

Responsiveness of the International Knee Documentation Committee Subjective Knee Form in Comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, Modified Cincinnati Knee Rating System, and Short Form 36 in Patients With Focal Articular Cartilage Defects

Nicholas J. Greco,^{*} Allen F. Anderson,[†] MD, Barton J. Mann,[‡] PhD, Brian J. Cole,[§] MD, MBA, Jack Farr,^{||} MD, Carl W. Nissen,[¶] MD, and James J. Irrgang,^{*#} PT, PhD, ATC, FAPTA
From the ^{}University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, [†]Tennessee Orthopedic Alliance, Nashville, Tennessee, [‡]American Orthopaedic Society for Sports Medicine, Rosemont, Illinois, [§]Rush University Medical Center, Chicago, Illinois, ^{||}OrthoIndy Cartilage Restoration Center of Indiana, Indianapolis, Indiana, and [¶]Connecticut Children's Medical Center, Farmington, Connecticut*

Background: The International Knee Documentation Committee Subjective Knee Form (IKDC SKF) is a patient-reported knee-specific outcome measure that has been shown to be a reliable, valid, and responsive measure for patients with a variety of knee conditions. Further testing is required to compare the reliability and responsiveness of the IKDC SKF to other commonly used patient-reported outcome measures for patients with articular cartilage lesions.

Hypothesis: The IKDC SKF has equal or better levels of reliability and responsiveness than the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), modified Cincinnati Knee Rating System (CKRS), and the Short Form 36 in patients with articular cartilage lesions.

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

Methods: Reliability was assessed by administering the 4 patient-reported outcome measures to 17 individuals who had undergone articular cartilage surgery 5 years before participation in this study. Responsiveness was determined by administering the 4 patient-reported outcome measures to 51 individuals with diagnosed focal articular cartilage defects who were scheduled to undergo surgical treatment. In both groups, the outcome measures were administered at baseline and at 6 and 12 months' follow-up. Participants also provided a global rating of change in comparison to baseline at the 6- and 12-month follow-ups.

Results: Test-retest reliability coefficients were 0.91 and 0.93 for the IKDC SKF at the 6- and 12-month follow-ups, respectively. The effect sizes and standardized response means were large (>0.80) at 6 months after surgery for the WOMAC pain, physical function, and total scores and 12 months after surgery for the IKDC SKF; WOMAC pain, physical function, and total; and CKRS scores. Six months after surgery, significant differences between those who were improved compared with those who were unchanged or worse were found only for the IKDC SKF. Twelve months after surgery, significant differences between the improved and unchanged groups were found for all of the knee-specific patient-reported outcome measures. Finally, the IKDC SKF, WOMAC, and CKRS scores were able to differentiate between individuals who perceived themselves to be improved versus not improved and the minimum clinically important difference for the IKDC SKF was 6.3 at 6 months and 16.7 at 12 months.

Conclusion: The reliability and responsiveness of the IKDC SKF is comparable with other commonly used patient-reported outcome measures for patients with articular cartilage lesions. The IKDC SKF is a suitable alternative to other commonly used knee-specific instruments for measuring symptoms, daily function, and level of symptom-free sports activity in patients undergoing articular cartilage surgery.

Keywords: patient-reported outcomes; reliability; responsiveness; IKDC Subjective Knee Form

The International Knee Documentation Committee (IKDC) Subjective Knee Form was created as a patient-reported knee-specific assessment of symptoms, function during daily activity, and the level of symptom-free sports activity. Previous evaluation of interventions for the knee have centered on the use of empiric assessments. The subjective nature of these analyses had led to flawed conclusions concerning the efficacy of surgical and nonsurgical interventions for the knee. Likewise, the sheer number of different scales used to assess treatment makes it difficult to compare the efficacies of treatment. Therefore, the IKDC Standard Knee Form was developed to serve as a single assessment for which the effects of multiple knee conditions and treatments could be compared to each other.

The initial IKDC Standard Knee Form included single items for patient-reported knee symptoms and function as well as examination of range of motion and knee laxity.¹¹ Subsequent revisions of the IKDC Knee Form modified the examination procedures and added a demographic component, current health assessment (the Short Form 36 [SF-36]), and a patient-reported outcome measure, the IKDC Subjective Knee Form.

Prior psychometric testing has determined that the IKDC Subjective Knee Form is a reliable and valid instrument to measure symptoms, function during daily activities, and sports activity in patients with different knee problems such as ligament or meniscal injuries, articular cartilage lesions, osteoarthritis, and patellofemoral pain.¹³ The IKDC Subjective Knee Form score has also been shown to be responsive for detecting change over time,¹⁴ and age- and sex-specific normative data have been established to facilitate interpretation of the score.²

In the initial validation study, test-retest reliability was assessed by administering the IKDC Subjective Knee Form score twice over an average of 50 days to 33 patients who had undergone a variety of knee ligament and meniscus surgical procedures at an average of 2.9 years after the index surgery. Responsiveness of the IKDC Subjective Knee Form Score was assessed by evaluating the responses from 207 individuals who were treated operatively or nonoperatively for a variety of knee problems over an average follow-up of 19 months. Because the reliability and responsiveness of an outcome measure are not a fixed property of the outcome measure itself but are dependent upon the specific use and application of the outcome measure,¹ the reliability and responsiveness estimates that have been established for patients with a variety of knee problems may not generalize to patients who undergo articular cartilage surgery. To use the IKDC Subjective Knee Form to evaluate the outcome of articular cartilage surgery, it is important to determine its reliability and responsiveness in individuals who have undergone articular cartilage surgery. Additionally, it is important to compare the reliability and responsiveness of the IKDC Subjective Knee Form to

other knee-specific and general measures of health status for individuals undergoing articular cartilage surgery.

The purpose of this study was to determine the reliability and responsiveness of the IKDC Subjective Knee Form and compare its reliability and responsiveness to 3 other commonly employed patient-reported outcome measures to assess knee function in patients with articular cartilage lesions. In this study, reliability was operationally defined as the consistency of measurement over time in a cohort of subjects who were expected to remain stable over time, and responsiveness was defined as the ability of the outcome instruments to detect change over time in a cohort of subjects whose condition was expected to improve with time. The instruments that were chosen for comparison with the IKDC Subjective Knee Form included the modified Cincinnati Knee Rating System (CKRS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and the SF-36 Health Status Survey. To evaluate responsiveness, these patient-reported outcome measures were administered to patients with diagnosed focal articular cartilage defects who were scheduled to undergo surgical treatment. Reliability was assessed by administering the patient-reported outcome measures to patients who had undergone articular cartilage surgery several years before participation in this study and were therefore presumed to be clinically stable.

METHODS

Participants

Individuals between 18 and 65 years of age were included in this study, which consisted of 2 cohorts of patients who were expected to undergo differential rates of change during a 12-month follow-up period. A treatment cohort, consisting of 73 patients with a primary diagnosis of an articular cartilage defect of the knee who were scheduled to undergo surgical intervention to repair the defect, was used to assess responsiveness. Surgical treatment included debridement, shaving, drilling, autologous chondrocyte implantation (ACI), abrasion arthroplasty, microfracture, and cell therapy. Patients were included in the treatment cohort if they had at least 1 symptomatic full-thickness (Outerbridge grade III or IV) chondral lesion of the femoral condyle or trochlea requiring surgical treatment and if they had a grade II or less cartilage lesion on the tibia and patella. Potential participants were excluded from this group if they did not meet the above-mentioned criteria, or if they had widespread arthritis in the involved joint, history of total meniscectomy in the involved compartment of the knee, required treatment of both knees, had a bipolar defect in which there were opposing lesions on the femur and tibia, or required concurrent total meniscectomy or meniscal allograft in the involved knee.

#James J. Irrgang, PT, PhD, ATC, FAPTA, Associate Professor and Director of Clinical Research, Department of Orthopaedic Surgery, University of Pittsburgh School of Medicine, Room 911 Kaufman Building, 3471 Fifth Avenue, Pittsburgh, PA 15260 (e-mail: jirrgang@pitt.edu).

One or more authors has declared a potential conflict of interest: Support for this study was provided by an unrestricted grant from Genzyme Biosurgery.

A stable cohort was recruited to assess test-retest reliability. This cohort comprised 63 patients with a diagnosed articular cartilage defect of the knee who had been treated with autologous chondrocyte implantation (Carticel, Genzyme Biosurgery, Cambridge, Massachusetts) at least 5 years before the current study. Potential subjects were identified from the Genzyme Cartilage Repair Registry and were included in the group if the cartilage defect met the same criteria as stated for the treatment cohort before ACI and if the defect had been treated with ACI at least 5 years before participation in this study. It was hypothesized that because of the length of time from surgery to participation in this study, the individuals included in the stable cohort would have a relatively constant level of symptoms, daily activity, and sports function over the 12-month period of follow-up. Potential participants were excluded if they underwent ACI on both knees, the ACI treatment failed (ie, the patient had to undergo either another operation to remove the graft or to reimplant cells), or the patient had to undergo a procedure that violated the subchondral bone to treat the defect.

Procedure

Potential participants in the treatment cohort were asked to participate in the study during a preoperative consultation. Subsequently, the surgeon or a research associate at the participating site explained all study procedures, had the individual review the knee forms and consent form, and obtained his or her signature on the consent form. Individuals who agreed to participate in the study were then asked to complete the baseline measures, which included a demographic form, the SF-36, modified CKRS, WOMAC, and IKDC Subjective Knee Form. To avoid the effects that the order of presentation of the outcome measures may have on the results, the outcome instruments were randomly organized into booklets for each participant at each time of administration. As such, the order in which the outcome measures were completed was not the same for each individual and each participant completed the forms in a different order at each follow-up period. The surgeon was responsible for completing an orthopaedic history and recording data from the examination using the IKDC Knee History and Knee Examination forms. The history and examination data were used only to confirm eligibility and for descriptive purposes. No attempt was made to combine the history and examination data with the IKDC Subjective Knee Form to determine an overall IKDC rating of the knee. During surgery, the surgeon noted various characteristics of the involved knee that were documented using the IKDC Surgical Documentation Form. If the stated eligibility criteria were satisfied, the patient continued in the study and completed the 4 patient-reported outcome measures 6 and 12 months after surgery. Six and 12 months after surgery, the participants also concurrently rated their change in knee function from baseline using a 7-point global rating scale. The 7-point scale included the categories much worse, somewhat worse, slightly worse, no change, slightly better, somewhat

better, and much better relative to the first time they had completed the assessment forms. The standing on this scale was used to determine if the participant perceived his or her condition to be improved or not improved after undergoing treatment, and this classification served as the criterion measure of change in the analysis of responsiveness. The follow-up patient-reported outcome measures and global rating of change were completed either at a scheduled follow-up appointment or via mail. Participants who completed both the global rating of change and the patient-reported outcome measures 6 and 12 months after surgery were included in the final analysis.

Those eligible to be included in the stable cohort were mailed a packet containing a letter of invitation that explained the study, consent form, demographic form, and the 4 patient-reported outcome measures. Patients who returned these forms were provisionally included into the study. These patients were then excluded from the study if their modified Cincinnati Score was less than 5, their modified Cincinnati Score changed more than 2 points since their rating 5 years after ACI, or they reported an additional knee injury, surgery, or failed ACI during the time after their 5-year registry evaluation. Given these inclusion and exclusion criteria, we believed that the patient's condition had a high probability of remaining stable over the 12-month period of follow-up. Those who passed the exclusion criteria were accepted into the stable cohort and they were mailed the 4 patient-reported outcome instruments and the global rating of change question 6 and 12 months after the baseline measures were returned. Similar to the treatment cohort, the 4 patient-reported outcome measures were included in random order in a booklet so that the order of the outcome measures was different for each participant each time the outcome measures were administered.

Data Management and Analysis

All data were entered into a computerized database for subsequent analysis. The data analysis was completed using SPSS version 16.0 for a personal computer (SPSS Science Inc, Chicago, Illinois). The scores for each patient-reported outcome instrument were calculated according to the scoring criteria for each of the instruments. For the WOMAC, we calculated subscale scores for pain, stiffness, and physical function as well as a total WOMAC score that considered the item responses for all of the items. The WOMAC scores were transformed to a 0 to 100 point scale, where 100 represented the absence of disability. This was done so that the WOMAC scores could be interpreted in the same manner as the IKDC Subjective Knee Form, modified CKRS, and SF-36 scale scores, where 100 was the most optimal score. The 8 SF-36 scale scores were also combined into physical and mental component summary scores using established weights for each of the SF-36 scores. The physical and mental component summary scores were expressed as standardized scores, in which 50 represents the United States population with a standard deviation of 10. Thus, a physical summary score of 55 indicates the patient had a score that

was one half a standard deviation above the United States population average. Descriptive statistics, including frequencies for categorical variables and measures of central tendency (means, medians) and dispersion (standard deviations, ranges), were calculated for all variables. Change scores from baseline to 6 months and baseline to 12 months were calculated for each instrument. Baseline scores were subtracted from the follow-up scores so that a positive change score represented improvement in the involved knee from baseline.

Reliability Analysis

Participants in the stable cohort were used to estimate reliability of the 4 patient-reported outcome instruments over a 6- and 12-month period. Intraclass correlation coefficients, using a 2-way random effects model in which participants and time were both considered to be random, were used to estimate reliability. It was hypothesized that participants in the stable cohort would have little change in knee function during the study, and this would be reflected by scores on the instruments that would be stable over time. The standard error of measurement was also calculated to determine measurement precision by multiplying the square root of 1 minus the intraclass correlation coefficient by the standard deviation of the baseline scores for the instrument.^{3,8} The minimum detectable change based on the 95% confidence interval was then computed by multiplying the standard error of measurement by 1.96 (for the 95% confidence interval) and the square root of 2. This value represents the amount of change in the score that is necessary to be deemed greater than measurement error.³

Analysis of Responsiveness

Group-Level Analysis of Within-Subject Observed Change From Before to After Treatment. The first construct of change we used for the responsiveness analysis was a group-level analysis of the within-subject observed change from baseline to posttreatment for all subjects.^{3,18} This analysis of responsiveness assumed that all subjects in the treatment cohort would experience improvement 6 and 12 months after articular cartilage surgery. A 1-way repeated-measures analysis of variance was used to determine if, overall, subjects in the treatment cohort improved as hypothesized. The 6- and 12-month change scores were also used to calculate effect sizes and standardized response means. These statistics relate the average change in the scores before and after treatment to the standard deviation of the baseline and change scores, respectively.¹⁸ Effect size and standardized response mean values of 0.8 are considered to indicate a high degree of responsiveness and values of 0.5 are considered to indicate a moderate degree of responsiveness.^{5,7,10,12,16} It was believed that, overall, subjects would improve after treatment of their cartilage defect, and this would therefore result in a large effect size and standardized response mean 6 and 12 months after surgery.

Group-Level Analysis of Between-Group Differences. The second construct of change was a group-level analysis of the between-groups difference of the change scores.^{3,18} The criterion measure of change to create the groups for this analysis was the participant's global rating of change in function from baseline to follow-up. Three groups were created based on the patients' global rating of change. The "improved" group consisted of individuals who had a global rating of change of much better or somewhat better. The "unchanged" group consisted of individuals with a global rating of change of slightly better, not changed, and slightly worse. The "worse" group was composed of participants who had a global rating of change of somewhat worse or much worse. To evaluate responsiveness, the 6- and 12-month change scores were compared between the participants in each of the 3 groups. Because of the relatively small number of patients in each of these groups, we used the nonparametric Kruskal-Wallis test to detect significant differences in the change scores between the groups. Post hoc tests were performed with the Mann-Whitney test to detect significant differences in change scores between each of the groups. When performing the post hoc tests to account for the multiple comparisons, we used a Bonferroni approach to determine a priori the level of significance to be .017 (eg, α of .05 divided by 3 planned comparisons). It was hypothesized that the magnitude of the change scores would be positively associated with the patient's perceived global change after treatment. Specifically, it was hypothesized that the average change score in those who were improved would be greater than the change score in those who were unchanged and those who were worse. Additionally, it was hypothesized that the change score in those who were unchanged would be greater than the change score in those who were worse. Instruments that consistently showed significant differences in change scores between the groups would be considered to be more responsive than those instruments that did not.

Individual-Level Analysis of Between-Groups Difference. The third analysis of change involved an individual-level analysis of between-group differences of the change scores.^{3,18} The purpose of this particular analysis was to determine the minimum clinically important difference, which is the change score that serves as the optimal cut-off point for discriminating between individuals who perceive themselves to be improved from those who do not.^{9,15} To conduct this analysis, the participants' global ratings of change were dichotomized by grouping patients whose global rating of change was much better or somewhat better into 1 group ("improved" group), and those whose rating of global change was slightly better, not changed, slightly worse, somewhat worse, or much worse into another group ("not improved" group). Using the dichotomized criterion measure of change, the sensitivity and specificity were calculated for each observed change score for each instrument.⁹ Sensitivity of change is the proportion of the participants who were improved based on the dichotomized criterion measure of change that had a change score at or above the cut-off point for the change score. Specificity of change is the proportion of participants who did not improve and who had a change score below the

cut-off point. Sensitivity and specificity of change were calculated using different change scores to serve as the cut-off point between individuals who improved and those who were not improved. To determine the optimal cut-off point for the change scores between the improved and unimproved groups of participants (ie, the minimum clinically important difference), a receiver operating characteristic (ROC) curve was constructed by plotting sensitivity on the vertical axis and 1-specificity on the horizontal axis. The area under the ROC curve is defined as the probability of identifying an improved patient on the basis of the change score from randomly selected pairs of improved and unimproved patients^{9,12} and was used to compare responsiveness between instruments, with a larger area indicating a higher level of responsiveness. The value of the change score that was closest to the upper left-hand corner of the ROC curve was determined to be the minimum clinically important difference.^{3,18} To determine this point mathematically, the product function of the sensitivity and specificity was maximized for each instrument score. The change score corresponding to this maximal product of specificity and sensitivity was then taken to be the minimum clinically important difference.

RESULTS

Reliability Analysis

Description of the Stable Cohort. Of the 64 patients eligible to participate in the stable cohort, 49 completed the follow-up patient-reported outcome measures and rated their global change in function from baseline to 6 and 12 months. Contrary to our hypothesis, some subjects reported a change in their status over the 12-month follow-up period. As a result of this, we limited the analysis of reliability to 17 patients who reported their status was unchanged at the 12-month follow-up. The average age of the 17 individuals who were included in the reliability analysis was 43.8 years (standard deviation, 10.4 years; range, 21-60 years); 61.9% were male. The differences in age and gender between the patients who were included and excluded in the analysis of reliability were not significantly different. Likewise, there were no significant differences in the baseline patient-reported outcome measures between those who were included and excluded from the final analysis of the stable cohort.

The scores for each instrument at baseline, 6 months, and 12 months for the stable cohort are displayed in Table 1, as are the intraclass correlation coefficients (ICCs), standard error of measurement, and minimal detectable change. The ICCs for the IKDC Subjective Knee Form were 0.91 and 0.93 for the 6- and 12-month follow-up periods, respectively. The ICCs for the WOMAC physical function and total scores at the 6-month follow-up were both 0.93. At 12 months, the ICCs for the WOMAC physical function and total scores were both 0.86. The ICCs for the WOMAC pain and stiffness scales at 6 and 12 months were slightly lower. The ICCs for the modified CKRS were 0.91 and 0.80 at the 6- and 12-month follow-ups, respectively. The ICCs for the 8 scales of the SF-36 at 6 months ranged from 0.31 to 0.92,

and from 0.36 to 0.95 at 12 months. For the physical component summary score, the ICCs were 0.92 and 0.95 at 6 and 12 months, respectively, which were substantially higher than the ICCs for the mental component summary score, which were 0.58 and 0.46 at 6 and 12 months, respectively.

At 6 months, the standard error of measurement and minimal detectable change for the IKDC Subjective Knee Form and modified CKRS were similar; however, they were both larger than the standard error of measurement and minimal detectable change for the WOMAC physical function and total scores. At 12 months, the standard error of measurement and minimal detectable change were similar for the IKDC Subjective Knee Form and the WOMAC physical function and total scores and they were lower than the standard error of measurement and minimal detectable change for the modified CKRS score. The standard error of measurement and minimal detectable change for the SF-36 scale scores were generally larger than those found for the knee-specific patient-reported outcome measures at both 6 and 12 months' follow-up. The standard error of measurement for the SF-36 summary component scores ranged from 2.4 to 3.6, and the minimal detectable change ranged from 6.6 to 10.0. Because of the difference in the scale for the summary component scores and the other patient-reported outcome scores, it should be noted that the magnitude of the standard error of measurement and minimal detectable change are not directly comparable with those of the IKDC Subjective Knee Form, WOMAC, modified CKRS, and the 8 SF-36 scale scores.

Analysis of Responsiveness

Description of the Treatment Cohort. Of the 73 participants recruited for the treatment cohort, 51 completed the follow-up patient-reported outcome measures and the global rating of change at 6 and 12 months and were included in the analyses of responsiveness. Of the 51 participants included in the analysis, 1 did not provide his or her birth date. Of the 50 remaining participants, the mean age was 36.6 years (standard deviation, 9.7 years; range, 18-56 years); 60.8% were male. The mean age of the 22 patients who were excluded from the analysis was 35.1 years (standard deviation, 11.9 years; range, 19-58 years); 76.2% were men. The differences in age and gender between the individuals who were included and excluded in the analysis were not significant. Additionally, there were no significant differences in any of the baseline patient-reported outcome scores between those who were included and excluded in the final analysis of responsiveness. The physicians made diagnoses after completing a comprehensive history and physical and arthroscopic examination of the subjects. These diagnoses and surgical procedures are listed in Appendix 1 (see online Appendix for this article at <http://ajs.sagepub.com/supplemental/>).

Group-Level Analysis of Within-Subject Observed Change From Before to After Treatment. The scores for each instrument at baseline and at 6 and 12 months for the treatment cohort are presented in Appendix 2 (see online Appendix for this article at <http://ajs.sagepub.com/supplemental/>). A ceiling effect was experienced in all

TABLE 1
Scores, Reliability, Standard Error of Measurement, and Minimal Detectable Change for the
IKDC Subjective Knee Form, WOMAC, Modified CKRS, and SF-36^a

Outcome Measure	Baseline		6 Months			12 Months			
	Score	Score	ICC	SEM	MDC	Score	ICC	SEM	MDC
IKDC	75.9 ± 18.7	72.6 ± 24.8	0.91 (0.76, 0.97)	5.6	15.6	74.1 ± 19.2	0.93 (0.82, 0.98)	4.9	13.7
WOMAC									
Pain	10.3 ± 13.4	11.8 ± 18.1	0.81 (0.55, 0.93)	5.8	16.2	9.7 ± 14.9	0.85 (0.64, 0.94)	5.2	14.4
Stiffness	19.1 ± 22.1	19.9 ± 21.7	0.86 (0.64, 0.95)	8.3	22.9	17.2 ± 18.2	0.75 (0.41, 0.91)	11.1	30.6
Physical function	10.4 ± 14.5	11.5 ± 16.2	0.93 (0.81, 0.98)	3.8	10.6	9.5 ± 14.2	0.86 (0.65, 0.95)	5.4	15.0
Total	11.0 ± 14.8	11.8 ± 16.4	0.93 (0.82, 0.98)	3.9	10.9	10.2 ± 14.6	0.86 (0.65, 0.95)	5.5	15.3
CKRS	78.2 ± 18.4	75.5 ± 17.3	0.91 (0.77, 0.97)	5.5	15.3	76.4 ± 15.3	0.80 (0.48, 0.93)	8.2	22.8
SF-36									
Physical function	80.0 ± 23.0	83.5 ± 21.0	0.90 (0.76, 0.96)	7.3	20.2	82.4 ± 21.7	0.97 (0.90, 0.99)	4.0	11.0
Role physical	77.9 ± 38.4	82.4 ± 36.2	0.85 (0.65, 0.94)	14.9	41.2	85.3 ± 28.0	0.89 (0.70, 0.96)	12.7	35.3
Bodily pain	78.4 ± 22.0	75.3 ± 22.7	0.79 (0.51, 0.92)	10.1	27.9	80.8 ± 23.6	0.81 (0.55, 0.93)	9.6	26.6
General health	82.8 ± 12.5	79.0 ± 11.5	0.61 (0.22, 0.84)	7.8	21.6	83.2 ± 11.4	0.62 (0.18, 0.85)	7.7	21.4
Vitality	71.5 ± 16.1	70.6 ± 15.5	0.70 (0.34, 0.88)	8.8	24.4	74.7 ± 13.8	0.45 (-0.04, 0.77)	11.9	33.1
Social function	91.2 ± 14.5	91.9 ± 13.9	0.84 (0.61, 0.94)	5.8	16.1	93.4 ± 10.9	0.46 (-0.01, 0.77)	10.7	29.5
Role emotional	94.1 ± 13.1	96.1 ± 11.1	0.31 (-0.20, 0.69)	10.9	30.2	100.0 ± 0.0	0.36 (0.09, 0.58)	13.1	36.3
Mental health	82.9 ± 9.7	82.4 ± 11.7	0.65 (0.27, 0.89)	5.7	15.9	79.0 ± 18.0	0.40 (-0.06, 0.73)	7.5	20.8
PCS	48.7 ± 10.6	49.1 ± 9.7	0.92 (0.80, 0.97)	3.0	8.3	50.3 ± 10.8	0.95 (0.82, 0.98)	2.4	6.6
MCS	56.9 ± 4.9	56.1 ± 4.4	0.58 (0.16, 0.82)	3.2	8.8	56.2 ± 7.2	0.46 (-0.03, 0.77)	3.6	10.0

^aScores are presented as mean ± standard deviation. IKDC, International Knee Documentation Committee; CKRS, Modified Cincinnati Knee Rating System; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; SF-36, Short Form 36; ICC, intraclass correlation coefficient (95% confidence interval of ICC); SEM, standard error of measurement; MDC, minimum detectable change based on 95th percentile; PCS, physical component summary score; MCS, mental component summary score.

instruments except the IKDC Subjective Knee Form. Corresponding changes in the scores from baseline are presented in Table 2, along with the effect sizes and standardized response means at 6 months and 12 months. One-way repeated measures analysis of variance demonstrated that there were significant differences in the scores for all of the patient-reported outcome measures across time except for the SF-36 general health score and the SF-36 mental component summary score.

The effect sizes for the WOMAC pain, physical function, and total scores were greater than 0.80 at both follow-up time points. The effect sizes for the IKDC Subjective Knee Form score were 0.76 and 1.06 at 6 and 12 months, respectively. The effect size for the modified CKRS was 0.60 at 6 months and 1.09 at 12 months. At 6-month follow-up, the standardized response means were greater than 0.80 for the WOMAC pain, physical function, and total scores, while the standardized response means for the IKDC Subjective Knee Form and modified CKRS were 0.57 and 0.52, respectively. Twelve months after surgery, the IKDC Subjective Knee Form and WOMAC pain, physical function, and total scores all had standardized response means greater than 0.80. The effect sizes and standardized response means for the 8 SF-36 scale and 2 summary component scores were less than 0.70 at the 6- and 12-month follow-up points.

Group-Level Analysis of Between-Group Differences. As expected, the 6- and 12-month change scores generally increased as the participants' global rating of change improved (Appendix 3 and 4; see online Appendix for this

article at <http://ajs.sagepub.com/supplemental/>). Because of the small sample sizes for some of the groups defined by the global rating of change, we collapsed the categories for further analyses into 3 groups (improved, no change, worse). The change scores for these 3 groups are reported in Appendix 5 (see online Appendix for this article at <http://ajs.sagepub.com/supplemental/>). In general, the change scores followed the hypothesized trends except for the WOMAC stiffness score, which was lower in the "no change" group compared to the group of patients who rated themselves worse at 6 months. At 12 months, all of the WOMAC change scores were greater in the group of individuals who rated themselves worse compared to the group of those who rated themselves unchanged.

The results of the Mann-Whitney tests indicated that there were significant differences in the IKDC Subjective Knee Form change scores between the improved and unchanged and worse groups at both the 6- and 12-month follow-ups. For the other patient-reported outcome measures, there were only a few significant differences between the change scores of the groups at the 6-month follow-up. More specifically, only the WOMAC pain and modified CKRS scores demonstrated significant differences in the change scores at 6 months between those who were improved and those who were worse. At the 12-month follow-up, there were significant differences between the change scores of patients in the improved group relative to the unchanged group for all scores except the SF-36 bodily pain, general health, mental health, and mental component summary scores.

TABLE 2
Change Scores, Effect Sizes, and Standardized Response Means at 6 and 12 Months^a

Outcome Measure	6 Months			12 Months		
	Mean \pm SD	ES	SRM	Mean \pm SD	ES	SRM
IKDC	11.5 \pm 20.4	0.76	0.57	19.4 \pm 19.2	1.06	1.00
WOMAC						
Pain	17.8 \pm 19.6	0.98	0.91	20.8 \pm 22.8	1.14	0.94
Stiffness	10.1 \pm 24.9	0.51	0.40	16.9 \pm 26.3	0.72	0.64
Physical function	16.2 \pm 18.8	0.88	0.86	21.8 \pm 19.3	1.20	1.13
Total	16.0 \pm 17.6	0.96	0.91	21.2 \pm 18.7	1.19	1.13
CKRS	13.1 \pm 25.4	0.60	0.52	21.7 \pm 28.6	1.09	0.76
SF-36						
Physical function	10.2 \pm 31.1	0.35	0.33	18.4 \pm 31.4	0.67	0.59
Role physical	9.3 \pm 46.1	0.21	0.20	26.0 \pm 47.2	0.63	0.55
Bodily pain	13.8 \pm 24.0	0.63	0.57	14.2 \pm 31.6	0.60	0.45
General health	2.4 \pm 14.8	0.11	0.16	1.26 \pm 21.6	0.06	0.06
Vitality	6.2 \pm 21.3	0.32	0.29	8.6 \pm 21.6	0.43	0.01
Social function	8.1 \pm 28.0	0.30	0.29	15.2 \pm 26.1	0.57	0.58
Role emotional	14.4 \pm 29.6	0.34	0.49	15.7 \pm 37.9	0.48	0.41
Mental health	2.3 \pm 19.2	0.13	0.12	6.1 \pm 18.0	0.32	0.34
PCS	4.5 \pm 11.1	0.44	0.41	6.4 \pm 12.1	0.57	0.53
MCS	3.0 \pm 10.0	0.30	0.30	2.8 \pm 9.7	0.27	0.29

^aSD, standard deviation; ES, effect size; SRM, standardized response mean; IKDC, International Knee Documentation Committee; CKRS, Modified Cincinnati Knee Rating System; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; SF-36, Short Form 36; PCS, physical component summary score; MCS, mental component summary score.

Individual-Level Analysis of Between-Groups Difference. The ROC curves for the IKDC Subjective Knee Form, WOMAC total, and modified CKRS scores for the 6- and 12-month follow-ups are presented in Figures 1 and 2. The area under the ROC curve, minimum clinically important difference, and the sensitivity and specificity for the minimum clinically important differences are displayed in Table 3. At 6 months, the area under the ROC curve was significantly different from 0 for the IKDC Subjective Knee Form; WOMAC pain, physical function, and total; modified CKRS; and SF-36 physical function, vitality, and physical component summary scores. The area under the curve for all these scores as well as for the WOMAC stiffness score was significantly different from 0 at 12 months after surgery. Six months after surgery, the area under the curve for the IKDC Subjective Knee Form (0.75) was slightly larger than the area under the curve for the WOMAC pain (0.73) and total (0.71), modified CKRS (0.72), and SF-36 physical function (0.70) and physical component summary (0.70) scores. Similarly, 12 months after surgery, the area under the curve for the IKDC Subjective Knee Form (0.78) was larger than the area under the curve for the WOMAC pain (0.70), physical function (0.70), and total (0.71); modified CKRS (0.75); and the SF-36 physical function (0.76), role physical (0.71), vitality (0.71), and physical component summary (0.75) scores. However, because of the overlapping confidence intervals, it is not possible to state that the IKDC Subjective Knee Form was more responsive than the other patient-reported outcome measures.

The minimum clinically important differences and the associated sensitivity and specificity of change for the

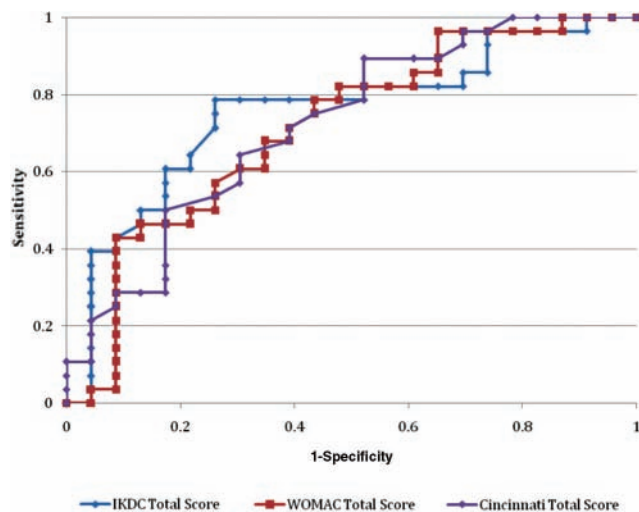


Figure 1. Receiver operating characteristic curve for change scores at 6-month follow-up. IKDC, International Knee Documentation Committee; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

minimum clinically important differences were determined for the outcome measures that had a significant area under the curve (Table 3). The minimum clinically important difference for the IKDC Subjective Knee Form was 6.3 and 16.7 at 6 and 12 months, respectively. For comparison, the minimum clinically important difference for the WOMAC total score was 11.5 at both the 6- and 12-month follow-up and 14.0 and 26.0 for the modified CKRS 6 and

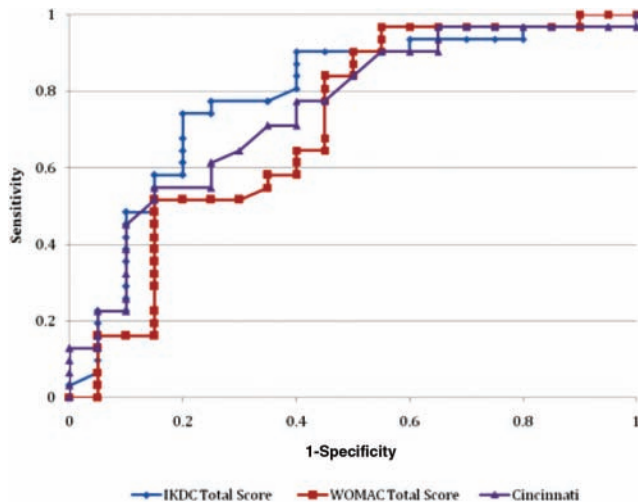


Figure 2. Receiver operating characteristic curve for change scores at 12-month follow-up. IKDC, International Knee Documentation Committee; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

12 months after surgery, respectively. At both the 6- and 12-month follow-ups, the minimum clinically important difference for the IKDC Subjective Knee Form score had the highest combination of sensitivity and specificity in comparison to all the other outcome measures, indicating that the minimum clinically important difference for the IKDC Subjective Knee Form may better discriminate between those who perceive themselves to be improved from those who do not perceive themselves to be improved.

DISCUSSION

The results of this study indicate that the IKDC Subjective Knee Form has comparable reliability and responsiveness to the WOMAC, modified CKRS, and the SF-36 physical function and physical component summary scores when evaluating patients following articular cartilage surgery. As such, the IKDC Subjective Knee Form is a suitable alternative to other commonly used knee-specific instruments for measuring symptoms, daily function, and level of symptom-free sports activity in patients undergoing articular cartilage surgery.

Reliability Analysis

Test-retest reliability is the degree to which the score remains stable when the condition being measured remains stable.¹³ To assess reliability, we identified a cohort of patients who underwent ACI 5 years before participation in this study and hypothesized that their level of symptoms, daily activity, and sports function would remain constant over the 1-year course of follow-up. However, contrary to our hypothesis, only 17 individuals rated the status of their knee as unchanged 12 months after completion of the baseline forms. Therefore, to accurately

estimate reliability, we limited the analysis of test-retest reliability to only those participants who reported no change in knee function. In the stable cohort, there was little change in the IKDC Subjective Knee Form scores from baseline to the 6- and 12-month follow-ups. The ICC for test-retest reliability for the IKDC Subjective Knee Form score was 0.91 at the 6-month follow-up and 0.93 at the 12-month follow-up. The ICCs for the WOMAC physical function and total scores at 6 months were similar (both 0.93), but were substantially lower than the IKDC Subjective Knee Form at 12 months. The ICCs for the modified CKRS were 0.91 and 0.80 at 6 and 12 months, respectively. Reliability of the 8 SF-36 scale and 2 summary component scores was more variable. At 6 and 12 months of follow-up, the physical function and physical component summary scores had the highest level of reliability. The reliability for the IKDC Subjective Knee Form in this study is similar to the ICC for test-retest reliability of the IKDC Subjective Knee Form of 0.94 when it was assessed over a shorter period of time (50 days) in a sample of subjects participating in a variety of long-term clinical outcome studies approximately 2.9 years after surgery.¹³

The standard error of measurement and minimal detectable change for the IKDC Subjective Knee Form were 5.6 and 15.6, respectively, at 6 months and were 4.9 and 13.7, respectively, at 12 months after surgery. Thus, over a 12-month period of time, a change of the IKDC Subjective Knee Form score greater than 13.7 points can be interpreted as a true change beyond measurement error in an individual's level of symptoms, function, and sports activity. At the 6-month follow-up, the standard error of measurement and minimal detectable change for the WOMAC physical function and total scores were slightly smaller than the standard error of measurement and minimal detectable change for the IKDC Subjective Knee Form. This was primarily due to the smaller baseline standard deviations for the WOMAC physical function and total scores in comparison to the IKDC Subjective Knee Form. At the 12-month follow-up, the standard error of measurement and minimal detectable change for the WOMAC physical function and total scores were slightly larger than those for the IKDC Subjective Knee Form. The standard error of measurement and minimal detectable change for the modified CKRS were similar to the IKDC Subjective Knee Form at 6 months, but were substantially higher than the IKDC Subjective Knee Form at 12 months.

Some may question our rationale to estimate test-retest reliability over a 6- and 12-month period of follow-up; however, we assert that this provides a more realistic estimate of measurement error over a period of follow-up that is more typical when evaluating the outcome of surgery. It is important to note that modern psychometric theory implies that the psychometric properties of a patient-reported outcome measure are not a fixed property of the outcome measure itself but are dependent upon the specific use and application of the outcome measure.¹ As such, the reliability of a patient-reported outcome measure is dependent on the circumstances in which the instruments are used. In prior research, test-retest reliability of

TABLE 3
Results of Receiver Operating Characteristic Curve Analysis^a

Outcome Measure	6 Months				12 Months			
	Area Under Curve (95% Confidence Interval)	MCID	Sensitivity	Specificity	Area Under Curve (95% Confidence Interval)	MCID	Sensitivity	Specificity
IKDC	0.75 ^b (0.62-0.89)	6.3	0.79	0.74	0.78 ^b (0.65-0.92)	16.7	0.74	0.8
WOMAC								
Pain	0.73 ^b (0.59-0.87)	17.5	0.68	0.70	0.72 ^b (0.57-0.87)	7.5	0.87	0.5
Stiffness	0.64 (0.49-0.801)	6.3	0.68	0.48	0.67 ^c (0.53-0.82)	18.8	0.55	0.65
Physical function	0.68 ^c (0.53-0.83)	8.1	0.79	0.52	0.70 ^c (0.54-0.86)	5.89	0.94	0.50
Total	0.71 ^b (0.57-0.86)	11.5	0.79	0.57	0.71 ^c (0.55-0.86)	11.5	0.84	0.55
CKRS	0.72 ^b (0.58-0.86)	14.0	0.64	0.70	0.75 ^b (0.61-0.89)	26.0	0.55	0.85
SF-36								
Physical function	0.70 ^c (0.56-0.85)	.75	0.61	0.78	0.76 ^b (0.62-0.90)	17.5	0.58	0.92
Role physical	0.63 (0.49-0.79)				0.71 ^c (0.56-0.85)	12.5	0.50	0.77
Bodily pain	0.59 (0.43-0.74)				0.57 (0.41-0.74)			
General health	0.57 (0.41-0.74)				0.58 (0.41-0.74)			
Vitality	0.67 ^c (0.52-0.83)	2.6	0.66	0.72	0.71 ^c (0.56-0.86)	2.6	0.65	0.62
Social function	0.60 (0.45-0.76)				0.63 (0.47-0.793)			
Role emotional	0.49 (0.32-0.65)				0.62 (0.46-0.78)			
Mental health	0.62 (0.46-0.78)				0.44 (0.28-0.60)			
PCS	0.70 [*] (0.56-0.84)	4.8	0.67	0.74	0.75 (0.62-0.87)	4.6	0.71	0.74
MCS	0.55 (0.40-0.71)	2.7	0.57	0.63	0.50 (0.35-0.66)	0.3	0.65	0.48

^aMCID, minimum clinically important difference (not calculated for those scores for which the area under the curve was not significantly different from 0); IKDC, International Knee Documentation Committee; CKRS, Modified Cincinnati Knee Rating System; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index; SF-36, Short Form 36; PCS, physical component summary score; MCS, mental component summary score.

^b $P < .01$.

^c $P < .05$.

patient-reported outcome measures has typically been estimated over a relatively short period of time (usually 1 to 4 weeks). We believe that this underestimates the error associated with repeat testing over a longer period of time. When evaluating outcome of knee surgery, such as after articular cartilage surgery, we are interested in changes experienced by the patient over a longer period of time. Given this, we chose to estimate test-retest reliability over 6- and 12-month periods of time. We selected a cohort of individuals who had undergone articular cartilage surgery 5 years before their participation in this study under the assumption that the status of the knee in these individuals would have reached a plateau and should remain relatively constant over a 12-month period of time. However, contrary to our hypothesis, a substantial number of participants reported a change in their knee status. An underlying assumption for estimating test-retest reliability is that the construct being measured should remain stable. Given this, we eliminated those individuals in the stable cohort who reported a change in the status of their knee and estimated reliability for the remaining participants, who reported their status was unchanged. We believe that the reliability of the outcome scores estimated over this longer period of time is a better reflection of the true measurement error that would be observed in individuals who are unchanged over a 6- or 12-month period of time. Thus, we believe that evaluating the stable cohort over a 6- or 12-month period does not introduce a bias, but rather provides a more realistic estimate of measurement error over

a longer period of follow-up that is more typical when evaluating the outcome of surgery.

Responsiveness Analysis

Multiple methods and statistics have been proposed to assess responsiveness of patient-reported outcome measures.^{12,18} Because there is no consensus on the optimal methods and statistics to demonstrate responsiveness, we chose to use multiple methods to provide evidence for responsiveness. This included evaluation of change of a group of patients expected to improve over time, comparing the amount of change between groups that were expected to undergo different rates of change, and determining the amount of change in patients who perceived themselves to be improved from those who did not.

Group-Level Analysis of Within-Patient Observed Change From Before to After Treatment. The first construct of change for analysis of responsiveness was that all participants in the treatment cohort would experience improvement in their knees over the course of 12 months after surgery and that this would be associated with improvements in the scores for the 4 patient-reported outcome measures. This hypothesis was substantiated by improved scores across time for all of the patient-reported outcome measures. Statistical analysis indicated that the scores for all of the patient-reported outcome measures increased over time except for the SF-36 general health score.

Effect sizes and standardized response means were calculated to relate the magnitude of the change scores to the standard deviation of the baseline scores and standard deviation of the change scores, respectively. Six months after surgery, only the WOMAC pain, physical function, and total scores had large (>0.80) effect sizes and standardized response means. The IKDC Subjective Knee Form, WOMAC stiffness, modified CKRS, and SF-36 bodily pain scores demonstrated moderate (>0.50) effect sizes and standardized response means 6 months after surgery. Twelve months after surgery, the IKDC Subjective Knee Form; WOMAC pain, physical function, and total; and modified CKRS scores all demonstrated large effect sizes and standardized response means, while the WOMAC stiffness and SF-36 physical function, role physical, bodily pain, social function, and physical component summary scores demonstrated moderate effect sizes and standardized response means. The decreased responsiveness of the general health status measure in comparison with the knee-specific outcome measures was expected and, as such, general health status measures should not be used as the primary outcome measure for surgical treatment of articular cartilage lesions of the knee.

In retrospect, the decreased responsiveness of the IKDC Subjective Knee Form and modified CKRS scores in comparison with the WOMAC scores should have been expected because of differences in the content of the outcome measures. Both the IKDC Subjective Knee Form and modified CKRS include questions related to sports activity and participation (eg, running, jumping, cutting, pivoting), while the WOMAC contains questions related to function during a variety of daily activities (eg, walking on level surfaces, ascending and descending stairs, putting on socks). Given the differences in the difficulty of the content, greater change in the WOMAC scores in comparison with the IKDC Subjective Knee Form and modified CKRS scores, as evidenced by the smaller effect sizes and standardized response means, should have been expected 6 months after surgery. Twelve months after articular cartilage surgery, many individuals would be expected to have returned to sports activities and participation, which would be reflected in greater change from baseline in the IKDC Subjective Knee Form and modified CKRS scores, resulting in large effect sizes and standardized response means 12 months after surgery. Because of the relative ease of the WOMAC questions, individuals may be expected to have higher WOMAC scores sooner after surgery. This is supported by the relatively large changes from baseline to 6 months and small changes from 6 to 12 months in the WOMAC scores, as well as by the greater ceiling effects for the WOMAC 12 months after surgery. Understanding how the content of patient-reported outcome measures influences responsiveness lends greater understanding to the meaning (ie, the validity) of the outcome score at different time points in recovery after articular cartilage surgery.

Group-Level Analysis of Between-Group Differences. For the second construct of change, participants provided a global rating of change on a 7-point scale, ranging from greatly better to greatly worse from baseline at 6- and 12-month follow-up. We hypothesized that individuals

who had greater improvement from baseline would have greater improvement in the patient-reported outcome measures. The results of this study generally supported our hypothesis; however, there were several exceptions. The group rated slightly better at the 6-month follow-up had greater change scores than the group that was somewhat better, and at 12 months, the group rated somewhat worse had greater change scores than the group rated slightly worse. These inconsistencies are likely the result of the small number of patients in each of these groups as well as the difficulty subjects have in providing a retrospective global rating of change.¹⁷ The group rated slightly better at the 6-month follow-up consisted of 4 patients, while there was only 1 patient in the group rated slightly worse at the 12-month follow-up.

A more focused analysis of the IKDC Subjective Knee Form scores of the individuals in the slightly better group at the 6-month follow-up helps to explain why this group had mean change scores greater than the somewhat better group at 6 months. The mean change score for the IKDC Subjective Knee Form for this group was 13.3, and 3 of the 5 patients had change scores greater than 15. These 3 patients all reported baseline scores less than 30 and 6-month scores in the range of 20 to 51. The other 2 individuals in this group had baseline IKDC Subjective Knee Form scores of 48.3 and 52.9, and 6-month follow-up scores of 52.9 and 50.6, respectively. Reviewing these scores, it is clear that the 3 patients with change scores greater than 15 would have been expected to have a global rating of change of somewhat or much better. These discrepancies illustrate the difficulties in using the global rating of change as a construct of change. It may be difficult for individuals to decide between the 7 categories of the scale when assessing their change in knee function from baseline. It may also be challenging for them to mentally decipher the change in their condition from baseline by quantifying their condition at each time point and then computing the difference. Additionally, individuals may have difficulty recalling their initial condition at baseline and confuse this with their status before injury.¹⁷ Finally, the 3 patients in this group who had change scores greater than 15 still had relatively low IKDC Subjective Knee Form scores that ranged from 21 to 51 at the 6-month follow-up and, compared with their expectation for improvement after surgery, believed that they were only slightly better. These factors may explain why these individuals only rated their change in knee function to be slightly better when their mean change scores appeared to imply a greater improvement. In the future, alternatives for the use of a retrospective global rating of change are needed for responsiveness studies. One such alternative to the global rating of change has been proposed by Beaton et al⁶ and involves determining if the patient's ability to cope with his or her problem has changed from before to after treatment. Improvement is suggested when the patient who was not able to cope with his or her problem before treatment is able to cope with the problem after treatment.

The small number of patients in each group designated by the individual's global rating of change made it impractical to perform statistical testing to determine if the differences in mean change scores between the groups were

significant. Because of this, we collapsed the participants into 3 groups (improved, unchanged, or worse). In general, at the 6-month follow-up, the mean change scores for those who were improved were greater than the change scores for those who were unchanged, and those who were unchanged had greater change scores than those who were worse; however, these differences were not significant for all of the patient-reported outcome measures. For the IKDC Subjective Knee Form, the mean change score for those who were improved was greater than the mean change score for those who were unchanged as well as for those who were worse. Similar results were observed at the 12-month follow-up. At the 12-month follow-up, the IKDC Subjective Knee Form score was the only outcome measure that demonstrated significant differences between those who were improved and those who were unchanged or worse. All the WOMAC, modified CKRS, and the SF-36 physical function, role physical, vitality, social functioning, role emotional, and physical component summary scores demonstrated significant differences at the 12-month follow-up between those who were improved and those who were unchanged, but the mean change scores for those who were improved were not significantly different than for those who were unchanged. At the 6- and 12-month follow-ups, there were no significant differences in the mean change scores for any of the patient-reported outcome measures between those who were unchanged and those who were worse. This may be attributable to the fact that there were only 6 individuals who were worse at 6 months and 4 who were worse at 12 months.

Individual-Level Analysis of Between-Groups Difference. Finally, ROC curves were used to compare the responsiveness and establish the minimum clinically important difference for each patient-reported outcome measure. For this analysis, the participant's global rating of change served as the criterion measure of change for defining improvement from the patient's perspective.³ To construct the ROC curves, participants were dichotomized as improved (greatly or much better) versus not improved (slightly better to greatly worse). The IKDC Subjective Knee Form had the greatest area under the ROC curve at 6 and 12 months, with areas of greater than 0.75 and 0.78, respectively; however, because of the wide overlapping confidence intervals, it is not possible to state that the IKDC Subjective Knee Form was more responsive than the WOMAC or modified CKRS scores. This implies that the IKDC Subjective Knee Form score has a probability similar to the WOMAC and modified CKRS scores of selecting a patient who has perceived improvement in knee function from a patient who has not perceived improvement.

The ROC curves were also used to determine the minimum clinically important difference. The minimum clinically important difference is the value of the change score that best differentiates an improved from an unimproved patient. For the IKDC Subjective Knee Form Score, the minimum clinically important change was 6.3 at 6 months and 16.7 at 12 months. The sensitivity and specificity for the minimum clinically important difference for the IKDC Subjective Knee Form were 79% and 74% at 6 months and 74% and 80% at 12 months. When applying

these results in clinical practice, it would be appropriate to conclude that a patient perceives himself or herself to be improved if the change score at 6 months is greater than or equal to 6.3 or if the change score is greater than 16.7 at 12-month follow-up.^{3,15}

In previous research to evaluate responsiveness of the IKDC Subjective Knee Form, 2 estimates for the minimum clinically important difference were reported in patients with a variety of knee impairments, with an average follow-up of 19.0 months.¹⁴ A change score of 11.5 maximized sensitivity of change at the expense of specificity (sensitivity of 82% and specificity of 64%), and a change score of 20.5 maximized specificity of change at the expense of sensitivity (sensitivity of 64% and specificity of 84%). The minimum clinically important difference 12 months after surgery that was observed in this study of patients undergoing articular cartilage surgery is within the range of the minimum clinically important difference that was reported previously for the IKDC Subjective Knee Form and has a slightly higher combination of sensitivity and specificity. We are not aware of any studies that have provided estimates of the minimum clinically important difference, using methodology similar to that which we used in this study, for the WOMAC, modified CKRS, or SF-36.

The magnitude of the minimum clinically important difference for the IKDC Subjective Knee Form and modified CKRS were much smaller 6 months after surgery than at 12 months after surgery. In contrast, the minimum clinically important difference of the WOMAC total score was similar 6 and 12 months after surgery. We believe that the differences in the magnitude of the minimum clinically important difference at 6 and 12 months for the IKDC Subjective Knee Form and modified CKRS are due to the relatively greater difficulty of the questions for patients 6 months after articular cartilage surgery, which results in a smaller amount of change in these scores 6 months after surgery. Although the differences in the magnitude of the 6- and 12-month minimum clinically important differences for the IKDC Subjective Knee Form and modified CKRS may cause concern for some because they cannot apply a single value, we believe that these results highlight the importance of considering the context in which the outcome measures are used.

In conclusion, we have provided evidence for reliability and responsiveness of the IKDC Subjective Knee Form in patients who have undergone articular cartilage surgery. We believe that the psychometric properties of the IKDC Subjective Knee Form are sufficient to measure change in status for patients undergoing articular cartilage procedures. Further, we believe the IKDC Subjective Knee Form compares favorably to other patient-reported outcome measures that are commonly used to measure outcome after articular cartilage procedures including the WOMAC, modified CKRS, and SF-36. Some differences were observed in the reliability and responsiveness of the 4 patient-reported outcome measures that were evaluated in this study depending on the length of follow-up. We believe this supports the contemporary concept of validity that requires evidence to support the use and interpretation of patient-reported outcome scores for a specific

context that considers the patient population under consideration and the length of follow-up.

REFERENCES

1. American Educational Research Association. *Standards for Educational and Psychological Testing/American Research Association, American Psychological Association, National Council on Measurement in Education, 1-194*. Washington, DC: American Educational Research Association; 1999.
2. Anderson AA, Irrgang JJ, Kocher MS, Mann BJ, Harrast JJ, International Knee Documentation Committee. The International Knee Documentation Committee Subjective Knee Evaluation Form: normative data. *Am J Sports Med*. 2006;34:128-135.
3. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine*. 2000;25:3192-3199.
4. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. *J Rheumatol*. 2001;28:400-405.
5. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol*. 1997;50:79-93.
6. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther*. 2001;14:128-146.
7. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press; 1969.
8. Crocker L, Algina J. *Introduction to Classical and Modern Test Theory*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1986.
9. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: An analogy to diagnostic test performance. *J Chronic Dis*. 1986;39:897-906.
10. Garratt AM, Ruta DA, Abdalla MI, Russell IT. SF36 health survey questionnaire: II. responsiveness to changes in health status in four common clinical conditions. *Qual Health Care*. 1994;3:186-192.
11. Hefti F, Müller W, Jakob RP, Stäubli H-U. Evaluation of knee ligament injuries with the IKDC form. *Knee Surg Sports Traumatol Arthrosc*. 1993;1:226-234.
12. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53:459-468.
13. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med*. 2001;29:600-613.
14. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med*. 2006;34:1567-1573.
15. Irrgang JJ, Marx RG. Clinical outcomes in sport and exercise physical therapies. In: Kolt G, Snyder-Mackler L, eds. *Physical Therapies in Sport and Exercise*. 2nd ed. Edinburgh, UK: Elsevier; 2007: 205-218.
16. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Medical Care*. 1990;28: 632-642.
17. Norman GR. Issues in the use of change scores in randomized trials. *J Clin Epidemiol*. 1989;11:1097-1105.
18. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther*. 1996;76:1109-1123.

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