The American Journal of Sports Medicine

Validation of the Knee Injury and Osteoarthritis Outcome Score Subscales for Patients With Articular Cartilage Lesions of the Knee
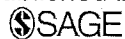
Luella Engelhart,* PhD, Lauren Nelson,†‡ PhD, Sandy Lewis,‡ BSN, Margaret Mordin,‡ MS, Carla Demuro-Mercon,‡ MS, PhD Candidate, Sharif Uddin,§ MS, Lori McLeod,‡ PhD, Brian Cole,‖ MD, and Jack Farr,¶ MD
Investigation performed at RTI Health Solutions, Research Triangle Park, North Carolina

Background: The Knee Injury and Osteoarthritis Outcome Score (KOOS) assesses acute and chronic knee injuries or early-onset osteoarthritis in young, active patients. The United States Food and Drug Administration guidelines recommend that patient-reported outcome instruments used to support clinical trial label claims should demonstrate content validity using patient input and have acceptable psychometric properties in the target population. To use the KOOS subscales in safety and efficacy trials assessing new treatments for patients with articular cartilage lesions, additional validation work, using input from patients with articular cartilage lesions, was necessary.

Purpose: Qualitative and quantitative evaluations of the KOOS subscales' validity among patients with articular cartilage lesions were conducted to support their use as clinically meaningful end points in clinical trials.

Study Design: Cohort study (diagnosis); Level of evidence, 2.

Methods: For qualitative analysis, cognitive interviews involving concept elicitation and cognitive debriefing with the KOOS items were conducted with 15 participants aged 25 to 52 years. Participants either were candidates for cartilage repair or had undergone cartilage repair 6 months or more before the study. For the quantitative analysis, a psychometric evaluation of the KOOS was conducted with clinical trial data from 54 patients, aged 18 to 55 years, evaluating the Cartilage Autograft Implantation System in the United States (n = 29) and the European Union (n = 25). Data were collected before surgery and at 7 postsurgical visits up to 12 months. Internal consistency and test-retest reliability, construct validity, responsiveness, and estimates of the minimal detectable change (MDC) were assessed. Test-retest reliability was assessed using data from months 2 and 3 on a subset of stable patients.

Results: Qualitative research confirmed that concepts measured on the KOOS are important to patients with articular cartilage lesions. Most participants reported the KOOS was comprehensive and appropriate. In the quantitative research, KOOS subscales showed excellent internal consistency reliability (range, .74-.97 at baseline) and test-retest reliability (range, .78-.82). Construct validity results supported hypothesized relationships, with significant correlations ($r \geq .50$) in the expected directions. Responsiveness analyses demonstrated excellent sensitivity to change; standardized response means ranged from 0.8 to 1.2, and MDC estimates ranged from 7.4 to 12.1.

Conclusion: The study results support the use of the KOOS subscales among patients with articular cartilage lesions.

Keywords: articular cartilage; KOOS; reliability; validity; ability to detect change

Patient-reported outcomes (PROs) provide valuable information from patients' perspectives and can often complement clinical outcomes in assessing the efficacy and safety of new treatments. For products intended to repair knee cartilage, the United States (US) Food and Drug Administration (FDA) guidance states that "clinically meaningful endpoints, such as improvement in pain and physical function, provide the most persuasive evidence of efficacy."[23] Furthermore, the guidance provides specific examples of PRO measures that may be appropriate to assess pain and function end points, including the Knee Injury and Osteoarthritis Outcome Score (KOOS).

The KOOS was designed to assess acute and chronic knee injuries or early-onset osteoarthritis in young, active patients.[17] The 42 items of the KOOS cover 5 patient-relevant dimensions: Symptoms (7 items), Pain (9 items), Activities of Daily Living (ADL) (17 items), Sports/Recreation (5 items), and knee-related Quality of Life (QOL) (4 items). Each subscale is scored from 0 to 100 on a worst-

to-best scale. In an effort to detect changes in a more active population, Roos[16,17] added questions to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), which was designed for evaluating the longer term consequences of knee injuries such as osteoarthritis.[3] The 24 items of the WOMAC Osteoarthritis Index were included in the KOOS in their original form: the WOMAC Function subscale is identical to the KOOS ADL subscale, the WOMAC Pain items are included in the KOOS Pain subscale, and the WOMAC Stiffness items are included in the KOOS Symptoms subscale. Items for the KOOS Sports/Recreation and QOL subscales are unique to the KOOS. Previous studies have presented the psychometric properties of the KOOS, such as the reliability, construct validity, and responsiveness, for use in clinical studies of patients requiring several orthopaedic interventions including articular cartilage lesion reconstruction,[17] meniscectomy,[16] and total knee and focal cartilage defect surgery.[2] More recently, the FDA released the final PRO guidance,[22] which outlines how the agency will review PROs used to support label or promotional claims. Specifically, the FDA's PRO guidance[22] requires demonstration of content validity and construct validity, in addition to psychometric properties, such as reliability and responsiveness to change, within the specific population of interest to the clinical study.

To date, the KOOS has undergone limited validation (content and psychometric) among patients with articular cartilage lesions.[2] The current study conducted both qualitative and quantitative research to evaluate the validity of the KOOS among patients with articular cartilage lesions. The research was performed in accordance with the FDA's PRO guidance[22] to facilitate future use of the KOOS subscales in clinical trials. The qualitative work evaluated the KOOS' ability to measure symptoms, function, and other aspects considered important to patients with articular cartilage lesions (content validity), while the quantitative work investigated whether the concepts represented on the KOOS conform to a priori hypotheses concerning logical relationships, in direction and magnitude, that should exist with external measures of related concepts in patients with articular cartilage lesions (construct validity).[22]

MATERIALS AND METHODS

Qualitative evaluations consisted of in-depth interviews and cognitive debriefing of the KOOS with patients who had undergone or were candidates for articular cartilage repair (ACR). Additionally, the evaluation included qualitative assessment of a numerical rating scale (NRS) for measuring knee pain intensity that requires patients to rate their pain on an 11-point scale, with 10 being the most severe, to provide support for content validity of these measures within this population. Quantitative evaluation included psychometric analyses of data obtained during 2 multicenter, randomized, pilot clinical trials. Patients in the qualitative evaluation were not clinical trial participants and thus consisted of an entirely separate sample from those included in the quantitative assessment. Each evaluation is further described below.

Qualitative Methodology

Using a standardized screening form, 3 US investigative sites recruited a total of 15 participants who were either a candidate for a cartilage procedure or a patient who had undergone a cartilage procedure more than 6 months ago, which is considered to be outside of the window in which they would still be experiencing postoperative pain. Inclusion of both presurgical (n = 3, 20%) and postsurgical (n = 12, 80%) patients allowed for feedback across the spectrum of injury and recovery to understand the relevance of items to patients over time and after intervention. Participants were eligible if they met the following criteria:

- were between the ages of 18 and 65 years;
- were either of the following: a candidate for cartilage repair (with knee pain symptoms for >6 months), or at least 6 months after a cartilage procedure (procedures included chondroplasty debridement, marrow stimulation, and others);
- had experienced at least mild pain over the past 7 days (NRS ≥ 1);
- were able to read and speak English fluently; and
- were willing to discuss their knee problems in an interview format.

Participants were ineligible for the following reasons:

- had generalized osteoarthritis, rheumatoid arthritis, inflammatory systemic arthritis, or disease involving the index knee;
- had symptoms greater than NRS level 2 for the nonindex knee;
- had a body mass index (BMI) greater than 40;
- had prior anterior cruciate ligament (ACL) or posterior cruciate ligament (PCL) reconstruction within the past 6 months;
- had prior tendon and/or ligament (non-ACL or -PCL) repair or patellar surgery within the past 6 months;

TABLE 1
Demographic Characteristics of Patients Participating
in the Qualitative Research[a]

Characteristic	Value
Age range, y	25-52
Male sex, n (%)	10 (67)
Education, n (%)	
High school diploma or equivalent	3 (20)
Some college	4 (27)
College degree or advanced degree	8 (53)
Race/ethnicity, n (%)	
Black/other	2 (14)
Hispanic	5 (33)
White	8 (53)
NRS (index knee), n (%)[b]	
1	4 (27)
2	3 (20)
3	3 (20)
4	0 (0)
5	0 (0)
6	4 (27)
7	1 (7)
NRS (nonindex knee), n (%)	
0	7 (47)
1	6 (40)
2	2 (13)
Surgical status, n (%)	
Presurgical	3 (20)
Postsurgical	12 (80)
Body mass index, median (range)	27 (20-38)

[a]NRS, numerical rating scale for pain.
[b]Percentages do not add up to 100 because of rounding.

- had prior total or subtotal meniscectomy or magnetic resonance imaging evidence of meniscal abnormalities;
- had prior partial meniscectomy to the index knee within the past 6 months;
- had a presence of bipolar or "kissing lesions" on the index knee;
- had prior exposure to the KOOS measure within the past year; and
- had a history of drug/alcohol abuse, recent steroid therapy, pain medications, or mental incompetence.

Participants in the in-depth interview ranged in age from 25 to 52 years, were predominantly white and male, and had educational backgrounds ranging from a high school diploma to a graduate-level degree. Demographic information, including surgical status, NRS scores for pain in the index and nonindex knee, and BMI are summarized in Table 1.

Concept Elicitation. To confirm that the KOOS has relevant constructs for measuring changes associated with repair of articular cartilage lesions, cognitive interviews of participants were performed using a standardized interview guide. A qualitative assessment of the NRS also was conducted. Each interview began with participants responding to a series of open-ended questions designed to get them talking about their experiences with their knee problems. Participants were asked to describe the symptoms associated with their knee and discuss ways in which their life may have been affected, such as their ability to perform daily activities or participate in leisure activities.

KOOS Cognitive Testing. The interview then focused on a cognitive debriefing review of the KOOS and the NRS. A "think-aloud" technique was utilized, wherein the interviewer asked the participant to read the instructions aloud and describe, in his or her own words, how he or she interpreted them. Participants also were asked about the relevance and importance of each item in relation to their knee problems.

Quantitative Methodology

Study Design. Data used in the quantitative analyses were collected in 2 multicenter, randomized, pilot clinical trials conducted by Advanced Technologies and Regenerative Medicine LLC (ATRM, Raynham, Massachusetts) in the US and European Union (EU) to evaluate the safety and performance of the Cartilage Autograft Implantation System (CAIS) for primary surgical treatment of articular cartilage lesions of the knee (ranging from ≥ 1 cm^2 to ≤ 10 cm^2). A total of 54 patients (US sample: n = 29; EU sample: n = 25) with articular damage of the femoral condyle who were candidates for primary surgical treatment were randomized in a 2:1 scheme to receive the CAIS (treatment) or a microfracture procedure (control). Full details of these studies are provided in Cole and colleagues,[5] and key inclusion and exclusion criteria are outlined in Appendix 1 (available online at http://ajs.sagepub.com/supplemental/).

Candidates were excluded if they met any of 22 distinct criteria including having a clinical and/or radiographic disease diagnosis of the index joint, including osteoarthritis, rheumatoid arthritis, and avascular necrosis. The majority of participants were male (n = 32, 59.3%) and white (n = 50, 92.6%), with a BMI of 25.9, and the average age was 34.1 years (range, 18-54 years). Data were collected using standardized case report forms. All participants were assessed before surgery and at 7 follow-up visits (at 1 and 3 weeks and at 2, 3, 6, 9, and 12 months after baseline) for US sites and 6 follow-up visits (at 1 and 3 weeks and at 2, 3, 6, and 12 months; there was no 9-month assessment) at EU sites.

Measures. Two knee-specific outcome instruments, (1) the International Knee Documentation Committee Subjective Knee Form (IKDC) and (2) the Tegner activity level scale (Tegner), and 2 generic health-related quality of life (HRQOL) measures, (1) the 36-item Short Form Health Survey (SF-36) and (2) the EQ-5D, were used in establishing construct validity.

International Knee Documentation Committee Subjective Knee Form. The IKDC has 17 items related to knee symptoms, knee function, and sports activity and is used to evaluate patients who have a variety of knee conditions.[1,9] The IKDC has been shown to have acceptable psychometric properties (internal consistency and test-retest reliability) as well as construct validity in several populations,[6] including patients with articular cartilage injuries.[8]

Tegner. The Tegner was developed to assess the current activity level of patients with knee disorders and instructs patients to select the single best-fitting description of their

activity level (prior to injury and currently), where 0 is defined as "on sick leave/disability" and 10 is "participation in competitive sports such as soccer at a national or international elite level."[20]

SF-36. The 8 SF-36 subscales (Physical Functioning, Role Limitations–Physical Health, Bodily Pain, Social Functioning, General Mental Health, Role Limitations–Emotional, Vitality, and General Health) and 2 SF-36 component scores (Physical Component and Mental Component) are widely used to assess general health status.[10,24,25]

EQ-5D. The EQ-5D (formerly known as the EuroQoL)[21] was used to assess general health status. Patients are asked to reflect on their health state "today."

Additionally, a visual analog scale for pain (VAS Pain) and an 11-point Patient-Reported (PR) Functional Status item (0 = "cannot perform activities" to 10 = "no limitation") were used to assess the KOOS responsiveness. Along with the KOOS subscale measurements, longitudinal assessments were obtained at baseline, weeks 1 and 3, and months 2, 3, 6, 9, and 12 for the Tegner, EQ-5D, VAS Pain, and PR Functional Status. The IKDC and SF-36 were assessed at baseline and months 2, 6, and 12. The NRS was not included in the protocols for the pilot studies.

Reliability

Internal consistency reliability estimates of the 5 KOOS subscales were evaluated by computing the Cronbach coefficient α using item-level data at baseline. The Cronbach α ranges between 0 and 1 and corresponds to the average interitem correlation. Moderately high interitem correlations (ie, .70-.90) indicate ideal internal consistency, and the items can be grouped together. When interitem correlations are low (ie, <.70), then the items may be measuring different constructs, while high interitem correlations (ie, >.95) suggest that the items may contain redundancy and allow for item reduction.

Test-Retest Reliability

The original pilot studies did not include baseline test-retest components. However, assessments from months 2 and 3 were used in the current analysis to examine the test-retest reliability of the KOOS subscales because this time interval in postoperative patients was expected to be relatively stable, with some patients showing small improvements. Intraclass correlation coefficients (ICCs) were computed for each of the 5 KOOS subscales using data from month 2 as the "test" administration and data from month 3 as the "retest" administration for patients who did not change on the PR Functional Status item. Following the recommendation of Schuck,[19] a 2-way (patient × time) random-effects analysis of variance (ANOVA) was used to compute the ICC estimates of test-retest reliability for the KOOS subscales. It is generally recommended that ICCs be at least .70 for multiple-item scales (eg, Nunnally and Bernstein[13]).

Construct Validity

To establish construct validity, patients were administered additional measures designed to assess the same or different constructs. Correlation analyses based on scores from the 2 knee-specific measures (IKDC and Tegner), 2 generic HRQOL measures (SF-36 and EQ-5D), and the KOOS allowed for the evaluation of the predicted relationships. The majority of hypothesized directions and magnitudes of correlation coefficients (r) anticipated between the KOOS subscales were based on previous studies.[2,15-18] The strength of the correlations was assessed using the Cohen criteria,[4] where a correlation <.30 is considered weak, a correlation between .30 and .50 is considered moderate, and a correlation >.50 is considered strong. A conservative type I error rate of 1% was applied to each individual hypothesis test.

The following study measures of somewhat similar concepts were used to demonstrate convergent validity ($|r| > .50$):

- the KOOS Symptoms score with the SF-36 Physical Functioning subscale score;
- the KOOS Pain score with the IKDC Pain score and the SF-36 Bodily Pain subscale score;
- the KOOS ADL score with the SF-36 Physical Functioning subscale score;
- the KOOS Sports/Recreation score with the IKDC Activity score, the Tegner score, and the SF-36 Physical Functioning subscale score; and
- the KOOS QOL score with the SF-36 General Health subscale score and the EQ-5D index score.

Dissimilar concepts were used to demonstrate divergent validity ($|r| < .30$). It was anticipated that the KOOS Symptoms, Pain, ADL, and Sports/Recreation scores would not correlate as highly with the SF-36 mental health subscales (Vitality, Role Emotional, Social Functioning, and Mental Health) as they correlate with SF-36 physical subscales.

Known-Group Analyses

A known-group analysis was conducted by classifying patients based on the PR Functional Status levels 0 to 3 (low) versus levels 4 to 7 (high). The hypotheses tested at 6 and 12 months were that participants with higher levels on the PR Functional Status would have significantly better KOOS Pain, Symptoms, ADL, Sports/Recreation, and QOL scores. Using ANOVA, the means of each of the KOOS subscales in the low and high functioning groups were compared, and a type I error rate of 1% was applied.

Responsiveness

Responsiveness was investigated using a variant of the Guyatt responsiveness statistic,[7] which compares differences between predefined groups that are anticipated to show change. Patients who improved from baseline to month 12 by 3 points on the PR Functional Status or improved on the VAS Pain by at least 30% from baseline to month 12 were

classified as changed. Patients who reported less than a 3-point improvement or worsened on the PR Functional Status or reported changes of less than 30% on the VAS Pain were classified as not changed. The Guyatt responsiveness statistic is a standardized change score, similar to an effect size, and was calculated as the difference in the change between these 2 groups (ie, changed and not changed) divided by the standard deviation (SD) of the group who did not change, essentially yielding an effect size in units based on the SD of the group with no change (SD_{NC}). Cohen[4] provides a general guideline for the interpretation of such effect size estimates: effect sizes of about .20 represent small effects, those of about .50 represent moderate effects, and those greater than about .80 represent large effects.

Guidance for Interpretation of Change

Several methods have been proposed in the psychometric literature to assess meaningful change,[11,14] including methods that estimate change that is minimally detectable above the measurement error (ie, distribution-based methods) or change that is meaningful to patients or clinicians (ie, anchor-based methods). Based on the data collected in this study, 2 distribution-based methods recommended in the FDA's PRO guidance were employed: (1) the half SD as recommended by Norman and colleagues,[12] and (2) the standard error of measurement (SEM) as recommended by Wyrwich and colleagues[26] and computed as SEM = SD * sqrt(1 − r), where SD is the standard deviation of the subscale score and r is the test-retest reliability estimate.

RESULTS

Qualitative Results From In-depth Patient Interviews and Cognitive Debriefings

Concept Elicitation. Participants spontaneously reported a variety of knee symptoms. These symptoms predominantly included pain, pressure, soreness, cramping in the leg or calf, locking of the knee, weakness or instability of the knee, swelling, popping, grinding, clicking, cracking, and stiffness, which largely corresponded with the first 3 sections of the KOOS (Symptoms, Stiffness, and Pain). While some symptoms abated or at least improved in frequency/severity after surgery for many patients, some still noted lingering sensations as described previously.

Participants were asked to describe ways in which their lives had been affected by their knee problems. All but 1 of the 15 participants shared that they have lingering functional limitations because of their knee. A major focus for the majority of participants was the inability to participate in activities that they previously enjoyed. Others described activities that they simply no longer attempted to avoid pain or because of fear of reinjury. Participants discussed both the physical limitations that they have regarding certain activities and the mental limitations or fears that they have regarding reinjury of their knee.

Participants also discussed the impact of their knee problems on their day-to-day activities, and responses included things such as carrying and running after children, squatting or kneeling to get up and down from the floor, cleaning the house, carrying laundry (or other heavy items), and using stairs or walking on sloping surfaces. Four participants shared that their knee injury had affected their position at work, either causing them to perform lighter duties or switch careers because they could no longer do such things as climb ladders, kneel, or maneuver around to perform their job. These functional impacts were represented in the daily living items presented in the KOOS.

Almost half of the participants described frustration about muscular atrophy that they experienced in the area surrounding their injured knee, noting that their knee "doesn't feel as strong as [it] used to be." Other effects also noted by the majority of participants included difficulty when sleeping, stiffness from maintaining certain positions, and inability to fully extend or bend their knee.

Cognitive Interview. After the open-ended portion of the interview, participants were asked to review the KOOS and the NRS. With few exceptions (such as US-centric suggestions to change terms such as the British term "ticking" in the instructions, recommendations to add transitional statements to highlight the shift between concepts of pain and function, and questions about the relevance of the notion of getting in/out of a bath), all participants reported that the KOOS and NRS items were clear and easy to understand. Participants indicated that the instructions were straightforward and that the response options provided were reasonable and appropriate. The majority of participants noted that they would be able to answer questions in either measure based on the current recall period of 1 week for the KOOS or the past 24 hours for the NRS. At the conclusion of the interview, a few of the participants indicated that the emotional component (how you felt about the injury/recovery, depressive feelings, how you are feeling mentally) was missing, although 1 section does have a question related to this area.

Quantitative Results From Analysis of Pilot Trial Data

Descriptive Statistics. Baseline descriptive statistics for the KOOS subscales, IKDC total, IKDC Pain, IKDC Activity, Tegner, 8 SF-36 subscales and 2 SF-36 component scores, EQ-5D, VAS Pain, and PR Functional Status are shown in Table 2. The distribution of each KOOS subscale score at baseline and months 6 and 12 is displayed in the box plots of Figure 1. As seen in Figure 1, the mean of each of the KOOS subscales was very low at baseline and showed great improvement across the 12-month trial.

Reliability. Estimates of the internal consistency reliability for the KOOS baseline subscale scores were .97 for ADL, .92 for Pain, .91 for Sports/Recreation, .84 for Symptoms, and .74 for QOL. These estimates indicate that the items composing each subscale can be grouped, although the elevated value for the ADL subscale provides some evidence that this scale may contain item redundancy. The ICC test-retest reliability estimates for the

TABLE 2
Descriptive Statistics for Measures at Baseline[a]

Measure	n	Mean	SD	Median	Minimum	Maximum
KOOS						
Symptoms	54	57.8	23.6	58.9	7.1	96.4
Pain	54	53.4	22.4	52.8	0.0	94.4
ADL	54	63.2	22.2	65.4	0.0	100.0
Sports/Recreation	54	28.3	24.3	25.0	0.0	90.0
QOL	54	26.4	17.4	25.0	0.0	68.8
IKDC						
Total	54	38.4	16.8	36.8	5.8	73.6
Pain	54	4.39	1.8	4.00	1.00	9.00
Activity	54	2.19	0.8	2.00	1.00	5.00
Tegner	53	2.47	1.7	2.00	0.00	7.00
SF-36						
Physical Functioning	54	35.99	10.0	36.17	15.19	57.14
Role Limitations–Physical Health	54	39.74	11.3	38.57	27.95	56.24
Bodily Pain	54	37.19	8.6	37.48	19.93	55.90
Social Functioning	54	46.68	11.6	51.71	13.71	57.14
General Mental Health	53	50.85	10.3	54.99	11.82	64.07
Role Limitations–Emotional Health	54	49.68	11.0	55.34	23.74	55.34
Vitality	53	52.94	9.9	53.79	25.39	67.99
General Health	54	54.17	8.9	55.57	33.57	64.00
PCS	54	37.67	9.4	35.10	22.73	57.76
MCS	54	55.62	10.9	57.51	17.44	70.93
EQ-5D						
Index	54	0.68	0.2	0.78	0.17	1.00
Scale	54	71.56	18.7	74.50	29.00	100.0
VAS Pain	54	6.77	1.7	6.80	2.90	10.00
Patient-Reported Functional Status	54	3.48	2.2	3.00	0.00	9.00

[a]SD, standard deviation; ADL, Activities of Daily Living; QOL, Quality of Life; IKDC, International Knee Documentation Committee; PCS, Physical Component Score; MCS, Mental Component Score; VAS, visual analog scale.

KOOS using scores from a stable group (n = 33) of patients based on the PR Functional Status score across the early postsurgical months 2 and 3 were .78 for Symptoms, .82 for Pain, .79 for ADL, .80 for Sports/Recreation, and .82 for QOL.

Construct Validity. Appendix 2 (available online) presents the correlation coefficients for the planned convergent and divergent correlation analyses in bold as well as correlation coefficients between the KOOS subscales and measures that were not specifically hypothesized. Observed correlations were generally supportive of the hypothesized relationships between the 5 KOOS subscales and the additional measures. However, the correlation between SF-36 General Health and the KOOS QOL subscale was lower than anticipated and nonsignificant (tested at $P \leq .01$), ranging from .07 at baseline to .39 at month 12.

The planned divergent correlations were larger than anticipated. At baseline, the KOOS subscales and SF-36 Social Functioning subscale scores ranged from $r = .47$ to $r = .64$, and the correlations between SF-36 General Mental Health and the KOOS Pain and ADL subscale scores at baseline were $r = .53$ and $r = .62$, respectively. However, by month 12, the majority of correlations between the KOOS subscales and the SF-36 subscales representing mental health subscale scores (Vitality, Role Limitations–Emotional, and General Mental Health) had decreased (ie, $r < .40$), while the correlations between the SF-36 subscales representing physical outcomes (Physical Functioning, Role Limitations–Physical Functioning, and Bodily Pain) and the KOOS subscales remained strong (ie, $r > .50$), demonstrating divergent validity.

Known-Group Analysis. Results from the known-group analysis were consistent with anticipated findings. Participants who had greater limitations in performing daily activities according to the PR Functional Status (scoring 3 or less) scored significantly worse on all KOOS subscales. At baseline, the majority of participants (n = 33) reported problematic limitations; their KOOS subscale scores were significantly worse than the KOOS subscale scores of the 18 participants who were not as limited. At month 12, the majority of participants reported fewer limitations (n = 24) and had higher KOOS subscale scores than those reporting more limitations (n = 7).

Responsiveness. According to the Cohen rule of thumb for effect sizes, Guyatt statistics were large for the Sports/Recreation (1.06) and QOL (0.91) subscales, indicating that improvement in the changed group on these subscales is on the order of 1 SD_{NC} above the improvement in the not-changed group. The responsiveness statistics
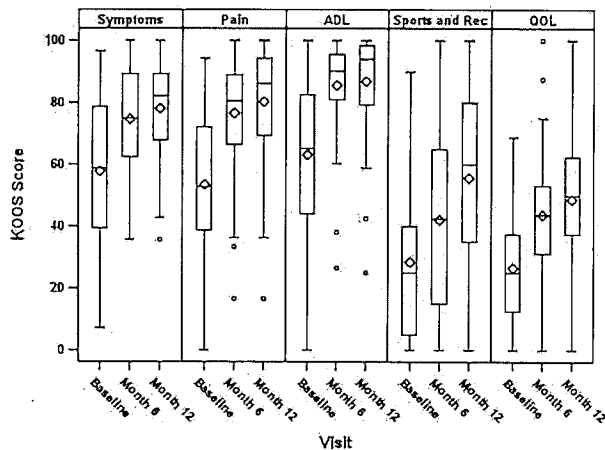
Figure 1. Box plots of Knee Injury and Osteoarthritis Outcome Score (KOOS) subscale scores at baseline and 6 and 12 months. Each box plot represents a distribution of KOOS subscales, with the lower and upper edges of the gray box denoting the 25th and 75th percentiles, respectively, sometimes also referred to as the interquartile range (IQR). The line in the box is the 50th percentile (ie, median) of the score distribution, and the diamond shape in the box is the mean of the distribution. The "whiskers" on each box display the range of score values that are not identified as outliers, while the open circles represent scores that are identified as outliers using a criterion >1.5 IQR. ADL, Activities of Daily Living; QOL, Quality of Life.

were moderate for the Symptoms (0.55), Pain (0.51), and ADL (0.51) subscales, indicating that improvement in the changed group on these subscales is on the order of half SD_{NC} above the improvement in the not-changed group.

Guidance for Interpretation of Change. Table 3 presents the half SD and the SEM estimates for the amount of change beyond the measurement error, which are comparable.

DISCUSSION

The 2011 FDA guidance for products intended to repair knee cartilage recognized that PROs assessing pain and physical function provide the clearest evidence for efficacy.[23] The guidance mentions several potential PRO measures, including the KOOS. However, the FDA's final PRO guidance[22] requires assessment of PRO instruments for content and construct validity within the population of interest. The current study evaluated the KOOS both qualitatively and quantitatively in patients with articular cartilage lesions. Overall, the qualitative research confirmed the concepts measured on the KOOS as relevant to patients with articular cartilage lesions, and all aspects of physical functioning that are important to participants were noted to be addressed in the KOOS.

Previous research by Hambly and Griva[8] points to the IKDC as the knee-specific instrument of choice for ACR

TABLE 3
Minimum Detectable Change Estimates[a]

KOOS Subscale	Half SD	SEM
Symptoms	11.8	11.1
Pain	11.2	9.50
ADL	11.1	10.7
Sports/Recreation	12.1	10.8
QOL	8.7	7.4

[a]KOOS, Knee Injury and Osteoarthritis Outcome Score; SD, standard deviation at baseline; SEM, standard error of measurement; ADL, Activities of Daily Living; QOL, Quality of Life.

patients, demonstrating superiority in terms of inclusion of the most important items to patients for measurement of symptoms and disabilities. However, the authors also acknowledge that the KOOS contains a large number of items that are experienced by and are important to patients after ACR. The methodology used for the qualitative research conducted as part of the current study allowed for individual in-depth analysis of the importance and relevance of each item presented as part of the KOOS, and results underscore the relevance for use of the measure in this patient population. Items were individually debriefed with each participant utilizing a procedure known as "think aloud." Participants were asked to read each item aloud and discuss first their interpretation of the item (ie, what the item means to them in their own words). Participants were also asked if the items presented are relevant and important to someone experiencing ACR. With few exceptions, all of the items in the KOOS were unanimously endorsed as both important and relevant to patients experiencing ACR (Appendix 3, available online). There were 9 items that, while not unanimously, were endorsed by the majority of the sample as relevant and important.

The quantitative evaluation in the current study builds on previous work conducted to validate the KOOS for patients with a variety of knee conditions, including articular cartilage lesions,[17] meniscectomy,[18] and total knee replacement.[16] Recently, Bekkers and colleagues[2] validated the KOOS using a sample of 40 patients who were retrospectively identified as having undergone either articular cartilage implantation or microfracture for a focal cartilage lesion at some point during a 4-year span. Data from the current study provided an opportunity to evaluate the KOOS in a cartilage repair population in the context of a clinical trial designed to assess the safety of the CAIS procedure over microfracture.[5]

The majority of the quantitative evaluation findings support results reported in previous KOOS validation studies.[2,8,18] The difference in KOOS means across subscales at month 12 in this study was consistent with results reported by Bekkers et al[2] after repair of focal cartilage defects. However, the baseline KOOS mean scores were considerably worse (mean scores ranging 20-60) than those reported in the Bekkers et al[2] sample (mean scores ranging 49-77), indicating that patients in the current study at baseline (before surgery) were much worse off than patients in the Bekkers et al[2] study.

The KOOS subscales demonstrated adequate internal consistency reliability, ranging from .74 to .97 in the overall sample at baseline. These estimates are consistent with previously reported estimates ranging from .71 to .95[18] and from .74 to .96.[2] Although the original pilot study designs lacked a standard test-retest component (using 2 occasions during a stable period such as presurgery), test-retest reliability was estimated using a subset of patients in the early postsurgery months who reported little change on functioning. The ICCs ranged from .78 to .82 and are larger than the .70 criterion set in the psychometric literature.

Correlation coefficients between the KOOS subscales and the 2 knee-specific instruments (IKDC and Tegner), and the 2 generic HRQOL instruments (SF-36 subscales and the EQ-5D), followed patterns that were explicitly hypothesized a priori, providing evidence of convergent and divergent validity. The construct validity evidence of the present analysis confirms findings reported in several previous studies. Specifically, strong correlations between KOOS subscales and the SF-36 physical subscales and weaker correlations between KOOS subscales and the SF-36 mental health subscales have been reported in a study evaluating patients with advanced osteoarthritis[15] and, more recently, in a study of patients with focal lesions of the knee.[2] More importantly, the strong correlation between the change from baseline and month 12 in the IKDC and the KOOS subscales suggests that both these measures have potential in assessing treatment benefit in patients in need of ACR. The advantage of using the KOOS in clinical trials would be that the constructs of pain, symptoms, functioning, athletic ability, and QOL can be assessed separately, allowing consideration of treatment benefit for each of them.

In terms of discriminating ability, the KOOS subscale scores differed predictably and significantly across known groups of interest, for example, when participants were classified according to PR Functional Status. All 5 of the KOOS subscales demonstrated satisfactory responsiveness to change, based on Guyatt responsiveness statistics using a definition requiring improvement in both functional status and pain.

As a guide for interpretation of changes on the 2 KOOS subscales recommended by the FDA for use as clinical trial end points, the KOOS Pain and Sports/Recreation subscales, the magnitude of a meaningful change was estimated using distribution methods. According to the FDA's PRO guidance, these 2 methods are acceptable ways to use for defining responder thresholds. The half SD estimates ranged from a low of 8.7 for the QOL subscale to a high of 12.1 for the Sports/Recreation subscale. Using the ICCs estimated in this study yielded SEM estimates smaller than the half SD. These 2 distribution methods for assessing change suggest a minimal change of approximately 10 points as meaningful, which is comparable with the clinically meaningful difference observed in the original WOMAC scale.[15]

There are several limitations to our qualitative and quantitative analyses. Although we tried to recruit participants for the interviews across age, race, sex, educational status, and surgical status, the majority of participants were male and postsurgical patients because of the strict protocol criteria. The pilot clinical trial sample size used in the quantitative evaluation was small. Also, the design did not include a patient-reported global rating item, limiting this study's ability to estimate a clinically meaningful difference using anchor-based methods. Despite these limitations, the qualitative and quantitative evaluations performed in the present study provide evidence to support the appropriateness of the KOOS subscales among patients with articular cartilage lesions.

REFERENCES

1. Anderson AF, Irrgang JJ, Kocher MS, Mann BJ, Harrast JJ. The International Knee Documentation Committee Subjective Knee Evaluation Form: normative data. *Am J Sports Med.* 2006;34(1):128-135.
2. Bekkers JE, de Windt TS, Raijmakers NJ, Dhert WJ, Saris DB. Validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) for the treatment of focal cartilage lesions. *Osteoarthritis Cartilage.* 2009;17(11):1434-1439.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15(12):1833-1840.
4. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
5. Cole BJ, Farr J, Winalski CS, et al. Outcomes after a single-stage procedure for cell-based cartilage repair: a prospective clinical safety trial with 2-year follow-up. *Am J Sports Med.* 2011;39(6):1170-1179.
6. Garratt AM, Brealey S, Gillespie WJ; DAMASK Trial Team. Patient-assessed health instruments for the knee: a structured review. *Rheumatology (Oxford).* 2004;43(11):1414-1423.
7. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40(2):171-178.
8. Hambly K, Griva K. IKDC or KOOS? Which measures symptoms and disabilities most important to postoperative articular cartilage repair patients? *Am J Sports Med.* 2008;36(9):1695-1704.
9. Irrang JJ, Anderson AF, Boland AL, et al. Development and validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2001;29(5):600-613.
10. McHorney CA, Ware JE Jr, 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(3):247-263.
11. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hayes RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2):163-169.
12. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41(5):582-592.

13. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill; 1994.

14. Roos EM, Engelhart L, Ranstam J, et al. ICRS Recommendation Document: patient-reported outcome instruments for use in patients with articular cartilage defects. *Cartilage*. 2011;2(2):122-136.

15. Roos EM, Lohmander LS. The Knee Injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1:64.

16. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee Injury and Osteoarthritis Outcome Score (KOOS): validation of a Swedish version. *Scand J Med Sci Sports*. 1998;8(6):439-448.

17. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998;28(2):88-96.

18. 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.

19. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Qual Life Res*. 2004;13(3):571-586.

20. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res*. 1985;198:43-49.

21. The EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208.

22. US Food and Drug Administration. Guidance for Industry. Patient-reported outcome measures: use in medical product development to support labeling claims. December 2009. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf. Accessed May 14, 2012.

23. US Food and Drug Administration. Guidance for Industry. Preparation of IDEs and INDs for products intended to repair or replace knee cartilage. December 2011. Available at: http://www.fda.gov/downloads/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/cellularandgenetherapy/ucm288011.pdf. Accessed May 14, 2012.

24. Ware JE Jr, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires: methods from the IQOLA Project. International Quality of Life Assessment. *Int J Technol Assess Health Care*. 1995;11(3):525-551.

25. Ware JE Jr, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey Manual and Interpretation Guide*. Boston: The Health Institute; 1993.

26. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol*. 1999;52(9):861-873.

For reprints and permission queries, please visit SAGE's Web site at http://www.sagepub.com/journalsPermissions.nav