Osteoarthritis and Cartilage 27 (2019) 762–770

Osteoarthritis and Cartilage

A 12-item short form of the Knee injury and Osteoarthritis Outcome Score (KOOS-12): tests of reliability, validity and responsiveness

B. Gandek † ‡ *, E.M. Roos §, P.D. Franklin †, J.E. Ware Jr. † ‡

† University of Massachusetts Medical School, Worcester, MA, USA
‡ John Ware Research Group, Watertown, MA, USA
§ Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Article history:
Received 17 April 2018
Accepted 23 January 2019

Keywords:
KOOS
Knee
Osteoarthritis
Patient-reported outcome measures
Psychometrics

SUMMARY

Objective: To evaluate reliability, validity and responsiveness of KOOS-12, a 12-item short form of the 42-item Knee injury and Osteoarthritis Outcome Score (KOOS) that provides Pain, Function and Quality of Life (QOL) scale scores and a summary knee impact score.

Design: Data from 1,392 knee osteoarthritis (OA) patients from the FORCE-TJR research cohort who completed KOOS before and 6 and 12 months after total knee replacement (TKR) were analyzed. KOOS-12 includes a pain frequency item and three items measuring pain during increasingly difficult (sitting/lying, walking, stairs) activities; function items about standing, rising from sitting, getting in/out of a car, and twisting/pivoting; and the 4-item KOOS QOL scale. Percent computable scale scores, floor and ceiling effects, internal consistency reliability, validity (scale correlations, tests of known groups validity using one-way analysis of variance (ANOVA)) and responsiveness (effect sizes, standardized response means) were compared for the KOOS-12, full-length KOOS, KOOS-PS and KOOS, JR.

Results: Internal consistency reliability was above 0.70 for all KOOS-12 scales and ≥0.90 for the KOOS-12 Summary score. Validity and responsiveness of KOOS-12 Pain, Function and QOL scales was satisfactory and reached similar conclusions as comparable full-length KOOS scales. The KOOS-12 Summary score was most responsive in discriminating between groups who differed in global ratings of post-TKR change in physical capabilities and had the highest effect sizes and standardized response means.

Conclusions: KOOS-12 was a reliable and valid alternative to KOOS in TKR patients with moderate to severe OA and provided three domain-specific and summary knee impact scores with substantially reduced respondent burden.

Introduction

Knee-specific patient-reported outcome measures (PROMs) provide important information from the patient about the impact of knee osteoarthritis (OA) and other knee disorders and their treatment effectiveness.[1] During the past decade, knee-specific PROMs have increasingly been collected in registries to monitor outcomes after total knee replacement (TKR)[2] and anterior cruciate ligament surgery.[3] One of the most frequently-used knee-specific PROMs is the Knee injury and Osteoarthritis Outcome Score (KOOS) which was developed for adults who have knee OA or knee injuries.[4,5] However, at 42 items, KOOS is often viewed as too lengthy for routine use in clinical care or registries. Two short forms have been derived from the KOOS, but both have limitations. KOOS-PS[6] is a 7-item form that only measures physical function, while KOOS, JR[7] is a 7-item form that includes items measuring symptoms, pain and function but only provides a summary score. Because treatment can vary for persistent knee pain as opposed to functional limitations, a short form that does not provide the specificity of separate pain and function measures is less useful clinically. In addition, neither short form measures knee-specific quality of life (QOL), which has been shown to be highly responsive to treatment.[8,9] In clinical settings, a brief but comprehensive knee-specific PROM that is reliable, valid and responsive, while providing domain-specific measures along with an overall knee impact score, would be optimal.

* Address correspondence and reprint requests to: B. Gandek, University of Massachusetts Medical School, Department of Orthopedics and Physical Rehabilitation 55 Lake Avenue North Worcester, MA 01655, USA.
E-mail addresses: barbara.gandek@umassmed.edu (B. Gandek), eroos@health.sdu.dk (E.M. Roos), patricia.franklin@umassmed.edu (P.D. Franklin), john.ware@jwrginc.com (J.E. Ware).

https://doi.org/10.1016/j.joca.2019.01.011
1063-4584/

KOOS-12 is a new 12-item short form that reduces respondent burden by 70% from the full-length KOOS and provides domain-specific scores for pain, function, and knee-specific QOL while representing content across domains sufficiently to support construction of a summary measure of overall knee impact. It contains 12 items from the KOOS, including four Pain items, four Function items, and four QOL items, which are scored as three scales and an overall KOOS-12 Summary score (Fig. 1). Selection of KOOS-12 items was based on item content; coverage of a wide range of measurement and high item information in item response theory (IRT) models; computerized adaptive test (CAT) simulations to identify items that best matched patients' pain and function levels; scale-level internal consistency reliability, validity and responsiveness; and qualitative input from patients, clinicians, and translation developers. The process of selecting items for the KOOS-12, along with item selection for a 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12), is described in detail in a separate paper.[10] This paper evaluates the reliability, validity and responsiveness of KOOS-12 and compares KOOS-12 psychometric properties to those of the full-length KOOS and its derivative forms. A companion paper examines psychometric properties of HOOS-12.[11]

Methods

Study design and participants

Data came from the Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) research cohort, which has more than 30,000 patients of 200 diverse surgeons throughout the U.S.[12] FORCE-TJR surveys were completed by patients pre-TJR and 6 and 12 months post-TJR, either as paper-pencil surveys or on the Internet, at their surgeon's office or at home. Data from a random sample of $n = 1{,}395$ knee OA patients who had a TKR between 2011 and 2014 (Item Selection sample) was used to select items for KOOS-12.[10] An independent random sample of $n = 1{,}392$ knee OA patients who had a TKR between 2011 and 2014 (Cross-Validation sample) was analyzed in this paper, to independently evaluate the psychometric properties of KOOS-12. FORCE-TJR and this study were approved by the University of Massachusetts Medical School Institutional Review Board.

The Cross-Validation sample had a mean age of 66.5 (SD = 8.9, range 37–100), and 68.2% were female. 84.5% were White non-Hispanic, 10.1% Black non-Hispanic, 2.6% Hispanic, and 2.8% of other race and ethnicity. Overall, 25.4% were a high school graduate or had less education, 27.7% had some education post-high school, 19.3% were college graduates, 22.6% had some post-college education, 2.3% had other education and 2.7% were missing. Sociodemographic characteristics were similar to those of the Item Selection sample.[10]

Measures

KOOS-12 contains three domain-specific scales that measure Pain (number of items k = 4), Function (k = 4) and knee-specific Quality of Life (k = 4) (Fig. 1). This breadth of content is in line with recommendations from OARSI, OMERACT and ICHOM regarding core outcome domains for knee OA patients.[13–16] KOOS-12 scales are scored using the method of summated ratings[17] in which item responses in a scale are simply summed. Scale scores calculated using this method and using more complex IRT-based scoring correlated 0.98 and had similar known groups validity (see Methods). Therefore, the summated ratings method was adopted for scoring KOOS-12. A person-specific value is imputed for missing item data within a scale, if $\geq$ 50% of items in the scale are answered. To facilitate interpretation, scale scores are transformed so 0 is the worst and 100 is the best possible score. This is similar to the full-length KOOS scales, which are also scored using the method of summated ratings and a similar imputation method for missing item data and transformed to a 0 to 100 scale. KOOS provides Pain (k = 9), Symptoms (k = 7), Function in Activities of Daily Living (ADL) (k = 17), Function in Sport/Recreation (k = 5), and QOL (k = 4) scale scores.[18]

The KOOS-12 summary knee impact score (KOOS-12 Summary score) was calculated as the average of the KOOS-12 Pain, Function and QOL scale scores. To compare methods for constructing a summary measure from the three KOOS-12 scales, a principal components analysis of their inter-correlations was conducted on the Item Selection data, to determine if loadings for each scale were equivalent or if scale scores needed to be standardized and weighted prior to calculating a summary score. Loadings were equivalent and substantial (0.928–0.940) across the three KOOS-12 scales, indicating that each scale contributed equally to measuring the higher-order summary construct of knee impact. Correlations between a summary score calculated as the simple average of the three scale scores vs a summary score calculated using weighted scale scores were substantial ($r = 0.999$) at pre-TKR and at 6 and 12 months post-TKR. Therefore, the KOOS-12 Summary score was calculated using the simpler method of averaging the three KOOS-12 scale scores. The KOOS-12 Summary score also ranges from 0 to 100, where 0 is the worst possible and 100 is the best possible score. A summary score is not calculated if any of the three scale scores are missing.

Two additional knee-specific measures constructed from the original KOOS by others were scored for comparative purposes (Fig. 1). KOOS-PS is a 7-item measure of physical function that includes four ADL and 3 Sport/Recreation items and was developed as an OARSI/OMERACT initiative.[6] KOOS, JR contains four Pain, 2 ADL and 1 Symptom items; it is scored to provide an overall measure of knee health and does not have separate Pain, ADL or Symptom scores.[7] KOOS, JR has been included by the U.S. Centers for Medicare & Medicaid Services as one of the PROMs in its Comprehensive Care for Joint Replacement model.[19] KOOS-PS and KOOS, JR were scored following their developers' methods, which require that all items in a scale must be answered to calculate a score;[7,20] the worst and best possible scores are 0 and 100, respectively.

To evaluate construct validity, KOOS-12 scales were evaluated in relation to the SF-36 Health Survey, a general measure of physical and mental functioning and well-being.[21] SF-36 is scored as eight scales, including scales that measure physical function (number of items k = 10), bodily pain (k = 2) and mental health (k = 5), and as two summary Physical (PCS) and Mental (MCS) Component Scores.[22] SF-36 (Version 2.0) scales were scored so that 50 was the mean and 10 was the standard deviation in the U.S. general population.[23]

Statistical analysis

Analyses focused on comparing the psychometric properties of KOOS-12 with those of the full-length KOOS from which the KOOS-12 was derived. In addition, properties of KOOS-PS and KOOS, JR were evaluated to provide information about their performance in relation to KOOS-12.

The percent of respondents for whom scale scores could be computed pre-TKR and 6 and 12 months post-TKR was examined. Internal consistency reliability of all scales was evaluated with Cronbach's coefficient alpha, which is based on the number of items in a scale and the mean inter-item correlation.[24] An alpha of 0.70 or higher is recommended for group-level comparisons, while a

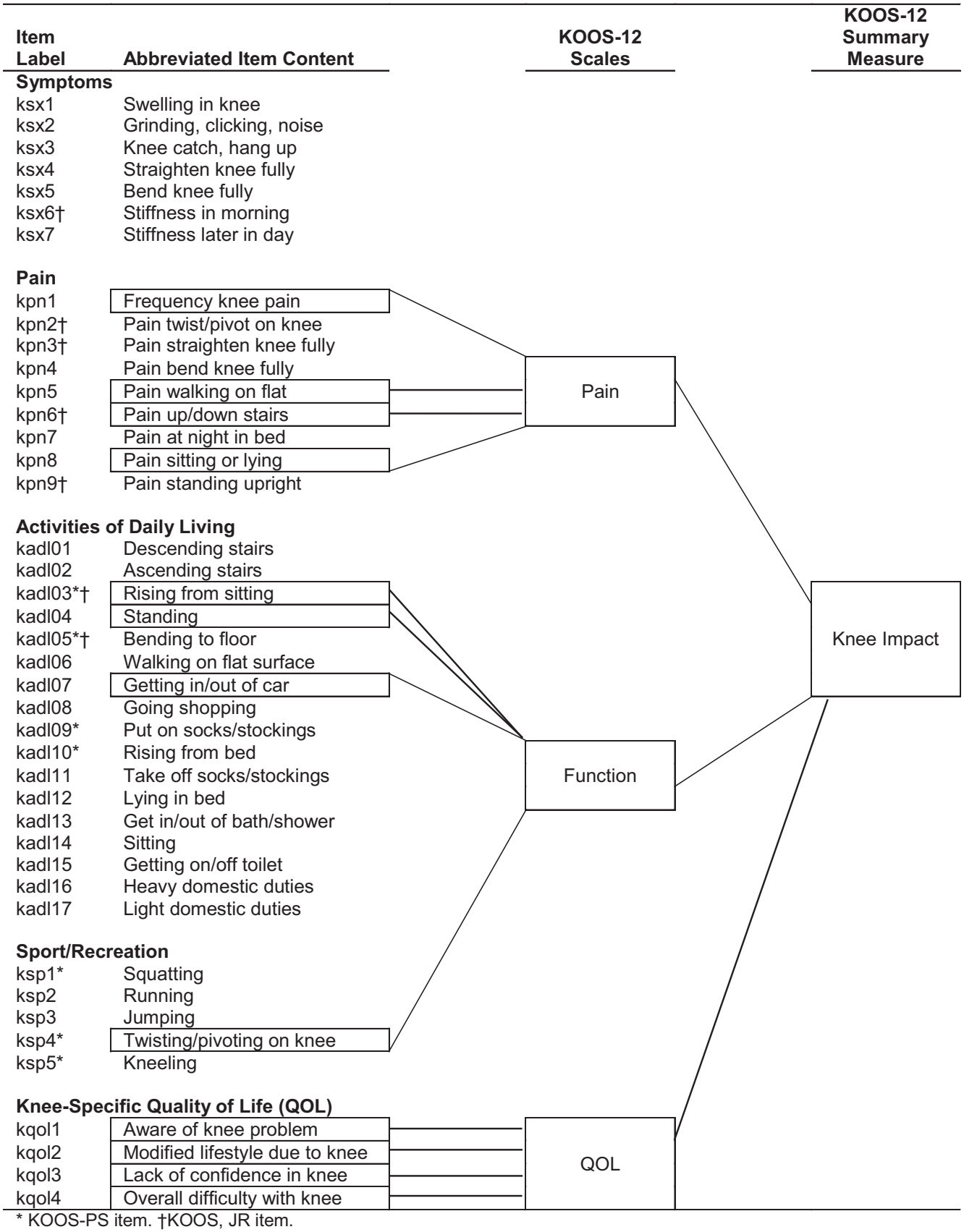

Fig. 1. KOOS-12 measurement model.

minimum reliability of 0.90–0.95 is considered acceptable for individual patients.[25,26] Reliability of the KOOS-12 Summary score was calculated based on Cronbach's alpha for the three KOOS-12 scales, their component weights and their covariances, using methods similar to those used to calculate reliability for SF-36 PCS and MCS.[22] In addition, because responsiveness of a measure is constrained if a high percentage of patients have the lowest or highest possible scores, floor and ceiling effects were evaluated and considered present if more than 15% of respondents had the lowest or highest possible scale scores, respectively.[27] However, because many patients could be viewed as "disease-free" after successful surgery, some ceiling effects were expected post-TKR, as in previous studies.[8]

Validity was evaluated in several ways. Construct validity was examined by estimating Pearson product–moment correlations between KOOS-12 scales and the full-length KOOS and general SF-36 measures, to determine if the KOOS-12 scales had higher correlations with measures of the same construct (convergent validity) than with measures of different constructs (discriminant validity). Correlations <0.30, 0.30–0.69, and ≥0.70 (equivalent to shared variances of approximately <10%, 10% to <50%, and ≥50%) were considered as low, moderate and high, respectively. Because items in the KOOS-12 Pain and Function scales are subsets of the full-length KOOS Pain, KOOS ADL and Sport/Recreation scales, high correlations between KOOS-12 scales and full-length KOOS scales in the same domain were expected. Moderate to high correlations were expected between KOOS-12 scales and SF-36 physical health measures and low correlations between KOOS-12 scales and SF-36 mental health measures.

Known groups validity of the KOOS-12 and other knee-specific measures was compared using one-way analysis of variance (ANOVA), in terms of the measures' responsiveness to patient-reported global change in capability to do everyday physical activities at 6 months compared to pre-TKR (lot more, more, same, less/lot less capable). In each ANOVA, the change score for the scale was the dependent variable, while the self-evaluated change in capability was the independent variable. The ANOVA F-statistic indicated how strongly a scale discriminated between global change groups and provided information about that scale's statistical efficiency. To aid in comparisons, results were summarized across scales using relative validity (RV) statistics (ratio of the F-statistic for each scale divided by the F-statistic for the full-length KOOS scale in domain-specific (pain or function) comparisons and divided by the KOOS-12 Summary score in comparisons across all scales), as in previous analyses.[28] In each set of comparisons, the denominator scale had RV = 1.0; 95% confidence intervals were derived using empirical bootstrap.[29] Within Pain and Function domains, the null hypothesis of equal validity of KOOS-12 and full-length KOOS scales was tested. In addition, because the KOOS-12 summary measure combines multiple indicators of joint impact, it was hypothesized to be the most valid of all measures. KOOS, JR also provides a summary score, but its Pain and Function items do not cover as wide a measurement range as KOOS-12[10] plus KOOS, JR does not include QOL items.

As a measurement property, responsiveness is best interpreted in relation to another measure captured simultaneously[30] using an approach such as the anchor-based method described above.[31,32] In addition, responsiveness of all knee-specific scales and summary measures was compared using the standardized effect size (ES; observed change score (post-minus pre-TKR) divided by the standard deviation of the pre-TKR score)[33] and the standardized response mean (SRM; observed change score divided by the standard deviation of the change score).[34]

All analyses were performed using Stata Statistical Software: Release 11 (StataCorp LP, College Station, TX). Two-tailed tests were used to determine significant ($P < 0.05$) differences.

Results

KOOS-12 Pain, Function and QOL scale scores could be calculated for 98.7–99.1%, 97.3–98.9% and 97.7–98.1% of patients (Table I). These percentages were similar for the KOOS-12 Pain and full-length KOOS Pain scales. The percent computable was similar for KOOS-12 Function and KOOS ADL (97.4–98.7%), but somewhat lower for KOOS Sport/Recreation (94.6–97.3%). The KOOS-12 Summary score could be calculated for 95.9–96.6% of patients. Scores for KOOS-PS could only be computed for 88.3–93.0% of patients, because the developer's scoring algorithm requires that all seven items must be answered and the squatting and kneeling items were missing 4–6% of the time post-TKR. KOOS, JR scores could be computed for only 90.8–92.0% of patients, also because all items must be answered to calculate a score.

Internal consistency reliability of all KOOS-12 scales was above 0.70 at pre-TKR and both post-TKR time points (Table I). Cronbach's alpha was lower for KOOS-12 Pain and Function than for corresponding full-length KOOS scales. However, average inter-item correlations did not differ greatly between the KOOS-12 Pain and KOOS Pain scales or the KOOS-12 Function and KOOS ADL scales, indicating that differences in reliability were primarily due to differences in scale length.[35] Reliability of the KOOS-12 Summary score was 0.90–0.93 and thus met the minimum level recommended when measures are used with individual patients.

Floor effects (percent with the lowest (worst) possible score) for all measures were very low (<1%) pre- and post-TKR, with the exception of the KOOS Sport/Recreation and QOL scales, which had

Table I
Percent computable scales, internal consistency reliability and average inter-item correlations for knee-specific measures

	k	% Computable			Cronbach's alpha			Inter-Item Correlation		
		Pre	6 m	12 m	Pre	6 m	12 m	Pre	6 m	12 m
KOOS-12 Pain	4	99.1	98.7	98.8	0.75	0.79	0.82	0.43	0.48	0.53
KOOS Pain	9	99.1	98.6	98.6	0.88	0.90	0.91	0.45	0.50	0.53
KOOS-12 Function	4	97.3	98.7	98.9	0.78	0.78	0.82	0.47	0.47	0.53
KOOS ADL	17	97.4	98.7	98.6	0.95	0.95	0.96	0.53	0.53	0.59
KOOS Sport/Recreation	5	97.3	94.8	94.6	0.88	0.88	0.90	0.59	0.59	0.64
KOOS-PS	7	93.0	89.8	88.3	0.84	0.84	0.86	0.43	0.43	0.47
KOOS/KOOS-12 QOL	4	98.1	98.0	97.7	0.80	0.83	0.84	0.44	0.49	0.51
KOOS-12 Summary	12	96.6	96.0	95.9	0.90	0.92	0.93	–	–	–
KOOS, JR	7	91.7	90.8	92.0	0.85	0.87	0.89	0.45	0.49	0.54

$N = 1{,}392$ (pre-TKR), 985 (6 month post-TKR), 814 (12 month post-TKR).
k, Number of items; % Computable, Percent of respondents for whom scale score could be computed at pre-TKR, 6 month and 12 months post-TKR; ADL, Activities of Daily Living; QOL, Quality of Life.

Table II
Floor and ceiling effects for knee-specific measures

	k	% at Floor			% at Ceiling		
		Pre-TKR	6 month	12 month	Pre-TKR	6 month	12 month
KOOS-12 Pain	4	0.4	0.2	0.2	0.8	22.3	32.4
KOOS Pain	9	0.2	0.2	0.2	0.6	14.6	25.0
KOOS-12 Function	4	0.8	0.2	0.0	0.2	14.2	23.3
KOOS ADL	17	0.2	0.0	0.0	0.4	8.7	15.7
KOOS Sport/Recreation	5	22.7	3.3	3.3	0.2	2.7	4.3
KOOS-PS	7	0.4	0.2	0.0	0.0	3.1	5.0
KOOS-12/KOOS QOL	4	9.1	0.8	0.8	0,0	7.6	11.6
KOOS-12 Summary	12	0.4	0.0	0.0	0.0	4.5	6.6
KOOS, JR	7	0.4	0.2	0.0	0.2	8.9	16.7

$N = 485$ patients with all scale scores at all three time points.
k, Number of items; ADL, Activities of Daily Living; QOL, Quality of Life; % Floor, % with lowest possible score; % Ceiling, Percent with highest possible score. All measures scored so 0 = worst possible and 100 = best possible score.

floor effects of 23% and 9% before TKR (Table II). Ceiling effects (percent with the highest (best) possible score) were negligible before TKR. Post-TKR, there were notable ceiling effects for KOOS-12 Pain (22–32% at 6 and 12 months) and KOOS Pain (15–25%). For Function, post-TKR ceiling effects generally were low, but were 23% for KOOS-12 Function and 16% for KOOS ADL at 12 months post-TKR. Post-TKR ceiling effects were low for the KOOS-12 Summary score (4–7%) but higher for KOOS, JR (9–17%), indicating that the KOOS-12 Summary score was better at distinguishing between patients at the highest measured levels of knee health.

Tests of construct validity, which supported the convergent and discriminant validity of KOOS-12 with generally similar results at all three time points, are presented for pre-TKR data (Table III). The correlation of KOOS-12 Pain and KOOS Pain was very high ($r = 0.93$), indicating that all reliable variance in the KOOS Pain scale was captured by the KOOS-12 Pain scale. KOOS-12 Function had a high correlation with KOOS ADL ($r = 0.89$) but only a moderate correlation with KOOS Sport/Recreation ($r = 0.61$), pre-TKR. The correlation of the KOOS-12 Function and KOOS Sport/Recreation scales increased to $r = 0.69$ and $r = 0.74$ at 6 and 12 months post-TKR, respectively. KOOS-12 Pain, Function and QOL scales had similar patterns of moderate correlations with SF-36 scales primarily measuring physical health (Bodily Pain, Physical Functioning, Physical Component Summary) and, in support of discriminant validity, low correlations with SF-36 scales primarily measuring mental health (Mental Health, Mental Component Summary).

Tests of known groups validity indicated that all knee-specific measures were responsive to differences between groups varying in global change in capability to do everyday physical activities at 6 months post-TKR (Table IV). Within the Pain domain, the KOOS-12 scale (RV = 0.87, 95% CI (0.72, 1.01)) was not significantly ($P < 0.05$) less responsive in detecting group differences than the full-length KOOS Pain scale (RV = 1.0), as hypothesized. Within the Function domain, RV for KOOS-12 Function, KOOS ADL, KOOS Sport/Recreation and KOOS-PS did not differ significantly. Across all measures, the KOOS-12 Summary score (RV = 1.0) was significantly ($P < 0.05$) more responsive than KOOS, JR (RV = 0.70, 95% CI (0.58, 0.83)) in detecting differences between groups, as hypothesized. Comparison of mean change scores indicated that the KOOS-12 Summary score detected about a three-quarter SD greater improvement on average than KOOS, JR for those who rated themselves as most improved.

Effect sizes (ES) at 6 and 12 months post-TKR were slightly higher for KOOS-12 Pain than KOOS Pain, while standardized response means (SRM) were similar (Table V). ES were similar for KOOS-12 Function, KOOS ADL and KOOS Sport/Recreation, although KOOS Sport/Recreation had a lower SRM than the other function measures. KOOS-PS had the lowest ES of all measures. The ES for the QOL scale were the highest of the three KOOS-12 scales. ES for the KOOS-12 Summary score (2.52–2.70 at 6 and 12 months) were higher than the ES for KOOS, JR (2.04–2.29).

Table III
Descriptive Statistics, Reliability and Inter-scale correlations Among Knee-Specific and SF-36 Measures, pre-TKR

	k	Mean	SD	Reliability	Pain		Function			QOL
					KOOS-12	KOOS	KOOS-12	KOOS ADL	KOOS Sport	KOOS QOL
Pain										
KOOS-12 Pain	4	42.9	16.3	0.75						
KOOS Pain	9	46.6	17.3	0.88	0.93					
Function										
KOOS-12 Function	4	42.9	18.2	0.78	0.71	0.74				
KOOS ADL	17	52.6	17.9	0.95	0.78	0.79	0.89			
KOOS Sport/Recreation	5	17.3	17.8	0.88	0.43	0.47	0.61	0.48		
KOOS-12/KOOS QOL	4	25.9	17.1	0.80	0.56	0.59	0.57	0.58	0.49	
SF-36 Bodily Pain	2	34.6	7.4	0.81	0.65	0.66	0.61	0.66	0.39	0.56
SF-36 Physical Functioning	10	31.1	8.9	0.87	0.52	0.51	0.56	0.61	0.43	0.47
SF-36 Mental Health	5	50.4	9.8	0.84	0.27	0.28	0.26	0.30	0.17	0.28
SF-36 PCS	35	32.7	8.3	0.91	0.51	0.50	0.54	0.57	0.42	0.49
SF-36 MCS	35	52.7	11.5	0.93	0.28	0.29	0.29	0.33	0.15	0.26

$N = 1,303$.
k, number of items. ADL, Activities of Daily Living; Sport, Sport/Recreation; QOL, Quality of Life; PCS, Physical Component Summary; MCS, Mental Component Summary. Reliability is internal consistency reliability (Cronbach's alpha), see text.
SE for all correlations = 0.028. All measures scored so 0 = worst possible and 100 = best possible score, except for SF-36 measures (US general population mean = 50, SD = 10; lower score = poorer health).

Table IV
Mean change scores (SD) and known-groups validity tests by self-evaluated change in physical activity at 6 months

	k	Mean (SD) Change Score by Change in Capability in Everyday Physical Activities*				F	RV Within Domain (95% CI)	RV Across Domains (95% CI)
		Lot More ($n = 431$)	More ($n = 204$)	Same ($n = 64$)	Less ($n = 52$)			
KOOS-12 Pain	4	41.6 (18.7)	32.2 (19.9)	23.7 (18.4)	17.0 (19.0)	41.40	0.87 (0.72, 1.01)	0.52 (0.41, 0.62)
KOOS Pain	9	40.1 (17.7)	30.4 (19.0)	22.5 (18.4)	14.3 (18.3)	47.82	1.00	0.60 (0.49, 0.71)
KOOS-12 Function	4	41.2 (17.5)	28.7 (18.8)	23.1 (19.2)	13.8 (19.2)	57.96	1.20 (0.99, 1.48)	0.72 (0.56,0.87)
KOOS ADL	17	35.0 (16.4)	25.6 (15.4)	18.5 (14.9)	13.4 (17.2)	48.19	1.00	0.60 (0.46, 0.74)
KOOS Sport/Recreation	5	39.3 (23.5)	23.4 (22.0)	18.4 (23.0)	8.0 (18.3)	50.93	1.06 (0.70, 1.55)	0.63 (0.44, 0.87)
KOOS-PS	7	24.4 (14.0)	15.8 (11.8)	11.6 (12.2)	8.2 (12.7)	45.54	0.94 (0.75, 1.20)	0.57 (0.43, 0.72)
KOOS-12/KOOS QOL	4	47.6 (20.4)	31.7 (21.4)	23.9 (22.6)	12.8 (25.2)	69.11	–	0.86 (0.72, 1.04)
KOOS-12 Summary	12	43.5 (15.9)	30.8 (16.4)	23.6 (16.6)	14.5 (17.7)	80.23	–	1.00
KOOS, JR	7	31.7 (14.8)	21.4 (14.2)	16.9 (13.1)	10.8 (15.2)	56.27	–	0.70 (0.58, 0.83)

k, Number of items; F, ANOVA F-statistic; RV, relative validity; CI, confidence interval; ADL, Activities of Daily Living; QOL, Quality of Life. All measures scored so 0 = worst possible and 100 = best possible score. All F-statistics $P < 0.001$.

* Item text (response options): Thinking about your everyday physical activities today (such as walking, climbing stairs, carrying groceries, or participating in sports); Compared to before your joint surgery, are you more or less capable now in your everyday physical activities because of your joint surgery? (A lot more capable now, somewhat more capable now, about the same, somewhat less capable now, a lot less capable now; fourth and fifth response groups combined in ANOVA).

Table V
Effect sizes and standardized response means for knee-specific measures

	k	Mean Score (SD)				Effect Size		Standardized Response Mean	
		6-month post-TKR ($n = 763$)		12-month post-TKR ($n = 659$)					
		Pre-TKR Score	Change Score	Pre-TKR Score	Change Score	6 Month	12 Month	6 Month	12 Month
KOOS-12 Pain	4	44.7 (16.0)	35.7 (20.6)	44.4 (16.3)	39.5 (20.4)	2.22	2.44	1.74	1.93
KOOS Pain	9	48.3 (16.8)	34.0 (19.8)	48.2 (16.9)	37.4 (19.6)	2.03	2.22	1.72	1.90
KOOS-12 Function	4	44.3 (17.3)	34.3 (20.1)	44.4 (17.5)	36.7 (20.0)	1.99	2.13	1.70	1.84
KOOS ADL	17	54.4 (17.1)	29.4 (17.6)	54.6 (17.8)	31.6 (17.9)	1.72	1.81	1.67	1.76
KOOS Sport/Recreation	5	17.7 (16.8)	31.0 (25.0)	18.0 (17.8)	34.3 (26.7)	1.85	1.94	1.24	1.28
KOOS-PS	7	52.0 (13.0)	19.9 (14.5)	52.2 (13.1)	21.7 (14.7)	1.53	1.69	1.37	1.48
KOOS-12/KOOS QOL	4	27.3 (16.7)	38.7 (24.0)	27.2 (17.2)	42.8 (24.4)	2.31	2.47	1.62	1.75
KOOS-12 Summary	12	38.8 (14.4)	36.2 (18.7)	38.7 (14.7)	40.0 (18.7)	2.52	2.70	1.94	2.12
KOOS, JR	7	49.7 (12.8)	26.1 (16.1)	49.6 (13.1)	30.0 (16.5)	2.04	2.29	1.63	1.82

ADL, Activities of Daily Living; QOL, Quality of Life. All measures scored so 0 = worst possible and 100 = best possible score.

Discussion

The objective of this paper was to evaluate the measurement properties of the KOOS-12. Selection of KOOS-12 items was informed by use of modern psychometric methods including IRT models and CAT simulations and by feedback from patients, clinicians, and researchers who developed translations of the KOOS and its companion measure, the HOOS.

This study demonstrated that a relatively short 12-item knee-specific survey can be constructed from KOOS items in a manner that: (1) substantially reduces respondent burden (12-item surveys can be completed by most patients in 2 min or less, since most patients can answer at least 6 items per minute as long as items are not too complicated); (2) allows for scoring a profile of separate Pain, Function, and QOL scales with satisfactory reliability, which reach similar conclusions in tests of validity and responsiveness as the full-length KOOS scales; and (3) achieves the advantages of a summary score with satisfactory validity and responsiveness to change after TKR.

The KOOS-12 Summary score provides an aggregate measure of knee impact across the Pain, Function and QOL domains. It was the only measure that had internal consistency reliability at or above 0.90, which is the minimum level often recommended when a measure is used with individual patients. It also had very low ceiling effects (4–7%) after TKR and was best at detecting differences between groups differing in self-reported evaluation of change in capability to do everyday physical activities post-TKR. A summary score that can be disaggregated into component scales provides the best of both worlds. It reduces the need for multiple comparisons and interpreting multiple outcomes while enabling the interpretation of specific outcomes as needed, for example in patient-clinician communication or systematic reviews and meta-analyses. However, because there are many different ways to arrive at any particular KOOS-12 Summary score, it should be interpreted in relation to the scale scores.

KOOS, JR also provides an overall summary measure of knee health. However, KOOS, JR could not be scored for 8–9% of patients because all seven KOOS, JR items need to be answered in order for a score to be calculated, vs 3–4% of patients for whom a KOOS-12 Summary score could not be calculated. In addition, KOOS, JR had higher ceiling effects at 6 and 12 months post-TKR (9–17% for KOOS, JR vs 4–7% for KOOS-12 Summary) and did not discriminate as well as the KOOS-12 Summary score between groups differing in post-TKR change in capability to do everyday physical activities. KOOS-12 Pain items measured a wider threshold range (−2.30 to 1.30 SD) than the KOOS, JR Pain items (−1.83 to 1.30 SD) in IRT analyses, as did KOOS-12 Function items (−2.02 to 1.45 for KOOS-12 vs −1.56 to 1.05 for KOOS, JR).[10] The KOOS-12 also includes the KOOS QOL scale, which has been shown to be highly responsive to TKR in past studies[8,36] and was highly responsive in this study. Finally, KOOS, JR does not allow for computation of domain scores, which limits its clinical usefulness.

The 4-item KOOS-12 Pain and 9-item KOOS Pain scales correlated highly ($r \geq 0.93$) at all time points. Thus, it was not surprising

that performance of the short form and full-length Pain scales was generally comparable. KOOS-12 Pain had somewhat higher ceiling effects post-TKR than KOOS Pain, but also had slightly higher ES. Both scales had similar performance in discriminating between groups differing in self-evaluated change in post-TKR capability to do physical activities. The KOOS-12 Pain scale is a parsimonious measure that includes the least painful (sitting/lying) and most painful (stairs) KOOS items plus an item selected frequently in CAT simulations (walking on flat) to predict the overall pain score, while also giving more proportional weight to pain frequency than the full-length KOOS Pain scale. The KOOS-12 Pain scale also excludes items that were of limited usefulness in estimating a pain score in CAT simulations. Thus, the KOOS-12 Pain scale appears to be an efficient measure of knee pain.

The KOOS-12 Function scale was reliable, valid and responsive, and performed as well as KOOS ADL and KOOS Sport/Recreation in terms of responsiveness (ES, SRM) and in discriminating between groups differing in self-evaluated change in post-TKR capability to do everyday physical activities. However, the 4-item KOOS-12 Function scale had a higher ceiling effect post-TKR (14–23% compared to 9–16% for the 17-item KOOS ADL and 3–4% for the 5-item KOOS Sport/Recreation scales). The KOOS-12 Function scale did not include some of the most difficult Sport/Recreation items such as running and jumping. These activities are not performed by some knee OA patients, particularly older patients, and were missing for 6–8% of the sample post-TKR. However, these activities are important for younger, more active patients. Thus, it is recommended that the four additional items from the KOOS Sport/Recreation scale be administered in addition to the KOOS-12 for patients who aspire to high-level function, such as those undergoing ACL evaluation.[36] Administering a total of 16 items to younger and more active patients allows for calculation of KOOS-12 scale scores and KOOS-12 summary measure along with the KOOS Sport/Recreation scale. Calculating both the KOOS-12 Function and KOOS Sport/Recreation scales in these patients is recommended, particularly because it allows for long-term follow-up of patients who may decline in knee function over time. The ES of KOOS-PS was not as high as that of KOOS-12 Function or the KOOS ADL and Sport/Recreation scales, and KOOS-PS scores could not be constructed for 7–12% of patients.

As with any new measure, additional research is needed into the performance of KOOS-12. KOOS-12 should be studied in other TKR samples as well as in populations other than knee OA patients who had a TKR, such as early-stage OA patients and patients with ligamentous injuries, and in countries outside the US. This research may be expedited because the KOOS-12 Pain, Function, QOL and summary scores can be calculated from the full-length KOOS. Therefore, interested researchers can glean insight into the performance of KOOS-12 using their existing KOOS data to retroactively monitor results for the KOOS and KOOS-12 in parallel. Documentation as to how to do this will be made available on the www.koos.nu website. Furthermore, analyses reported in this paper used KOOS-12, KOOS-PS and KOOS, JR items that were embedded within the full-length KOOS. Additional studies should be conducted to confirm that KOOS-12 psychometric properties are similar when KOOS-12 is administered by itself. The performance of KOOS-12 also should be evaluated in relation to other knee-specific measures, such as the Oxford Knee Score[37] and Knee Society Score.[38,39] This study also could not evaluate test-retest reliability of the KOOS-12, although intraclass correlation coefficients for full-length KOOS scales ranged from 0.85 to 0.90 in a pooled test-retest analysis of 26 unique cohorts[36] and it is likely that KOOS-12 also would have satisfactory test-retest reliability.

KOOS-12, KOOS, and KOOS-PS are freely available through the www.koos.nu website along with user's guides and scoring instructions. Additional information will be made available on this site, such as a guide to help users determine the KOOS form best suited for their needs and information on comparing scores for KOOS and KOOS-12. The full-length KOOS may offer some advantages in TKR populations, including when used in research, in prediction analyses with models including clinical symptoms, and when TKR is performed in patients with high physical activity levels that are better captured by the KOOS Sport/Recreation scale. However, KOOS-12 is a promising alternative to the full-length KOOS and KOOS derivatives and uniquely, it allows for estimation of domain-specific scores of pain, function and QOL plus an overall knee impact summary score, with reduced respondent burden compared to the full-length KOOS. While the KOOS-12 domain-specific scales are important for clinical interpretation and systematic reviews of OA treatment, the KOOS-12 Summary score demonstrated potential to serve as an aggregate outcome measure for use in clinical trials, registries and quality initiatives.

Distribution

KOOS-12 is available free of charge from www.koos.nu. This site also includes a guide to the different KOOS versions and information on comparing KOOS and KOOS-12 scores. No licensing or permission to use KOOS-12 or other KOOS questionnaires available from www.koos.nu (original KOOS, KOOS-PS) is required.

Authors' contributions

All authors contributed to study conception and design, analysis and interpretation of the data, and drafting the article or revising it critically for important intellectual content. Dr. Gandek assembled the data and performed the data analysis. All authors read and approved the final manuscript. Dr. Gandek takes responsibility for the integrity of the work as a whole.

Conflict of interests

Professor Roos is developer of the KOOS, which is freely available with no licensing required for academic or commercial use. Other authors report no conflicts of interest.

Role of the funding source

This research was supported by AHRQ grant R03 HS024632 (Gandek PI) and a FORCE-TJR program project award (P50 HS018910, Franklin PI) to the Department of Orthopedics and Physical Rehabilitation at the University of Massachusetts Medical School. The funding sources did not play any role in the study design, collection, analysis or interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication. The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or the U.S. Department of Health and Human Services.

Acknowledgements

The authors thank Jakob Bjorner, MD, PhD for psychometric consultation; Nina Deng, EdD for developing the relative validity bootstrapping software; Celeste Lemay RN, MPH, Wenyun Yang, MS and Hua Zheng, PhD for data and computer support; and the researchers, clinicians and patients who provided feedback on KOOS and HOOS item content and translations.

References

1. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function. Arthritis Care Res 2011;63(Suppl 11): S208–28.

2. Rolfson O, Eresian Chenok K, Bohm E, Lubbeke A, Denissen G, Dunn J, *et al*. Patient-reported outcome measures in arthroplasty registries. Acta Orthop 2016;87(Suppl 1):3–8.
3. Engebretsen L, Forssblad M, Lind M. Why registries analysing cruciate ligament surgery are important. BrJSports Med 2015;49:636–8.
4. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS)–development of a self-administered outcome measure. JOrthopSports PhysTher 1998;28:88–96.
5. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. HealthQualLifeOutcomes 2003;1:64.
6. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, *et al*. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:542–50.
7. Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. ClinOrthopRelatRes 2016;474:1461–71.
8. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003;1:17.
9. Ingelsrud LH, Terwee CB, Terluin B, Granan LP, Engebretsen L, Mills KAG, *et al*. Meaningful change scores in the Knee Injury and Osteoarthritis Outcome Score in patients undergoing anterior cruciate ligament reconstruction. Am J Sports Med 2018;46:1120–8.
10. Gandek B, Roos EM, Franklin PD, Ware Jr JE. Item selection for 12-item short forms of the Knee injury and Osteoarthritis Outcome Score (KOOS-12) and Hip disability and Osteoarthritis Outcome Score (HOOS-12) 2019;27:746–753. https://doi.org/10.1016/j.joca.2018.11.011.
11. Gandek B, Roos EM, Franklin PD, Ware Jr JE. A 12-item short form of the Hip disability and Osteoarthritis Outcome Score (HOOS-12): tests of reliability, validity and responsiveness 2019;27:754–761. https://doi.org/10.1016/j.joca.2018.09.017.
12. Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: a patient-centered research consortium for comparative effectiveness in total joint replacement. J Am Med Assoc 2012;308:1217–8.
13. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. JRheumatol 1997;24:799–802.
14. Pham T, Van Der Heijde D, Lassere M, Altman RD, Anderson JJ, Bellamy N, *et al*. Outcome variables for osteoarthritis clinical trials: the OMERACT-OARSI set of responder criteria. JRheumatol 2003;30:1648–54.
15. Singh JA, Dowsey MM, Dohm M, Goodman SM, Leong AL, Scholte Voshaar MMJH, *et al*. Achieving consensus on total joint replacement trial outcome reporting using the OMERACT filter: endorsement of the final core domain set for total hip and total knee replacement trials for endstage arthritis. JRheumatol 2017;44:1723–6.
16. Rolfson O, Wissig S, van Maasakkers L, Stowell C, Ackerman I, Ayers D, *et al*. Defining an international standard set of outcome measures for patients with hip or knee osteoarthritis: consensus of the International Consortium for Health Outcomes Measurement Hip and Knee Osteoarthritis Working Group. Arthritis Care Res (Hoboken) 2016;68:1631–9.
17. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;140:5–55.
18. KOOS scoring 2012, www.koos.nu August 2012.
19. Federal Register. Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services. Fed Regist 2015;80:73273–554. Federal Register 73273.
20. KOOS-PS User's Guide. Updated April 2016. www.koos.nu.
21. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
22. Ware Jr JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care 1995;33:AS264–79.
23. Ware Jr JE. SF-36 health survey update. Spine 2000;25:3130–9.
24. Cronbach LJ. In: Coefficient alpha and the internal structure of tests 1951;16:297–334.
25. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. QualLife Res 2002;11:193–205.
26. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, *et al*. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22–31.
27. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.
28. McHorney CA, Ware Jr JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247–63.
29. Deng N, Allison JJ, Fang HJ, Ash AS, Ware Jr JE. Using the bootstrap to establish statistical significance for relative validity comparisons among patient-reported outcome measures. Health Qual Life Outcomes 2013;11:89.
30. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.
31. Ware Jr JE, Keller SD. Interpreting general health measures. In: Spilker B, Ed. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd edn. Philadelphia, PA: Lippincott-Raven Publishers; 1996:445–60.
32. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo ClinProc 2002;77:371–83.
33. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178–89.
34. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632–42.
35. Nunnally JC, Bernstein IH. Psychometric Theory. New York: Mc-Graw Hill; 1994.

36. Collins NJ, Prinsen CA, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthritis Cartilage 2016;24:1317–29.
37. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J.Bone Joint Surg.Br. 1998;80:63–9.
38. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. Clin Orthop Relat Res 1989;248:13–4.
39. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, *et al*. Development of a new Knee Society Scoring System. Clin Orthop Relat Res 2012;470:20–32.