with other people of my age', as patients said that their response would differ according to which individuals they thought about; and an item concerned with pain medication, which many patients refused to use, even when their symptoms were quite severe, while other patients used them for the alleviation of other coexisting conditions.

Stages two & three: Item reduction and scale generation; testing scale properties

– prospective study.

Inclusion criteria were consecutive patients aged between 30 and 70, attending a pre-admission clinic within four weeks prior to primary unilateral TKR or UKR, with primary or secondary OA, who were able to understand consenting procedures, fluent/literate in English.

Sample sizes for questionnaire development typically require at least three respondents per questionnaire item tested¹⁰. However, some loss to follow-up (including due to cancelled/delayed operations) is usual, so the aim was to recruit at least 100 patients, between October 2010 and November 2011.

Assessments

Patient-completed questionnaire that included (i) the OKS^{1,11}, items scored from 0 to 4; summary score range 0 (worst) to 48 (best); ii) The new 17 candidate questionnaire items, each scored 0 (worst) to 4 (best); iii) The SF-36 generic health questionnaire^{12 13}: 36 items, comprising 8 dimensions: physical functioning, social functioning, role physical, role emotional, mental health, energy/vitality, bodily pain and general health perceptions, over the last 4 weeks. Scores for each dimension range from 0 (poor health) to 100 (good health). Summary scores calculated concerning limitations related to physical (Physical Component Summary, PCS) and mental (Mental Component Summary, MCS) factors¹⁴ standardised to general population scales with mean of 50 (SD 10).

Seventy-three patients were given a second copy of the new candidate questionnaire items (with a pre-paid return envelope) to complete 48 hours later (to assess test retest reliability). Not all patients received this due to some interviews having to end quickly as the patients were required to attend appointments. In such cases the request to undertake the test/retest exercise (and its necessary explanation) was set aside.

Clinician assessed American Knee Society Score (AKSS)¹⁵, by the study researcher, comprising a “Knee score” (pain, stability, range of motion) and a “Functional score” (walking distance and stair climbing) each scored 0 (worst) to 100 (best).

The same assessments were repeated, in an out-patient clinic at 6 months post-surgery. Patients who did not attend an appointment were sent the patient’s questionnaire with a letter and pre paid reply envelope.

Statistical analysis

Data were analysed using SPSS release 20 (IBM SPSS statistical software), unless otherwise stated. Non-parametric tests were used in the analysis where data were found to be non-Normally distributed. The significance level for all analyses was set at two-sided $p < 0.05$.

Item reduction

The formal process of item reduction used Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) and Rasch analysis. This was an iterative process. As the number of items reduced, analyses were revisited to confirm the validity of the measure across methods. Due to the highly skewed responses on the pre-operative, but not the post-operative, OKS-APQ items, item reduction criteria were applied chiefly (and given primacy) to the post-operative data.

Exploratory Factor Analysis (EFA)

EFA was conducted with pooled OKS and OKS-APQ candidate items in order to identify items that had clear loadings on a separate factor from the OKS. In EFA, items most strongly interrelated tend to gain high loadings on a single factor suggesting the items are assessing the same underlying concept¹⁶.

EFA was conducted using SPSS 20 with the R-Factor plugin, based on the matrix of polychoric correlations¹⁷. The number of factors was determined where eigenvalues (the proportion of total variance explained by the factor relative to that explained by a single item) were > 1 . Factor extraction used the Principal Axis Factoring extraction, with oblique rotation, method. If an item correlated with a factor of more than 0.3, this was considered a significant loading and the item would be assigned to that factor. ¹⁶

Confirmatory Factor Analysis (CFA)

CFA (using Lisrel V8.80 software) was used to test the fit of the hypothesized factor structure (resulting from the EFA) of the OKS-APQ. χ^2 test was used to assess the fit of the model to the data. The following fit indices were considered satisfactory: (i) Root Mean Square Error of Approximation (RMSEA) (range 0 to 1), <0.05 representing close fit, <0.08 good fit, <0.1 satisfactory fit; (ii) RMSEA p value of close fit $p > 0.05$; (iii) Comparative Fit Index (CFI) > 0.95 ¹⁸.

Rasch analysis

The dimensionality of the OKS-APQ, plus the functioning and fit of individual items, was examined using a Rasch unidimensional measurement model (using RUMM2010 software; RUMM Laboratory Pty Ltd; Western Australia 6023). The Rasch model is equivalent to a test of the theoretical construct validity and adequacy of a scale. It assumes that as a person's disability or symptoms increase, the probability of a maximum score (worst) on the item increases¹⁹.

Criteria for item reduction

Items were excluded from the list of candidate questionnaire items according to pre-determined criteria:

- 1) Non-response greater than 10% (pre- or post-operative assessments);
- 2) Particularly high ceiling or floor effect at the 6 month follow-up assessment (i.e. $\geq 50\%$ of responses taking either of the two most extreme responses i.e. 'strongly agree; or 'strongly disagree'). Pre-operative responses were expected to be highly skewed in the severe direction ('strongly agree'), and more so than the OKS, because the new questionnaire was designed to assess relatively higher level function;
- 3) Clear factor structure (that excluded any item);
- 4) Internal consistency reliability (Cronbach's alpha) improved by item's removal;
- 5) Poor item fit: taking together results of Rasch analysis, EFA and CFA.

6) Differential item functioning (DIF) on gender or procedure (TKR vs UKR) in Rasch analysis.

Once finalised, the OKS-APQ scale was assessed for:

Test retest reliability (repeatability): Intraclass Correlation Coefficients (ICC) comparing pre-operatively assessed OKS-APQ domain scores with measures obtained 48 hours later.

Convergent validity An *a priori* hypothesis was that at least a moderate correlation (Spearman *rho*) would be obtained between the new OKS-APQ measure, the knee-specific OKS and AKSS clinical assessment scores. Generic SF-36 domains were expected to be less highly correlated with the new measure than the knee-specific measures, but the Pain, Physical function, Role Physical and Social Function sub-scales would be the most highly correlated.

Divergent validity Dissimilar SF-36 domains (e.g. Mental health, and General health perception) were expected to be least correlated with the new measure.

Responsiveness was assessed within the cases where outcomes data were provided at both pre- and 6 month post-operative assessments. Mean pre- and 6 month post-operative scores, change scores and effect sizes between OKS-APQ, OKS, AKSS and SF-36 domains were compared. Effect size is a method of calculating the extent of change measured by an instrument in a standardised way that allows comparison between instruments.²⁰ It was calculated as the difference between the sample's mean pre-surgical and 6 month post-surgical scores, divided by the standard deviation (SD) of pre-surgical scores. An effect size of 1.0 is equivalent to a change of one SD in the sample; values of 0.2, 0.5 and 0.8 are typically regarded as indicating small, medium and large degrees of change respectively.

The hypothesis was that at least moderate effect sizes (≥ 0.5) would be obtained by the knee-specific measures and relevant domains of the SF-36 (ie. pain and physical functioning) and that the knee specific measures would be the most responsive.

RESULTS

A sample of 122 consecutive patients was recruited: mean age 61.5 years (SD 6.68 min 42, max 71. One patient was included in error, who had a birthday between letter of invitation and clinic attendance); 53 (43.4%) male. 121 completed study questionnaires, while 114 (93.4%) received a study clinical assessment. Of these, 45 (/114, 39.5%) had a unilateral knee problem, 61 (53.5%) bilateral symptoms and 8 (7%) had multiple/systemic problems that affected walking ability. Ninety-two patients (80.7%) received surgery within the study period and completed follow-up questionnaires at six months, of whom 87 (94.6%) were clinically assessed. Of these (n=92), 38 (41.3%) had had TKR, 53 (57.6%) UKR and one (1.1%) required high tibial osteotomy, due to intra-operative findings. Mean pre- and post-operative OKS scores were 19.36 (SD 7.69) and 33.59 (SD 9.51) respectively (compares with mean values from National HES data: 19.00 and 33.76 respectively²¹).

Table 1 shows patients' pre and post-operative responses to each of the 17 new candidate questionnaire items.

- Table 1 –

Pre-operative responses to all but two of the candidate questionnaire items were highly skewed (as anticipated), with between 57.0% and 84.3% endorsing the most severe response ('strongly agree'). Responses at the follow-up assessment were much more evenly distributed.

Item reduction

Pre- and post-operative response rates were very high: none approaching the '10% or above' criterion. Item 10 was excluded, as 62% of post-operative responses took an extreme (strongly disagree) category. Preliminary EFA of pooled post-operative OKS and candidate OKS-APQ items revealed two factors which explained 76% of variance. The OKS-APQ items had mostly significant loadings on the first factor and the OKS loaded significantly on the second factor. OKS-APQ items 7, 9, 14 and 15 were excluded as they were loading significantly on both factors ($r > 0.3$).

Rasch analysis led to the removal of a further 4 items (items 5, 6, 13 and 16) [see supplementary table S3 and accompanying report for further details] leaving 8 items exhibiting good fit and factor structure.

EFAs of pre-operative, then post-operative, pooled OKS and remaining eight candidate items, revealed two underlying factors which explained 55% and 79% of variance respectively. Table 2 shows factor loadings ($>.3$) for these analyses.

– Table 2–

Internal consistency reliability

The ICR of the pre-operative OKS-APQ scale was examined in SPSS, where the Cronbach's alpha was .85. Item to total correlations (r) ranged from .43 to .73, and removal of any single item did not improve the overall alpha.

CFA and Rasch Fit statistics for the finalised 8-item OKS-APQ (post-operative data).

The unidimensional, 8 item OKS-APQ model was subjected to CFA, which indicated its excellent fit: the χ^2 test was not significant ($\chi^2 = 19.09$, $df=20$, $p=0.52$); RMSEA of 0.0 demonstrated close fit (90% CI 0.0-0.086), p value for close fit of RMSEA was 0.75, and the CFI was 1.

The overall fit of the OKS-APQ 8 items to the Rasch model was examined by assessing the item-trait interaction statistic. Here, the item-trait interaction statistic was (appropriately) non-significant ($\chi^2=6.9$, $df=16$, $p=0.98$). Fit of all eight individual items to the model was excellent, with no items having residuals beyond ± 2 .

None of the OKS-APQ items demonstrated DIF within gender and procedure (TKA vs. UKA) groups and overall, the unidimensionality of the new 8-item measure was supported. [see supplementary material on the Rasch analysis].

Table 3 summarises the results of the item reduction process.

– Table 3–

Testing scale properties of the finalised 8-item OKS-APQ.

The OKS-APQ is scored by summing each individual item score (0 to 4), to produce an overall score of 0 to 32 (lower score representing greater severity). This is then converted to a metric of 0 – 100 [$100/32 \times \text{Actual score}$]. Lower scores represent greater severity.

Test-retest reliability

Sixty-seven (/73) patients returned repeat measures of the OKS-APQ scale items at 48 hours following baseline. Of these, one patient ('outlier') ticked all most extreme ('strongly agree') responses at time 1, but gave the opposite response ('strongly disagree') to all items at re-test. It is conceivable,

but unlikely, that their symptoms had changed in the interim. The ICC (agreement) comparing test and re-test measures was .79 (good) when the outlier was included, and .92 (excellent) when they were excluded.

Convergent and divergent validity

Table 4 shows correlations (Spearman Rho) between the OKS-APQ scale, the clinical AKSS and SF-36 domains.

– Table 4 –

Moderate to high correlations (i.e. $r > 0.5$, shown in bold typeface in table 4) were obtained with pre- and post-operative OKS and the clinical AKSS (post-operative, but not pre-operative) domains, and with the SF-36 Physical functioning, Role Physical, Pain, social functioning (particularly pre-operative) domains, and with the SF-36 PCS score.

Divergent validity of the OKS-APQ was shown by the fairly low ($r < .34$) correlations obtained with the SF-36 Mental health and General health perception domains.

Responsiveness

Table 5 shows the mean pre- and 6 month post-operative scores, change scores and effect sizes of the OKS-APQ, OKS, AKSS, and SF-36 domains for patients with scores at both assessments.

– Table 5 –

Most assessments exhibited highly statistically significant improvements in scores at 6 months, following surgery (exceptions: SF-36 Mental health and health perception domains). Effect sizes were

>1.0 for all knee-specific assessments, (highest: ES 4.16, for the new OKS-APQ), and for 3 of the SF-36 domains (Pain, Physical function and Role-physical), indicating substantial improvement and suggesting that the measures were highly responsive. The generic SF-36 domains had lower effect sizes than the knee-specific measures, indicating that the condition-specific scales were more responsive in this context.

DISCUSSION

Around the time of knee replacement, patients' perceptions of their knee pain and function have been demonstrated to be inextricably linked ²² and their measurement, (as occurs using the OKS), remains relevant and important in this context. Nonetheless, acknowledged changes to population characteristics and technological developments suggest that outcomes assessment may benefit from the additional evaluation of patients' perceptions of higher level knee functioning and participation.

We have here reported the developmental process and validation of a new brief questionnaire: the OKS-APQ, designed to complement the OKS. The methods that we have used are well described, widely accepted and recommended by recent FDA guidelines on PROMs' methodology⁹.

Patients recruited for interview, and those on whom the new measure was tested, were younger than the mean age (70.01 years) at which primary knee replacements are currently received in the UK²³. This sampling criterion was an attempt to obtain people who were - on average - likely to be more active, or to have higher expectations of the eventual outcome following surgery, although we acknowledge that younger individuals are not necessarily more active than older individuals.

Our approach with developing the OKS-APQ did not involve asking questions about specific activities, because no particular activities were prominent in interviews; and because it is known that activities (such as specific sports or interests) that are relevant or important to people undergoing knee replacement are highly variable and sports participation, *per se*, is of interest to only a very small

minority²⁴. Specific activities also vary in the level of intensity with which people approach them and individuals' preferred activities change over time: all of which can be highly problematic when measuring outcomes²⁵. More important, is the extent to which physical activity limitations are perceived as affecting individuals' ability to do the things that are important to them, including their ability to join in with family and friends.

The final 8-item questionnaire was found to be very acceptable to patients, as evidenced by high response rates. It represents one scale that is unidimensional, with a high level of internal consistency and no redundant items. The test-retest reliability was found to be good (excellent following the removal of one outlier). Assessment of convergent and divergent validity showed that the new questionnaire performed well in this regard, obtaining moderate to high correlations with the pre- and post-operative OKS, both post-operative AKSS domains (but not the pre-operative domains), and with relevant SF-36 (Physical functioning, Role Physical, Pain, Social functioning (particularly pre-operative) domains. Low correlations were associated with less relevant domains, such as General health perception. Thus *a priori* hypotheses were generally supported. The lack of correspondence found with the AKSS pre-operative assessment is difficult to explain, although the item content of the two measures address rather different aspects of knee pain and function. Thus, for instance, pre-operatively a patient may have quite a reasonable range of knee movement (AKSS item), but still fail to achieve a walking pace perceived as acceptable (OKS-APQ item).

A more detailed appraisal of responsiveness is beyond the scope of this paper, but based on effect size, the new measure appeared to be even more responsive than either of the other knee specific clinician-assessed or patient-reported measures.

Limitations of our study include the fact that while we believe that the new measure should perform well for assessing the impact of advanced knee disease on activity and participation levels in all adults undergoing knee replacement, we did not assess the measurement properties of this instrument across the full range of applicable ages. This will require further future investigation.

The OKS-APQ is designed to be used alongside the OKS and in similar relevant contexts. It is not designed as a standalone instrument (unlike the OKS). Despite the need for further evidence of acceptable measurement properties of this instrument across the full range of ages applicable to knee replacement, we recommend that the OKS-APQ is used with the OKS, with all adult patients, irrespective of their age, where further detail regarding patients' activity and participation levels is required. This would permit our findings to be compared with results obtained with an average TKR population.

Reference List

- (1) 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(1):63-69.
- (2) Dawson J, Fitzpatrick R, et al. User Manual for the OKS. Isis Innovation Ltd, the University of Oxford. 2010. www.isis-innovation.com.
- (3) Conaghan PG, Emerton M, Tennant A. Internal construct validity of the Oxford Knee Scale: evidence from Rasch measurement. *Arthritis Care and Research* 2007; 57(8):1363-1367.
- (4) Department of Health. Our NHS Our future: NHS next stage review - interim report. London: 2007.
- (5) Carr AJ, Robertsson O, et al. Knee Replacement. *Lancet* 2012; 379(9823):1331-1340.
- (6) Ghomrawi HMK, Ferran NF, et al. How Often are Patient and Surgeon Recovery Expectations for Total Joint Arthroplasty Aligned? Results of a Pilot Study. *HSSJ* 2011; 7:229-234.
- (7) Nilsson AK, Toksvig-Larsen S, Roos EM. Knee arthroplasty: are patients' expectations fulfilled? *Acta Orth* 2009; 80(1):55-61.
- (8) Scott CEH, Howie CR, et al. Predicting dissatisfaction following total knee replacement. *J Bone Joint Surg [Br]* 2010; 92(9):1253-1258.
- (9) US Department of Health and Human Services Food and Drug Administration. Guidance for industry: PRO measures: use in medical product development to support labeling claims. 2009.
- (10) Barrett P, Kline P. The observation to variable ratio in factor analysis. *J Personality and Group Behaviour* 1981; 1:23-33.

- (11) Murray DW, Fitzpatrick R, et al. The use of the Oxford Hip and Knee Scores. J Bone Joint Surg [Br] 2007; 89-B:1010-1014.
- (12) Ware-JE J, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30(6):473-483.
- (13) Jenkinson C, Stewart-Brown S, et al. Evaluation of the SF-36 Version II in the United Kingdom. J Epidemiol Community Health 1999; 53:46-50.
- (14) Ware JE, Kosinski M, Keller S. SF-36 physical and mental summary scales: a user's manual. Boston, Massachusetts: New England Medical Center.; 1994.
- (15) Insall JN, Dorr LD, et al. Rationale of the Knee Society clinical rating system. Clin Orthop Rel Res 1989;13-14.
- (16) Kaiser HF. The application of electronic computers to factor analysis. Educ psychol measurement 1960; 20:141-151.
- (17) Holdago FP, Chae S, et al. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant 2010; 44:153-166.
- (18) Tabachnick BG, Fidell LS, Osterlind SJ. Using multivariate statistics. Boston: Allyn and Bacon; 2001.
- (19) Andrich D. Rasch Models for Measurement. London: Sage Publications; 1988.
- (20) Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989; 27(3 Suppl):S178-S189.

- (21) HES online. Finalised Patient Reported Outcome Measures (PROMs) in England April 2010 to March 2011: Pre- and postoperative data. <http://www.hesonline.nhs.uk> Last accessed 07/05/13.

- (22) Gooberman-Hill R, Woolhead G, et al. Assessing chronic joint pain: lessons from a focus group study. *Arthritis Care and Research* 2007; 57(4):666-671.

- (23) The 9th Annual Report of the NJR for England and Wales Surgical data to 31st December 2011.
http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Reports/9th_annual_report/NJR%209th%20Annual%20Report%202012.pdf. 2012.

- (24) Weiss JM, Noble PC, et al. What functional activities are important to patients with knee replacements? *Clin Orthop Rel Res* 2002; 404:172-188.

- (25) Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 2003; 37:197-206.

Table 1. Response frequencies for 17 candidate items: pre- and 6 months post-knee replacement.

OKS-APQ candidate questionnaire items	Responses	Pre-op N=121	Post-op 6 months N=92
Please consider these statements thinking about the past 4 weeks:		N (%)	N (%)
1. It is a problem for me to do activities (e.g. sports, dancing, walking) to the level I want, because of my knee.	Strongly agree	102 (84.3)	16 (17.4)
	Tend to agree	16 (13.2)	19 (20.7)
	Neither agree nor disagree	1 (0.8)	11 (12.0)
	Tend to disagree	1 (0.8)	25 (27.2)
	Strongly disagree	1 (0.8)	21 (22.8)
2. It is a problem for me to carry heavy things (e.g. items at work, shopping or a child), because of my knee.	Strongly agree	77 (63.6)	10 (10.9)
	Tend to agree	39 (32.2)	20 (21.7)
	Neither agree nor disagree	4 (3.3)	14 (15.2)
	Tend to disagree	1 (0.8)	22 (23.9)
	Strongly disagree	0 (0.0)	26 (28.3)
3. I need to modify my work or everyday activities, because of my knee.	Strongly agree	85 (70.2)	12 (13.0)
	Tend to agree	31 (25.6)	18 (19.6)
	Neither agree nor disagree	1 (0.8)	12 (13.0)
	Tend to disagree	3 (2.5)	20 (21.7)
	Strongly disagree	1 (0.8)	30 (32.6)
4. I need to plan carefully before going out for the day because of my knee (e.g. taking painkillers, using a knee brace or checking that there will be places to sit down).	Strongly agree	71 (58.7)	8 (8.7)
	Tend to agree	38 (31.4)	17 (18.5)
	Neither agree nor disagree	6 (5.0)	8 (8.7)
	Tend to disagree	5 (4.1)	26 (28.3)
	Strongly disagree	1 (0.8)	33 (35.9)
5. It is a problem for me to weave or dodge around people in a crowded place, because of my knee.	Strongly agree	69 (57.0)	10 (10.9)
	Tend to agree	41 (33.9)	15 (16.3)
	Neither agree nor disagree	5 (4.1)	12 (13.0)
	Tend to disagree	4 (3.3)	25 (27.2)
	Strongly disagree	2 (1.7)	30 (32.6)
6. It is a problem for me to stand for a long time, because of my knee.	Strongly agree	97 (80.2)	13 (14.1)
	Tend to agree	16 (13.2)	21 (22.8)
	Neither agree nor disagree	5 (4.1)	9 (9.8)
	Tend to disagree	1 (0.8)	28 (30.4)
	Strongly disagree	2 (1.7)	21 (22.8)
7. It is a problem for me that I experience delayed knee pain or stiffness (i.e. that night or next day) after doing an activity (e.g. walking, sport, work/housework).	Strongly agree	90 (74.4)	16 (17.4)
	Tend to agree	27 (22.3)	31 (33.7)
	Neither agree nor disagree	1 (0.8)	9 (9.8)
	Tend to disagree	3 (2.5)	20 (21.7)
	Strongly disagree	0 (0.0)	16 (17.4)
8. It is a problem for me to fully take part in activities with friends and family, because of my knee.	Strongly agree	72 (59.5)	8 (8.7)
	Tend to agree	30 (24.8)	15 (16.3)
	Neither agree nor disagree	12 (9.9)	15 (16.3)
	Tend to disagree	4 (3.3)	23 (25.0)
	Strongly disagree	3 (2.5)	31 (33.7)
9. It is a problem for me to walk up or down a slope, because of my knee.	Strongly agree	72 (59.5)	10 (10.9)
	Tend to agree	40 (33.1)	18 (19.6)
	Neither agree nor disagree	7 (5.8)	17 (18.5)
	Tend to disagree	2 (1.7)	22 (23.9)
	Strongly disagree	0 (0.0)	25 (27.2)
10. My knee problem can put a strain on my close relationships (partner, family or friends).	Strongly agree	26 (21.7)	4 (4.3)
	Tend to agree	23 (19.2)	2 (2.2)
	Neither agree nor disagree	26 (21.7)	8 (8.7)
	Tend to disagree	22 (18.3)	21 (22.8)
	Strongly disagree	23 (19.2)	57 (62.0)

OKS-APQ candidate questionnaire items	Responses	Pre-op N=121	Post-op 6 months N=92
Please consider these statements thinking about the past 4 weeks:		N (%)	N (%)
11. It is a problem for me to walk at the pace I would like, <u>because of my knee</u> .	Strongly agree	99 (81.8)	13 (14.1)
	Tend to agree	20 (16.5)	19 (20.7)
	Neither agree nor disagree	1 (0.8)	10 (10.9)
	Tend to disagree	1 (0.8)	27 (29.3)
	Strongly disagree	0 (0.0)	23 (25.0)
12. It is a problem for me to twist or turn, as <u>my knee</u> may give way or be painful.	Strongly agree	97 (80.2)	15 (16.3)
	Tend to agree	23 (19.0)	19 (20.7)
	Neither agree nor disagree	1 (0.8)	10 (10.9)
	Tend to disagree	0 (0.0)	26 (28.3)
	Strongly disagree	0 (0.0)	22 (23.9)
13. It is a problem for me that <u>my knee</u> restricts the type of holiday or day trip I can take.	Strongly agree	71 (60.7)	10 (10.9)
	Tend to agree	32 (27.4)	20 (21.7)
	Neither agree nor disagree	6 (5.1)	11 (12.0)
	Tend to disagree	4 (3.4)	25 (27.2)
	Strongly disagree	4 (3.4)	26 (28.3)
14. It is a problem for me to walk on uneven or rough ground, <u>because of my knee</u> .	Strongly agree	88 (73.3)	13 (14.1)
	Tend to agree	30 (25.0)	28 (30.4)
	Neither agree nor disagree	2 (1.7)	12 (13.0)
	Tend to disagree	0 (0.0)	21 (22.8)
	Strongly disagree	0 (0.0)	18 (19.6)
15. Wearing footwear of my choice is a problem for me, <u>because of my knee</u> .	Strongly agree	39 (32.2)	7 (7.7)
	Tend to agree	21 (17.4)	11 (12.1)
	Neither agree nor disagree	36 (29.8)	13 (14.3)
	Tend to disagree	14 (11.6)	22 (24.2)
	Strongly disagree	11 (9.1)	38 (41.8)
16. It is a problem for me to dash across a busy road, <u>because of my knee</u> .	Strongly agree	94 (77.7)	18 (19.6)
	Tend to agree	23 (19.0)	25 (27.2)
	Neither agree nor disagree	2 (1.7)	17 (18.5)
	Tend to disagree	1 (0.8)	16 (17.4)
	Strongly disagree	1 (0.8)	16 (17.4)
17. It is a problem for me that I need to take longer to do everyday activities, <u>because of my knee</u> .	Strongly agree	73 (60.3)	11 (12.0)
	Tend to agree	37 (30.6)	14 (15.2)
	Neither agree nor disagree	9 (7.4)	13 (14.1)
	Tend to disagree	1 (0.8)	27 (29.3)
	Strongly disagree	1 (0.8)	27 (29.3)

Table 2. Results of Exploratory Factor Analysis (EFA) (pre- and post-operative solutions) that included OKS items and the final 8 OKS-APQ items (significant loadings are in bold).

EFA Sorted Pattern Matrix					
PRE-OPERATIVE DATA			POST-OPERATIVE DATA		
	F1	F2		F1	F2
OKS-APQ1	.996	-.203	OKS-APQ 3	.986	-.057
OKS-APQ8	.967	-.218	OKS-APQ 1	.891	.018
OKS-APQ3	.790	.053	OKS-APQ 2	.889	.050
OKS-APQ4	.773	.114	OKS-APQ 8	.770	.170
OKS-APQ11	.747	-.017	OKS-APQ 4	.761	.187
OKS-APQ2	.682	.064	OKS-APQ 17	.755	.178
OKS-APQ17	.560	.248	OKS-APQ 11	.750	.173
OKS-APQ12	.397	.320	OKS-APQ 12	.688	.233
OKS5	-.142	.917	OKS 4	.668	.318
OKS9	.046	.791	OKS 3	-.109	.960
OKS1	-.072	.785	OKS 2	-.016	.910
OKS7	-.126	.780	OKS 8	.031	.829
OKS12	.027	.754	OKS 6	.073	.789
OKS3	-.068	.743	OKS 1	.166	.765
OKS8	-.050	.739	OKS 12	.132	.747
OKS11	.028	.673	OKS 7	-.005	.652
OKS10	-.042	.646	OKS 10	-.008	.645
OKS2	-.122	.641	OKS 9	.333	.613
OKS4	.160	.566	OKS 5	.335	.602
OKS6	.209	.436	OKS 11	.387	.518

Table 3. Item reduction process applied to 17 candidate items.

OKS-APQ (candidate items) <i>Please consider these statements thinking about the past 4 weeks:</i>	1	2	3	4	5	6	Item rejected
1. It is a problem for me to do activities (e.g. sports, dancing, walking) to the level I want, because of my knee.							
2. It is a problem for me to carry heavy things (e.g. items at work, shopping or a child), because of my knee.							
3. I need to modify my work or everyday activities, because of my knee.							
4. I need to plan carefully before going out for the day because of my knee (e.g. taking painkillers, using a knee brace or checking that there will be places to sit down).							
5. It is a problem for me to weave or dodge around people in a crowded place, because of my knee.					X		X
6. It is a problem for me to stand for a long time, because of my knee.					X		X
7. It is a problem for me that I experience delayed knee pain or stiffness (i.e. that night or next day) after doing an activity (e.g. walking, sport, work/housework).		X					X
8. It is a problem for me to fully take part in activities with friends and family, because of my knee.							
9. It is a problem for me to walk up or down a slope, because of my knee.		X					X
10. My knee problem can put a strain on my close relationships (partner, family or friends).	X						X
11. It is a problem for me to walk at the pace I would like, because of my knee.							
12. It is a problem for me to twist or turn, as my knee may give way or be painful.							
13. It is a problem for me that my knee restricts the type of holiday or day trip I can take.					X		X
14. It is a problem for me to walk on uneven or rough ground, because of my knee.		X					X
15. Wearing footwear of my choice is a problem for me, because of my knee.		X					X
16. It is a problem for me to dash across a busy road, because of my knee.					X		X
17. It is a problem for me that I need to take longer to do everyday activities, because of my knee.							
Key* <ol style="list-style-type: none"> 1. ≥50% most extreme response at 6 month follow-up assessment e.g. 'strongly disagree' (ceiling effect – no scope for further improvement). 2. Cross-loading $r > 0.3$ with OKS scale in exploratory factor analysis. 3. Clear factor structure (item has no allegiance to key factor/subscale). 4. Internal consistency reliability (α) improved by item's removal. 5. Poor item fit in Rasch analysis (fit residual $> \pm 2$). 6. Differential item functioning in Rasch analysis (e.g. men vs women, or UKR vs TKR, or Charnley grades A,B,C give significantly different responses) 							

Table 4: Pre and post-operative Spearman rho correlation coefficients between the new OKS-APQ scale, the OKS, clinical AKSS, the generic SF-36 domains, and SF-36 MCS and PCS scores.

	OKS	AKSS	AKSS function	SF-36 domains								SF-36 components	
				Physical functioning	Role Physical	Role Emotional	Social functioning	Mental health	Energy / vitality	Pain	General health	PCS	MCS
OKS-APQ:													
Pre-operative	.538***	.294**	.339***	.552***	.582***	.288**	.557***	.233*	.314**	.496***	.149	.564***	.231*
Post-operative	.807***	.533***	0.509***	.727***	.679***	.512***	.469***	.336**	.462***	.675***	.179	.715***	.243*

High correlation values ($r > 0.5$) are shown in bold typeface.

*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$

Table 5. Instrument responsiveness: mean pre- and 6 month post-operative scores, change scores and effect sizes comparing OKS-APQ, OKS, AKSS clinical and SF-36 domain scores.

	N ¹	PRE-SURGERY		6 MONTHS POST-SURGERY		Mean Change ²	SD	Significance test ³	Effect size ⁴
		Mean	SD	Mean	SD			p	
OKS-APQ (score range 0-100)	92	9.81	12.08	60.02	30.89	50.58	32.05	<0.001	-4.16
OKS (score range 0-48)	87	19.93	7.78	37.54	9.59	17.61	10.09	<0.001	-2.01
AKSS	83	36.05	16.14	84.98	13.26	48.93	21.05	<0.001	-3.03
AKSS Function	83	64.94	15.47	87.11	12.02	22.17	15.19	<0.001	-1.43
SF-36:									
Bodily Pain	89	36.33	19.44	67.54	24.29	-31.21	28.32	<0.001	-1.61
Physical function	80	33.88	20.00	65.81	24.91	-31.94	23.75	<0.001	-1.60
Role - physical	91	40.66	27.14	68.89	27.13	-28.23	33.39	<0.001	-1.04
Mental health	91	76.10	18.24	80.27	16.27	-4.18	15.80	0.01	-0.23
Role - emotional	89	75.94	29.34	87.92	19.22	-11.99	29.19	<0.001	-0.41
Vitality/energy	89	51.97	19.48	59.90	20.70	-51.94	19.48	<0.001	-0.41
Social function	91	58.10	28.10	83.38	24.51	-25.27	27.57	<0.001	0.90
Health perception	83	71.41	19.19	71.33	20.00	0.08	13.82	0.96	0.00

¹The n represents the number of cases with complete data provided at both pre- and post-operative assessments to permit the evaluation of scales and thus change in pain, function, overall health-related quality of life.

²The positive/negative direction of the sign relates to the direction in which each scale is scored to denote increasing severity and is irrelevant to the size of the effect.

³Paired t-tests

⁴Calculated as mean change/SD of pre-surgery score.

FUNDING

This work was supported by an internal source of funding [details withheld to ensure blinding for the peer review process].

ACKNOWLEDGMENTS

The research team would like to thank all study participants for assisting us with this study.

Supplementary material/

**Table S1. Characteristics of interview sample guided by
purposive sampling framework*.**

n=26

Sex		
	male	12
	female	14
Age		
	44 to <55	9
	55 to <65	14
	65-68	3
Pre-operative (n=16)		
	Being considered for KR	10
	Within 4 weeks of KR	6
Post-operative (n=10)		
	Uni	6
	TKR	4
Has current bilateral knee problems		3

*A purposive sampling framework guided recruitment to ensure that examples of males and females, including pre-operative ('under consideration' or 'listed') and post-operative subjects (both UKR and TKR) would be represented.

Table S2: Saturation (no new substantive themes forthcoming within interviews).

Code name/theme	INTERVIEW NUMBER																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Knee pain - continuous	x	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		✓		
- Description of severity	x	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		✓		
- Intermittent				x	✓		✓									✓			✓	✓			✓		✓	✓
Low tolerance of physical activities	x	✓	✓	✓			✓	✓	✓		✓	✓		✓		✓	✓		✓	✓		✓	✓	✓	✓	✓
Rough or uneven ground			x				✓				✓			✓		✓				✓						✓
Avoids twisting & turning	x	✓						✓		✓	✓			✓		✓			✓					✓		✓
Stiffness – after inactivity	x						✓			✓					✓				✓	✓		✓		✓	✓	
Knee locks or gives way	x		✓					✓			✓		✓						✓	✓			✓			
Paying later for today's activities	x			✓					✓					✓		✓			✓	✓			✓		✓	
Standing in one place									x	✓		✓	✓				✓		✓	✓		✓	✓	✓		✓
Bending/ crouching/ squatting		x	✓	✓	✓			✓		✓	✓	✓			✓	✓	✓		✓				✓			
Walking up/down a slope																x		✓	✓							
All of life affected	x		✓		✓				✓		✓	✓						✓	✓			✓				
Inconvenience	x							✓	✓			✓			✓		✓	✓	✓				✓			✓
Frequent need to sit down		x						✓	✓	✓		✓	✓				✓		✓					✓		
Need to adapt		x						✓				✓	✓		✓		✓		✓	✓				✓	✓	
Forward planning	x															✓	✓		✓				✓	✓		
Restrictions on foot wear																x										
Impact on holidays/day trips		x	✓						✓	✓						✓	✓		✓							✓
Reliance on painkillers	x				✓				✓	✓	✓	✓	✓		✓	✓			✓	✓	✓			✓		
Effect on general health	x	✓	✓									✓														
Reduced participation with family/ friends				x				✓	✓	✓		✓	✓				✓	✓	✓							✓
Loss of autonomy		x		✓							✓	✓	✓	✓	✓		✓			✓						
Loss of confidence /feels inadequate		x	✓	✓				✓		✓	✓						✓		✓				✓	✓		
Feeling prematurely old		x	✓	✓				✓		✓							✓		✓					✓	✓	
Worry about falling			x					✓								✓			✓				✓			
Knee swelling		x			✓	✓	✓	✓						✓					✓							
Noises made by knee joint		x	✓	✓				✓				✓				✓										

First mentioned x Mentioned again ✓

Table S3. Rasch analysis on final list of 8 items in the OKS-APQ scale, showing item location and fit statistics.

Item	Location*	SE	Fit Residual	DF	X ₂	DF	Probability
8	-0.603	0.158	0.100	61.75	1.731	2	0.42
4	-0.575	0.154	0.215	61.75	0.396	2	0.82
17	-0.147	0.155	0.194	61.75	0.238	2	0.89
3	-0.050	0.146	0.041	61.75	0.397	2	0.81
2	-0.036	0.154	-0.505	61.75	1.684	2	0.43
11	0.276	0.151	0.424	61.75	0.224	2	0.89
12	0.461	0.147	1.950	61.75	1.840	2	0.40
1	0.674	0.147	0.614	61.75	0.410	2	0.82

* Location represents level of 'difficulty' of an item across the continuum of the underlying trait/construct. Item 8 was the 'easiest' items in the scale (where participants are most likely to not consider the activity related to in the question to be a problem). Item 1 was the most difficult item.

Report on Rasch analysis.

Post-operative data were fitted to the Rasch model to examine the comprehensive fit of the OKS-APQ scale and its individual items to the Rasch model. Targeting of the scale, ordering of item thresholds, DIF of individual items, unidimensionality of the measure, and internal consistency reliability were also examined. It was assumed that the distances between the response category thresholds would vary, so the (unrestricted) partial credit model was used.

Further item reduction to final 8 items (following application of other procedures): Rasch analysis on 12 items not already rejected showed no items had item fit residual above the pre-defined criterion of ± 2 and none of the items demonstrated DIF. However, analysis of residual correlations demonstrated that item 13 (relating to holidays or day trips) and 11 (relating to achievable walking pace) had residual correlations more than 0.3 (0.305). As item 13 also had the largest fit residual it was removed from the analysis. The removal of this item improved fit indices of the overall score. New analysis demonstrated that item 6 (relating to standing) had large fit residual and this item was removed from further analysis. The remaining 10 item solution had a correlation between residuals of items 2 (carrying heavy items) and 16 (dashing across the road) of 0.302, so item 16 was excluded from further analysis due to its biggest fit residual. Similarly, Rasch analysis on remaining 9 items demonstrated that items 5 (dodging around people in crowded places) and 1 (doing activities to the desired level) had a high residual correlation and removal of item 5 (which had worse fit indices) demonstrated the best fit on Rasch analysis and EFA (when repeated) compared with the 12 item questionnaire.

Table 3 (above) summarises the final list of items in the OKS-APQ scale resulting from the item reduction process, showing item location and fit statistics resulting from the item reduction process. The average location of the sample was 0.41 (SD 1.86), which indicated that the measure was reasonably targeted, with the post-operative average of the patients in the sample at the slightly higher level of the construct.

Figure 1 (below) shows item response thresholds for the eight items comprising the OKS-APQ scale. The X-axis denotes the position on the underlying latent trait/variable and the Y-axis denotes the probability of a response category being observed. Most of the 8 items have 'disordered' thresholds (also called step calibrations); the exceptions are items 2, 8 and 17. This is not to be confused with disordered categories (which would indicate that the ordinal numbering of response categories did not accord with their substantive meaning) and, simply relates to the central 'neutral' response category being the least likely response option to be endorsed.

Reliability of the OKS-APQ was examined by the Person Separation Index (PSI), which estimates how well the scale performs in distinguishing between participants with different levels of trait. PSI is interpreted as the Cronbach's alpha, with minimum values of 0.7 required for the group use and 0.85 for the individual use. The PSI of OKS-APQ was 0.92.

Further examination of the unidimensionality of the OKS-APQ was performed using principal component analysis of the residuals. Items were divided into two groups (the ones that loaded negatively or positively on the first component) and independent t-tests were performed between the groups (for each person). If more than 5% tests were beyond the ± 1.96 range, this was taken as evidence of unidimensionality. In fact only 4.35% of the cases had significant t-tests outside the ± 1.96 range, which supported the unidimensionality of the measure.

Figure S1. Rasch analysis: item response thresholds for the eight items comprising the OKS-APQ scale.

Figure S1. Rasch analysis: item response thresholds for the eight items comprising the OKS-APQ scale.

