Published in final edited form as:

Arthritis Care Res (Hoboken). 2017 June ; 69(6): 817–825. doi:10.1002/acr.23193.

Validity and Responsiveness of the Knee injury and Osteoarthritis Outcome Score (KOOS): A Comparative Study among Total Knee Replacement Patients

Barbara Gandek, PhD^{1,2} and John E. Ware Jr., PhD^{1,2}

¹Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

²John Ware Research Group, Watertown, MA, USA

Abstract

Objective—To evaluate validity and responsiveness of the Knee injury and Osteoarthritis Outcome Score (KOOS) in relation to other patient-reported outcome measures before and after total knee replacement (TKR).

Methods—Pre-TKR and 6-month post-TKR data from 1,143 patients in a U.S. joint replacement cohort was used to compare the KOOS, WOMAC and SF-36 Health Survey. Validity was evaluated with multiple methods, including correlations of pre-TKR scale scores and ANOVA models that used pre-TKR data to compare the relative validity (RV) of scales in discriminating between groups differing in assistive walking device use and number of comorbid conditions and used post-TKR minus pre-TKR change scores to assess RV of scales in discriminating between groups rating themselves as better, same or worse (BSW) in their capability to do activities at 6 months. Responsiveness also was described using effect sizes (ES) and standardized response means (SRM).

Results—In support of convergent and discriminant validity, KOOS scale scores were worse for patients using an assistive device but only declined weakly with increasing comorbid conditions. While all knee-specific scales discriminated between BSW groups, the KOOS QOL scale was significantly ($p < 0.05$) better than all measures except the SF-36 Physical Component Summary. KOOS QOL also had the highest ES, while SF-36 measures had lower ES and SRM. KOOS Pain and Symptoms scales discriminated better than WOMAC Pain and Stiffness scales among BSW groups.

Conclusion—KOOS scales were valid and responsive in this cohort of U.S. TKR patients. KOOS QOL performed particularly well in capturing aggregate knee-specific outcomes.

Corresponding Author: Barbara Gandek, PhD, University of Massachusetts Medical School, Department of Quantitative Health Sciences, 368 Plantation Street, Worcester, MA 01605, Phone: 508.856.8999, Fax: 508.856.8993, barbara.gandek@umassmed.edu.

Disclosure Statement: The authors have declared no conflict of interest.

Author contributions

Both authors conceptualized and designed the study, analyzed and interpreted the data, drafted the article and revised it critically for important intellectual content, and read and approved the final version submitted for publication. BG takes responsibility for the integrity of the work as a whole.

The Knee injury and Osteoarthritis Outcome Score (KOOS) was published nearly 20 years ago as a patient-reported outcome (PRO) measure suitable for use among patients with knee osteoarthritis (OA) or knee injuries (1, 2). Subsequently, it has been used in numerous studies including a randomized controlled trial of treatments for knee OA patients eligible for total knee replacement (TKR) (3), a comparison of TKR patients in 22 U.S. states (4), and the Osteoarthritis Initiative (5). Notably, the U.S. Center for Medicare & Medicaid Services (CMS) allows submission of three KOOS scales (Pain, Function in Daily Living, and 2 Stiffness items) in the PRO component of its Comprehensive Care for Joint Replacement (CJR) model, which bundles payment and quality measures for episodes of care (6).

The KOOS has been shown to have adequate reliability, construct validity and responsiveness across a number of patient groups and countries (7). While the measurement properties of the KOOS have been evaluated among knee OA patients in Canada (8) and in European (9–15) and Asian (16, 17) countries, it has had limited psychometric evaluation in the United States. KOOS development included a small pilot study with anterior cruciate ligament (ACL) patients in Vermont (1), and Engelhart evaluated KOOS reliability, validity and responsiveness in a small study of ACL patients in the U.S. and Europe (18). Singh examined test-retest reliability and the minimum clinically important difference of the KOOS Quality of Life scale in a study of 141 U.S. knee OA patients (19), while Steinhoff compared the responsiveness of KOOS scales in 82 U.S. TKR patients (20). However, to the best of our knowledge, there has not been a large-scale evaluation of the validity and responsiveness of the KOOS among TKR patients in the United States.

While many PRO questionnaires are used with TKR patients, comprehensive information about their comparative validity and responsiveness is lacking. This study evaluated the validity and responsiveness of the KOOS in comparison to two of the most widely-used PRO measures in TKR (21, 22), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and SF-36 Health Survey, using data from a U.S. national joint replacement cohort. By conducting a variety of cross-sectional and longitudinal tests, for which there were strong hypotheses as to the results that would be expected for valid knee-specific versus generic measures (described further in Materials and Methods), this paper will increase understanding of the comparative performance of the KOOS in relation to other widely-used PRO measures.

Materials and Methods

Patients

Data came from the Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) cohort of more than 25,000 total joint replacement patients from more than 150 surgeons in 22 U.S. states (23). This analysis was based on data from 1,179 TKR patients randomly selected from high volume surgical centers. Questionnaires were self-administered via scannable paper-pencil (77%) or Internet (23%) administration at the surgeon's office or patient's home. FORCE-TJR and this study were approved by the University of Massachusetts Medical School Internal Review Board.

Measures

The KOOS contains 42 knee-specific items grouped into five scales that measure pain, other symptoms, function in activities of daily living (ADL), function in sport and recreation (Sport), and knee-related quality of life (QOL) (1). All KOOS items were asked in reference to the surgical knee. Scales were scored so that 0 was the worst possible and 100 was the best possible score (24). The KOOS includes all 24 items in the WOMAC (version LK3.0) (9). Therefore, the WOMAC 5-item Pain, 2-item Stiffness and 17-item Function scales (25) were scored from the KOOS items; of note, the KOOS ADL and WOMAC Function scales contain the same 17 items. To be consistent with the KOOS, the WOMAC scales were scored so 0 was the worst possible and 100 was the best possible score. Internal consistency reliability of all scales was evaluated using Cronbach's coefficient alpha (26).

Unlike the KOOS and WOMAC which are joint-specific, the SF-36 Health Survey is a generic measure of health status which is not specific to any diagnosis and thus captures the impact of comorbid conditions as well as the surgical knee (27). The eight SF-36 (Version 2.0) scales were scored so that 0 was the worst possible and 100 was the best possible score (28). Summary Physical (PCS) and Mental (MCS) Component Scores also were calculated from all eight scales (29); PCS and MCS were scored so 50 was the mean and 10 was the standard deviation in the U.S. general population (28). Reliability and validity of the SF-36 have been demonstrated in TKR (22, 30).

Analyses

Construct validity, or the extent to which a scale is more (convergent) or less (discriminant) related to other measures in a manner consistent with theory, was evaluated by conducting cross-sectional and longitudinal tests (31–33). Responsiveness also was described using the effect size and standardized response mean for change scores.

Concurrent validity was evaluated by examining Pearson product-moment correlations among measures of more and less conceptually related scales at baseline (pre-TKR), using a multitrait-multimethod approach (34). Patterns of higher and lower correlations were expected based on item content, the construct measured by each scale, and results from previous KOOS and WOMAC studies (7, 9, 10, 12, 15–17, 35). Knee-specific measures of the same construct (e.g., KOOS Pain, WOMAC Pain) were hypothesized to have higher correlations than correlations of these measures with generic measures of the same construct (e.g., SF-36 Bodily Pain). KOOS Symptoms, QOL and other knee-specific scales were hypothesized to be more highly inter-correlated than with generic SF-36 measures. In addition, while pain and function are conceptually distinct constructs and thus would not be expected to have a high correlation, the operational definitions used in the WOMAC Pain and Function scales are known to be confounded because items about the same activities are included in both scales (35). Similarly, the KOOS Pain and ADL scales are confounded. Therefore, the correlation of knee-specific pain and function scales was expected to be high. Finally, all knee-specific scales were expected to have relatively lower correlations with the SF-36 mental measures, because knee problems affect mental health less than physical health.

Cross-sectional and longitudinal tests of known groups validity were based on a theoretical foundation and hypotheses specified in advance as to the strength of relationships with external variables that would be expected for a valid measure. Cross-sectional tests compared pre-TKR scale scores for known groups defined by use of an assistive walking device (cane, walker or wheelchair) for any reason, and by the number (0, 1, 2+) of self-reported comorbid (non-arthritis) conditions using a modified Charlson index based on (36). Because conclusions about the validity of a measure should also be based on longitudinal tests (29) change scores (6 month post-TKR minus pre-TKR scale score) were compared for groups rating themselves as better, same or worse (BSW) at 6 months in their capability to do everyday physical activities and their ability to accomplish daily work, due to their surgery. For each BSW rating, patients were classified into four known groups (lot more, more, same, or less capable/able), as in previous analyses (29).

Group comparisons used one-way analysis of variance (ANOVA), with the known group as the independent variable and the pre-TKR scale scores (cross-sectional analyses) or change scores (longitudinal analyses) as the dependent variables (37, 38). Each ANOVA F-statistic indicates how strongly a scale discriminates between groups and thus provides information about that scale's validity. To facilitate comparisons across scales, relative validity (RV) statistics were calculated, based on the ratio of the F-statistic for each scale to the F-statistic for the best performing scale ($RV=1.0$) within each set of scale comparisons; 95% confidence intervals for RV statistics were estimated using empirical bootstrap (39, 40).

In cross-sectional known groups validity tests, substantial validity in discriminating between assistive walking device groups was hypothesized for all knee-specific scales, particularly KOOS ADL and WOMAC Function (12), and for SF-36 scales measuring physical but not mental health (41). All knee-specific scales were expected to discriminate weakly between groups defined by the count of comorbid conditions, while generic scales measuring physical health were hypothesized to discriminate substantially. In longitudinal known groups validity tests, all knee-specific measures were hypothesized be more valid than the SF-36 measures, because the BSW items asked patients to rate their overall change because of their joint surgery. In addition, the KOOS ADL and WOMAC Function scales, which ask about difficulty in performing specific physical activities, were hypothesized to be the most valid for longitudinal tests of capability to do physical activities. The KOOS QOL scale, which includes an item about lifestyle modifications due to knee problems, was hypothesized to be the most valid for longitudinal tests of ability to do daily work.

As a measurement property, responsiveness or the magnitude of change in a scale score is best evaluated in relation to the amount of change expected (42). This anchor-based method of evaluating responsiveness (or longitudinal validity), in which changes in a scale score are interpreted in relation to another measure (43, 44), was evaluated using the BSW items, as described above. In addition, traditional estimates of responsiveness, including the effect size (ES; observed change score divided by the standard deviation of the pre-TKR score) (45) and standardized response mean (SRM; observed change score divided by the standard deviation of the change score) (46), were calculated; both statistics are presented to facilitate comparisons with other studies. Because responsiveness to change over time is constrained if a high percent of respondents score at the floor (lowest possible score) or ceiling (highest

possible score) of a scale, floor and ceiling effects also were evaluated at baseline (pre-TKR) and 6 months.

All analyses were performed using Stata Version 11.2 (StataCorp, Irving, TX). Two-tailed tests were used to determine significant ($p<0.05$) differences.

Results

The mean age of the sample ($N=1,179$) was 66.1 ($SD=9.7$); 57% were age 65 or older and 12% were younger than age 55. Sixty-one percent were female. The majority (89.8%) were white, while 7.6% were black and 2.6% reported another race. The highest level of education was high school graduate or less for 28%, while 39% were college graduates or had post-college graduate education. Six-month post-TKR data was available at the time of analysis for $N=886$, who did not differ notably from the full sample in sociodemographic characteristics or pre-TKR KOOS scores. The primary reason that patients did not have 6-month data was that patients who had a second TKR within six months of their initial surgery did not complete a 6-month survey for the first TKR. By design they completed post-TKR surveys for their contralateral knee. Other patients who did not have a 6-month survey completed a follow-up survey at one year, which satisfied study goals as well as regulatory requirements. Conclusions of analyses that used pre-TKR data did not change when the sample was limited to patients who had 6-month data.

The amount of missing data per item at baseline (pre-TKR) was low, ranging from 0.5–3.0% per item for the KOOS (mean=1.3%) and 0.2–2.1% for the SF-36 (mean=0.9%). Baseline scale scores could be calculated for >99% of patients for all measures except the SF-36 PCS and MCS (98.9%) and KOOS Sport scale (98.6%). Six-month change scores (post-TKR minus pre-TKR) could be calculated for >98% of patients completing the 6-month survey for all measures except the SF-36 PCS and MCS (97.5%) and KOOS Sport scale (96.2%). Internal consistency reliability of all scales exceeded the minimum level of 0.70 recommended for group-level analyses (31) at baseline (Table 1) and was similar at 6 months (data not reported).

To maintain a constant sample size across scale comparisons, cross-sectional validity tests were limited to $N=1,143$ patients who had scores for all KOOS, WOMAC and SF-36 measures at baseline (pre-TKR). Longitudinal analyses were limited to $N=820$ patients for whom 6-month data were available and who had 6-month change scores for all measures.

In support of convergent validity, correlation of the KOOS and WOMAC Pain scales was high ($r=0.94$) and higher than correlations of these knee-specific pain scales with the SF-36 Bodily Pain (BP) scale (Table 1). When the five pain items that the KOOS and WOMAC scales have in common were removed from the KOOS Pain scale, the modified KOOS-WOMAC correlation still was high ($r=0.71$). Correlations of the KOOS ADL (same as WOMAC Function) scale with the KOOS Sport and SF-36 Physical Functioning (PF) scales were similar and moderate ($r=0.55$ – 0.57). In support of their validity in discriminating knee-specific from generic health problems, KOOS Symptoms and QOL scales generally had higher correlations with other knee-specific scales than with generic SF-36 scales. As

previously observed, correlations of the KOOS Pain and ADL scales ($r=0.78$) and WOMAC Pain and Function scales ($r=0.77$) were high, in part because of confounded item content. In contrast, unconfounded SF-36 pain (BP) and function (PF) scales had a correlation of $r=0.53$. Lower correlations between all KOOS scales and SF-36 Mental Health and MCS measures ($r=0.18$ – 0.38) indicated that they were measuring distinct constructs.

As hypothesized for valid measures, scores on all KOOS scales were significantly ($p<0.001$) worse for patients using an assistive walking device (Table 2). However, there are multiple reasons for using an assistive device and the SF-36 Physical Functioning scale ($RV=1.00$) and PCS ($RV=0.97$, 95% $CI=0.78$ – 1.20), which respond to conditions in addition to knee problems, showed greater validity in this test. In tests comparing groups defined by counts (0, 1, 2+) of comorbid (non-arthritis) conditions, the KOOS Symptoms and QOL scales had the best discriminant validity (did not discriminate significantly ($p>0.05$) between comorbid condition groups), while other KOOS and WOMAC scales also discriminated weakly (Table 2). In contrast, the SF-36 General Health (GH) scale ($RV=1.00$) was most valid in ordering groups differing in comorbid condition counts.

Longitudinal evidence of validity and responsiveness includes monotonic increases in mean change scores for all KOOS scales as groups made more favorable evaluations of their change in capabilities for doing everyday physical activities and accomplishing daily work at 6 months (Table 3). The KOOS QOL scale was the most responsive ($RV=1.00$) of all knee-specific and generic measures in both longitudinal validity tests, although RV s for the SF-36 PCS were not significantly different from RV s for KOOS QOL. KOOS and WOMAC function ($RV=0.47$ – 0.57), pain ($RV=0.38$ – 0.58), and symptom/stiffness ($RV=0.29$ – 0.45) measures were significantly less responsive than the KOOS QOL scale. RV s of knee-specific and generic measures for similar constructs (KOOS, WOMAC and SF-36 pain; KOOS, WOMAC and SF-36 function) did not differ significantly.

Although a relatively small percentage (about 10%) of patients rated their status as “worse” 6 months post-surgery, the mean change score for the “worse” group on all knee-specific scales was positive in both longitudinal validity tests, indicating improvement. In contrast, mean change scores for the SF-36 generally remained stable or declined for the “worse” group; the one exception was the SF-36 Bodily Pain (BP) scale, where patients in the “worse” group improved by around 0.3 SD units in both tests.

Six months after TKR, the KOOS QOL scale had the highest effect size, along with the KOOS and WOMAC Pain scales (Table 4). Effect sizes were slightly lower for the knee-specific function scales. Standardized response means were similar for most KOOS and WOMAC pain and function scales but were lower for the KOOS Sport scale. ES and SRM were lower for the SF-36 scales than most knee-specific scales.

Before TKR, floor and ceiling effects were negligible to low ($<15\%$) for most knee-specific scales, although there was a large floor effect for the KOOS Sport scale (28.8%) (Table 4). At 6 months, floor and ceiling effects also were low for most KOOS and WOMAC scales, although ceiling effects approached 10% for the KOOS ADL and WOMAC Function scales. In addition, a higher percentage scored at the ceiling on the WOMAC Pain scale than the

KOOS Pain scale at 6 months post-TKR (21.7% versus 13.7%), and the WOMAC Stiffness scale had a higher ceiling effect than the KOOS Symptoms scale (15.6% versus 3.6%). Among those scoring at the ceiling of the WOMAC Pain scale at 6 months, 13% reported some pain (monthly, weekly or daily) on the KOOS knee pain frequency item.

Discussion

This study evaluated the validity and responsiveness of the KOOS among TKR patients using various methods and criteria and compared the KOOS with other widely-used knee-specific and generic PRO measures. In support of its validity, KOOS scales were related to other measures as hypothesized; a few exceptions concerning hypotheses about their comparative performance are discussed below. In support of its responsiveness as a measure of knee-specific outcomes, KOOS had higher effect sizes and standardized response means at 6 months than generic SF-36 measures. Implications of these findings for PRO measurement in TKR are discussed below.

As in other studies (7), the KOOS QOL scale was highly responsive in terms of traditional responsiveness statistics (ES, SRM). In addition, this scale was strongest in discriminating among groups differing in post-TKR ratings of change in ability to do physical activities and daily work. Knee-specific function scales had been hypothesized to be the most valid in the longitudinal physical activities test, and thus the KOOS QOL scale had a stronger performance than hypothesized. KOOS QOL broadly conceptualizes the impact of knee problems, including their cognitive (*awareness of knee problem*), emotional (*troubled by knee problem*), functional (*modification of life style due to knee problem*) and overall (*general difficulty with knee*) consequences. The KOOS QOL scale currently is not submitted to CMS as part of the CJR model. While other scales are required to distinguish knee pain from knee function, because of its empirical performance and its focus on quality of life, KOOS QOL warrants consideration for inclusion in the CJR model to fully capture joint-specific outcomes.

Many TKR studies administer both a knee-specific and a generic questionnaire, to include measures that are specific to knee outcomes plus measures that allow outcomes to be compared across conditions. The PRO component of the CJR model also includes both joint-specific and generic measures. As in previous studies (47–49), knee-specific measures had higher responsiveness statistics (ES, SRM) than generic measures. However, the best SF-36 measure (PCS) was as valid as the most valid knee-specific scale (KOOS QOL) in relation to patient ratings of overall change in function after TKR. It also is notable that patients who rated their status as “worse” 6 months after TKR improved on average on all knee-specific scales, while mean scores for the “worse” group generally declined or remained stable on generic SF-36 measures. This difference in results for the “worse” group warrants further study to determine if it reflects the impact of comorbid orthopedic and other conditions on generic scores despite knee-specific improvement in the KOOS and WOMAC. Alternatively, patients may rate their overall outcomes as worse if their post-surgical improvement was not as great as they expected. Regardless, these results underscore the value of both knee-specific and generic measures for purposes of fully understanding patient outcomes after TKR.

As would be expected for two scales with five items in common, the KOOS and WOMAC Pain scales were highly correlated and their relative validity was not significantly different in cross-sectional and longitudinal tests. However, the trend in RV statistics was more favorable for the KOOS Pain scale (RV=0.56–0.58) than the WOMAC Pain scale (RV=0.38–0.43) in longitudinal validity tests. In addition, a notable percentage of patients who had the best possible score on the WOMAC Pain scale 6 months post-TKR reported some pain on the KOOS Pain scale; Roos found similar results (9). Collectively, these results support use of the KOOS Pain scale over the WOMAC scale, despite the KOOS having slightly higher respondent burden.

The KOOS Symptoms scale has relatively heterogeneous item content, and includes the two WOMAC Stiffness items plus five additional items. The KOOS Symptoms and WOMAC Stiffness scales only had a moderately high correlation ($r=0.72$). In addition, 6-months post-TKR a higher percentage of patients had the best possible score on the WOMAC Stiffness scale (15.6%) than the KOOS Symptoms scale (3.6%); Roos found similar results (9). The Symptoms scale's relatively low item homogeneity, which is often seen in scales of symptoms that largely vary independently, indicates that it may benefit from separate scoring and interpretation of its stiffness and non-stiffness components along with an overall score. A short stiffness scale also may be preferable when only a brief measure of this key OA symptom is needed; for example, the CJR model only includes two stiffness items rather than all 7 KOOS Symptoms items. However, information about the full profile of specific symptoms may be important in facilitating patient-surgeon discussions of TKR outcomes.

The KOOS Sport/Recreation scale did not discriminate well among known groups in the cross-sectional assistive device validity test, but performed as well as other function measures in longitudinal validity tests. However, the Sport scale had a much higher standard deviation post-TKR than pre-TKR. While the higher post-TKR variation in Sport scores may reflect differences in trajectories of functional recovery, it also may reflect differences in patient lifestyles. Roos for example found that Sport activities were extremely or very important to only about 50% of TKR patients (9). The ADL scale alone did not fully capture the full functional improvement of some TKR patients in this study, however; nearly 10% of patients had the highest possible KOOS ADL score post-TKR. To better capture the total benefit of TKR, additional items about activities which are more difficult than the ADL items but are more applicable to the broader TKR population than the Sport items may need to be developed.

This study has a number of limitations. Data was collected by both paper-pencil and electronic methods; however, self-reported PROs generally have been shown to be equivalent across these two data collection methods (50) and it is unlikely that this impacted results. Criteria used to establish known groups were based on patient self-report; additional analyses using clinician reports to define severity groups and to rate patient change after TKR also should be conducted. In addition, accumulating evidence of validity is an ongoing process. These analyses should be replicated and extended, including evaluation of patients one year or more post-TKR. Similar tests of validity also should be conducted for patients with milder knee OA and other knee disorders and patients from countries other than the U.S. Results of this study may not apply to these other patient populations.

In summary, this study found that the KOOS was reliable, valid and responsive in a large cohort of TKR patients in the U.S. By comparing various knee-specific measures with each other and with generic measures before and after TKR, this study confirmed the complimentary advantages of these measurement approaches. This study also provides information that will be useful in balancing the brevity, precision and interpretation of knee-specific PRO measures for TKR patients, as will be necessary for their routine clinical use.

Acknowledgments

The authors thank Jeroan Allison, MD, MPH, Milena Anatchkova, PhD, Patricia Franklin, MD, MBA, MPH and Courtland Lewis, MD for helpful comments on earlier drafts of this paper; Nina Deng, EdD for developing the bootstrapping software used to evaluate relative validity; and Wenyun Yang, MS and Hua Zheng, PhD for data support.

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

References

1. 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; 28:88–96. [PubMed: 9699158]
2. 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; 8:439–48. [PubMed: 9863983]
3. Skou ST, Roos EM, Laursen MB. A randomized, controlled trial of total knee replacement. *N Engl J Med.* 2016; 374:692.
4. Ayers DC, Li W, Harrold L, Allison J, Franklin PD. Preoperative pain and function profiles reflect consistent TKA patient selection among US surgeons. *Clin Orthop Relat Res.* 2015; 473:76–81. [PubMed: 24957788]
5. Faschingbauer M, Kasperek M, Schadler P, Trubrich A, Urlaub S, Boettner F. Predictive values of WOMAC, KOOS, and SF-12 score for knee arthroplasty: Data from the OAI. *Knee Surg Sports Traumatol Arthrosc.* 2016 Nov 11. [Epub ahead of print].
6. Federal Register. Medicare program; Comprehensive Care for Joint Replacement payment model for acute care hospitals furnishing lower extremity joint replacement services. 2015 80 Federal Register 73273.
7. Collins NJ, Prinsen CA, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee injury and Osteoarthritis Outcome Score (KOOS): Systematic review and meta-analysis of measurement properties. *Osteoarthritis Cartilage.* 2016; 24:1317–29. [PubMed: 27012756]
8. Stratford PW, Kennedy DM. A comparison study of KOOS-PS and KOOS function and sport scores. *Phys Ther.* 2014; 94:1614–21. [PubMed: 25035265]
9. 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. [PubMed: 12801417]
10. de Groot IB, Favejee MM, Reijman M, Verhaar JA, Terwee CB. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: A validation study. *Health Qual Life Outcomes.* 2008; 6:16. [PubMed: 18302729]
11. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM, et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome

- Score (KOOS) in knee osteoarthritis patients. *Osteoarthritis Cartilage*. 2008; 16:423–8. [PubMed: 17905602]
12. Goncalves RS, Cabri J, Pinheiro JP, Ferreira PL. Cross-cultural adaptation and validation of the Portuguese version of the Knee injury and Osteoarthritis Outcome Score (KOOS). *Osteoarthritis Cartilage*. 2009; 17:1156–62. [PubMed: 19303082]
  13. Monticone M, Ferrante S, Salvaderi S, Motta L, Cerri C. Responsiveness and minimal important changes for the Knee Injury and Osteoarthritis Outcome Score in subjects undergoing rehabilitation after total knee arthroplasty. *Am J Phys Med Rehabil*. 2013; 92:864–70. [PubMed: 23900017]
  14. Moutzouri M, Tsoumpas P, Billis E, Papoutsidakis A, Gliatis J. Cross-cultural translation and validation of the Greek version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with total knee replacement. *Disabil Rehabil*. 2015; 37:1477–83. [PubMed: 25335550]
  15. Paradowski PT, Keska R, Witonski D. Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in patients with osteoarthritis undergoing total knee replacement. *BMJ Open*. 2015; 5:e006947.
  16. Xie F, Li SC, Roos EM, Fong KY, Lo NN, Yeo SJ, et al. Cross-cultural adaptation and validation of Singapore English and Chinese versions of the Knee injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. *Osteoarthritis Cartilage*. 2006; 14:1098–103. [PubMed: 16814575]
  17. Nakamura N, Takeuchi R, Sawaguchi T, Ishikawa H, Saito T, Goldhahn S. Cross-cultural adaptation and validation of the Japanese Knee Injury and Osteoarthritis Outcome Score (KOOS). *J Orthop Sci*. 2011; 16:516–23. [PubMed: 21766211]
  18. Engelhart L, Nelson L, Lewis S, Mordin M, Demuro-Mercon C, Uddin S, et al. Validation of the Knee Injury and Osteoarthritis Outcome Score subscales for patients with articular cartilage lesions of the knee. *Am J Sports Med*. 2012; 40:2264–72. [PubMed: 22962288]
  19. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: A multicenter study. *J Rheumatol*. 2014; 41:509–15. [PubMed: 24429183]
  20. Steinhoff AK, Bugbee WD. Knee Injury and Osteoarthritis Outcome Score has higher responsiveness and lower ceiling effect than Knee Society Function Score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. 2016; 24:2627–33. [PubMed: 25416964]
  21. McAlindon TE, Driban JB, Henrotin Y, Hunter DJ, Jiang GL, Skou ST, et al. OARS Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. *Osteoarthritis Cartilage*. 2015; 23:747–60. [PubMed: 25952346]
  22. Alviar MJ, Olver J, Brand C, Tropea J, Hale T, Pirpiris M, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *J Rehabil Med*. 2011; 43:572–83. [PubMed: 21607295]
  23. Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: A patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA*. 2012; 308:1217–8. [PubMed: 23011710]
  24. KOOS scoring. 2012. www.koos.nu. Last accessed June 21, 2016
  25. Bellamy, N. WOMAC osteoarthritis index user guide VIII. Queensland, Australia: University of Queensland; 2007.
  26. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16:297–334.
  27. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992; 30:473–83. [PubMed: 1593914]
  28. Ware, JE., Jr, Kosinski, M., Dewey, JE. How to score Version 2 of the SF-36 Health Survey. Lincoln, RI: QualityMetric Incorporated; 2000.
  29. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: Summary of results from the Medical Outcomes Study. *Med Care*. 1995; 33:AS264–79. [PubMed: 7723455]

30. Veenhof C, Bijlsma JWJ, van den Ende CHM, Van Dijk GM, Pisters MF, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. *Arthritis Care Res.* 2006; 55:480–92.
31. Nunnally, JC., Bernstein, IH. *Psychometric theory*. 3rd. New York: Mc-Graw Hill; 1994.
32. Streiner, DL., Norman, GR., Cairney, J. *Health measurement scales: A practical guide to their development and use*. 5th. Oxford, UK: Oxford University Press; 2015.
33. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res.* 2002; 11:193–205. [PubMed: 12074258]
34. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959; 56:81–105. [PubMed: 13634291]
35. Gandek B. Measurement properties of the Western Ontario and McMaster Universities Osteoarthritis index: A systematic review. *Arthritis Care Res.* 2015; 67:216–29.
36. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care.* 1996; 34:73–84. [PubMed: 8551813]
37. 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; 31:247–63. [PubMed: 8450681]
38. Kerlinger, FN. *Foundations of behavioral research*. New York: Holt, Rinehart, and Winston; 1964.
39. Deng N, Allison JJ, Fang HJ, Ash AS, Ware JE Jr. Using the bootstrap to establish statistical significance for relative validity comparisons among patient-reported outcome measures. *Health Qual Life Outcomes.* 2013; 11:89. [PubMed: 23721463]
40. Henderson AR. The bootstrap: A technique for data-driven statistics using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta.* 2005; 359:1–26.
41. Hawker GA, Melfi CA, Paul JE, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol.* 1995; 22:1193–6. [PubMed: 7674255]
42. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol.* 2010; 10:22. [PubMed: 20298572]
43. Ware, JE., Jr, Keller, SD. Interpreting general health measures. In: Spilker, B., editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd. Philadelphia, PA: Lippincott-Raven Publishers; 1996. p. 445-60.
44. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002; 77:371–83. [PubMed: 11936935]
45. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989; 27:S178–89. [PubMed: 2646488]
46. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990; 28:632–42. [PubMed: 2366602]
47. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology (Oxford).* 1999; 38:870–7. [PubMed: 10515649]
48. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB, Kinemax Outcomes Group. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am.* 2001; 83-A:1856–64. [PubMed: 11741066]
49. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage.* 2007; 15:273–80. [PubMed: 17052924]
50. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value Health.* 2008; 11:322–33. [PubMed: 18380645]

Significance & Innovations

- This study compared the validity and responsiveness of the KOOS in relation to the WOMAC and SF-36 Health Survey in a large cohort of U.S. total knee replacement patients.
- KOOS was a valid and responsive joint-specific measure in this patient population.
- The KOOS Quality of Life scale warrants consideration as a short aggregate knee-specific QOL outcome measure in the CMS Comprehensive Care for Joint Replacement model.
- Joint-specific and generic measures demonstrated complimentary advantages in evaluating the outcomes of total knee replacement.

Table 1

Correlations among knee-specific and SF-36 measures, pre-TKR *

	k	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pain												
(1) KOOS Pain	9	46.4	18.0									
(2) WOMAC Pain	5	51.7	18.9	0.94								
(3) SF-36 BP	2	36.0	18.4	0.66	0.63							
Function												
(4) KOOS/WOMAC ADL	17	52.8	18.3	0.78	0.77	0.64						
(5) KOOS Sport	5	18.4	19.6	0.51	0.45	0.42	0.55					
(6) SF-36 PF	10	38.6	22.1	0.49	0.50	0.53	0.57	0.44				
Other Knee-Specific												
(7) KOOS Symptoms	7	48.6	19.8	0.67	0.57	0.46	0.55	0.43	0.38			
(8) WOMAC Stiffness	2	43.5	22.3	0.65	0.58	0.50	0.61	0.42	0.37	0.72		
(9) KOOS QOL	4	25.4	18.0	0.57	0.54	0.53	0.59	0.54	0.50	0.47	0.45	
Other Generic												
SF-36 MH	5	73.4	18.9	0.33	0.34	0.36	0.34	0.18	0.28	0.24	0.21	0.30
SF-36 PCS	35	33.2	8.4	0.52	0.51	0.69	0.56	0.46	0.82	0.37	0.40	0.50
SF-36 MCS	35	51.7	11.9	0.36	0.36	0.38	0.38	0.18	0.25	0.26	0.22	0.31

* N = 1,143. k = number of items. SE for all correlations=0.029. All measures scored so 0=worst/100=best possible score, except SF-36 PCS/MCS (US general population mean=50, SD=10; lower score=poorer health). Column headings in parentheses match rows; e.g., column (1) is for (1) KOOS Pain. KOOS ADL and WOMAC Function scales have identical content so data for both scales is presented in (4) KOOS/WOMAC ADL. Internal consistency reliability (Cronbach's alpha): KOOS Pain=0.88, WOMAC Pain=0.84, SF-36 BP=0.77; KOOS/WOMAC ADL=0.95, KOOS Sport=0.89; SF-36 PF=0.87; KOOS Symptoms=0.74, WOMAC Stiffness=0.78, KOOS QOL=0.81, SF-36 MH=0.85, SF-36 PCS=0.92, SF-36 MCS=0.92. BP = Bodily Pain; ADL = Function in Activities of Daily Living; PF = Physical Functioning; QOL = Quality of Life; MH = Mental Health; PCS = Physical Component Summary; MCS = Mental Component Summary.

Table 2

Mean scores (SD) and known-groups validity tests for assistive walking device and comorbid condition groups, pre-TKR *

	Assistive Walking Device				Number of Comorbid Conditions				
	No	Yes	F	RV (95% CI)	0	1	2+	F	RV (95% CI)
N	790	352			488	405	250		
KOOS									
Symptoms	50.4 (18.8)	44.7 (21.3)	20.59	0.11 (0.04–0.22)	48.9 (20.1)	48.8 (19.7)	47.8 (19.4)	0.29 [§]	0.01 (0.00–0.02)
Pain	49.3 (17.1)	39.8 (18.1)	73.04	0.38 (0.24–0.59)	47.8 (17.7)	46.6 (17.7)	43.4 (18.5)	5.04 [†]	0.11 (0.02–0.26)
ADL	56.5 (16.9)	44.4 (18.5)	118.17	0.62 (0.44–0.90)	54.7 (17.9)	53.3 (18.5)	48.3 (17.9)	10.57	0.22 (0.08–0.45)
Sport	21.1 (19.6)	12.2 (18.1)	52.93	0.28 (0.15–0.46)	19.5 (19.3)	18.8 (19.6)	15.5 (20.0)	3.64 [‡]	0.08 (0.00–0.22)
QOL	27.6 (17.8)	20.5 (17.5)	39.79	0.21 (0.11–0.34)	26.1 (18.3)	25.9 (18.1)	23.5 (17.0)	1.90 [§]	0.04 (0.00–0.12)
WOMAC									
Stiffness	45.6 (21.8)	38.7 (22.8)	23.70	0.12 (0.05–0.24)	44.8 (22.9)	44.2 (22.0)	39.9 (21.2)	4.33 [‡]	0.09 (0.01–0.24)
Pain	54.9 (17.9)	44.5 (19.3)	77.68	0.41 (0.25–0.61)	53.2 (18.7)	51.9 (18.8)	48.5 (19.2)	5.16 [†]	0.11 (0.02–0.25)
Function	56.5 (16.9)	44.4 (18.5)	118.17	0.62 (0.44–0.90)	54.7 (17.9)	53.3 (18.5)	48.3 (17.9)	10.57	0.22 (0.08–0.45)
SF-36									
PF	44.1 (20.9)	26.0 (19.3)	190.27	1.00 -	40.4 (21.2)	39.9 (23.2)	32.8 (21.1)	11.16	0.23 (0.08–0.42)
RP	49.7 (26.0)	29.4 (24.8)	153.08	0.80 (0.58–1.09)	46.5 (26.7)	43.6 (27.9)	37.4 (26.6)	9.38	0.20 (0.06–0.37)
BP	39.3 (17.7)	28.4 (17.4)	92.05	0.48 (0.32–0.72)	37.9 (18.5)	36.5 (18.7)	31.2 (16.9)	11.56	0.24 (0.09–0.45)
GH	73.8 (16.7)	61.4 (20.5)	116.63	0.61 (0.41–0.94)	75.5 (16.4)	68.4 (18.9)	62.1 (19.8)	47.91	1.00 -
VT	54.9 (19.8)	44.7 (21.2)	62.00	0.33 (0.18–0.52)	54.3 (20.1)	52.2 (20.9)	46.2 (20.9)	12.98	0.27 (0.12–0.48)
SF	72.8 (25.3)	53.7 (28.0)	130.74	0.69 (0.46–1.02)	69.6 (26.7)	68.2 (27.4)	60.0 (28.7)	10.83	0.23 (0.07–0.43)
RE	78.4 (26.1)	63.6 (31.1)	69.57	0.37 (0.21–0.59)	75.8 (27.9)	75.1 (27.8)	68.0 (30.3)	6.91 [†]	0.14 (0.03–0.33)
MH	75.7 (18.2)	68.4 (19.5)	36.94	0.19 (0.09–0.37)	74.8 (18.0)	74.0 (18.8)	70.0 (20.4)	5.72 [†]	0.12 (0.02–0.27)
PCS	35.3 (7.8)	28.5 (7.9)	184.63	0.97 (0.78–1.20)	34.5 (8.1)	33.1 (8.6)	30.6 (8.2)	18.11	0.38 (0.19–0.62)
MCS	53.3 (11.4)	48.1 (12.4)	47.38	0.25 (0.12–0.44)	52.6 (11.4)	52.2 (12.0)	49.2 (12.6)	7.38	0.15 (0.03–0.32)

* F = ANOVA F-statistic; RV = relative validity; CI = confidence interval. All measures scored so 0=worst/100=best possible score, except SF-36 PCS/MCS (US general population mean=50, SD=10). All F-statistics p<0.001 except:

[†] p<0.01,

[‡] p<0.05,

[§]
p>0.05.

ADL = Activities of Daily Living; QOL = Quality of Life; PF = Physical Functioning; RP = Role Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role Emotional; MH = Mental Health; PCS = Physical Component Summary; MCS = Mental Component Summary.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Mean change scores (SD) and known-groups validity tests for self-evaluated physical activity and daily work transition groups *

	Capability in Everyday Physical Activities [†]					Ability to Accomplish Daily Work [‡]						
	Lot more	More	Same	Less	F	RV (95% CI)	Lot more	More	Same	Less	F	RV (95% CI)
N	451	207	77	83			415	216	108	76		
KOOS												
Symptoms	30.6 (21.9)	22.4 (19.4)	14.5 (22.3)	8.9 (22.7)	33.28	0.45 (0.28–0.63)	31.3 (22.3)	23.4 (18.0)	15.1 (22.0)	7.3 (22.6)	38.21	0.43 (0.28–0.62)
Pain	38.4 (19.9)	30.0 (19.0)	23.6 (23.5)	14.0 (19.9)	42.97	0.58 (0.40–0.79)	39.5 (20.3)	30.7 (18.2)	21.4 (20.0)	14.9 (20.4)	49.83	0.56 (0.39–0.75)
ADL	32.6 (17.0)	26.3 (17.8)	20.5 (18.3)	12.4 (19.3)	37.40	0.51 (0.35–0.68)	33.4 (17.6)	26.8 (16.7)	19.3 (16.8)	12.5 (19.4)	42.43	0.47 (0.33–0.64)
Sport	37.4 (25.4)	22.8 (24.5)	15.5 (26.3)	11.1 (24.6)	41.96	0.57 (0.38–0.78)	38.0 (25.9)	24.7 (22.5)	16.0 (27.4)	10.3 (23.8)	43.69	0.49 (0.33–0.68)
QOL	46.1 (22.6)	31.1 (21.6)	23.3 (22.9)	12.0 (22.2)	73.90	1.00 -	47.6 (22.4)	32.7 (21.3)	22.2 (20.6)	10.5 (21.5)	89.39	1.00 -
WOMAC												
Stiffness	32.5 (26.0)	24.3 (25.5)	18.7 (27.4)	9.9 (29.3)	21.42	0.29 (0.16–0.44)	33.9 (26.3)	23.7 (25.1)	18.4 (26.1)	9.5 (28.7)	26.09	0.29 (0.18–0.44)
Pain	35.8 (19.7)	29.1 (18.8)	23.7 (23.2)	14.5 (19.4)	32.01	0.43 (0.28–0.61)	36.6 (20.2)	29.9 (18.0)	21.3 (20.4)	16.5 (20.2)	34.28	0.38 (0.25–0.53)
Function	32.6 (17.0)	26.3 (17.8)	20.5 (18.3)	12.4 (19.3)	37.40	0.51 (0.35–0.68)	33.4 (17.6)	26.8 (16.7)	19.3 (16.8)	12.5 (19.4)	42.43	0.47 (0.33–0.64)
SF-36												
PF	30.7 (22.2)	18.6 (19.5)	10.2 (25.1)	3.4 (22.1)	52.64	0.71 (0.50–0.99)	31.2 (21.8)	19.5 (20.5)	13.8 (23.8)	0.3 (22.4)	55.27	0.62 (0.44–0.89)
RP	32.7 (28.1)	18.7 (24.6)	9.5 (28.4)	1.9 (23.2)	44.96	0.61 (0.42–0.88)	33.2 (28.9)	20.3 (25.1)	12.5 (24.1)	–0.9 (22.8)	46.35	0.52 (0.35–0.74)
BP	30.5 (22.2)	18.9 (19.2)	13.4 (16.7)	7.9 (19.7)	42.18	0.57 (0.40–0.85)	31.8 (22.1)	18.1 (19.4)	15.0 (17.6)	6.8 (17.5)	51.31	0.57 (0.39–0.87)
GH	3.6 (12.7)	–0.5 (15.3)	–0.3 (12.8)	–4.3 (16.4)	10.17	0.14 (0.05–0.25)	3.8 (12.6)	–0.1 (15.2)	0.6 (13.7)	–6.2 (15.8)	12.65	0.14 (0.06–0.26)
VT	12.6 (17.7)	6.3 (15.2)	5.1 (17.8)	–3.2 (20.9)	22.56	0.31 (0.17–0.46)	13.0 (17.7)	6.5 (16.4)	5.8 (16.8)	–5.4 (18.0)	27.72	0.31 (0.20–0.45)
SF	17.6 (24.7)	9.3 (23.7)	5.8 (22.1)	0.8 (28.3)	15.72	0.21 (0.11–0.35)	18.5 (25.5)	8.9 (23.3)	6.3 (21.0)	–0.7 (25.9)	19.67	0.22 (0.12–0.36)
RE	11.1 (26.5)	9.8 (26.2)	3.8 (24.7)	0.2 (25.5)	5.22	0.07 (0.02–0.15)	11.7 (26.8)	9.0 (24.8)	4.0 (25.9)	0.3 (27.2)	5.66	0.06 (0.02–0.14)
MH	8.2 (16.0)	5.8 (14.0)	4.3 (13.7)	–0.3 (17.3)	7.70	0.10 (0.03–0.22)	8.5 (16.1)	5.4 (15.1)	5.1 (12.5)	–1.8 (15.3)	10.40	0.12 (0.05–0.20)
PCS	12.6 (8.7)	6.8 (7.4)	4.1 (8.2)	1.4 (7.9)	65.79	0.89 (0.66–1.22)	12.9 (8.7)	7.3 (7.8)	5.4 (7.5)	0.2 (7.4)	69.59	0.78 (0.56–1.10)
MCS	2.2 (10.2)	1.9 (9.5)	1.2 (7.5)	–1.0 (10.7)	2.61 [§]	0.04 (0.00–0.09)	2.4 (10.4)	1.5 (9.3)	1.0 (8.3)	–1.4 (9.8)	3.45 [§]	0.04 (0.01–0.08)

* F = ANOVA F-statistic; RV = relative validity; CI = confidence interval. All measures scored so 0=worst/100=best possible score, except SF-36 PCS/MCS (US general population mean=50, SD=10). All F-statistics p<0.001 except:

[§] p<0.05,

^{||} p>0.05.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ADL = Activities of Daily Living; QOL = Quality of Life; PF = Physical Functioning; RP = Role Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role Emotional; MH = Mental Health; PCS = Physical Component Summary; MCS = Mental Component Summary.

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

[‡]Item text (response options): Thinking about your daily work at home or in the workplace; Compared to before your joint surgery are you more or less able to accomplish your work now because of your joint surgery? (A lot more able to accomplish now, somewhat more able to accomplish now, about the same, somewhat less able to accomplish now, a lot less able to accomplish now; fourth and fifth response groups combined in ANOVA)

Table 4

Descriptive statistics for knee-specific and SF-36 measures at pre-TKR and 6 months post-TKR *

	Mean Score (SD)			Responsiveness		% Floor [†]		% Ceiling [†]	
	Pre-TKR	Post-TKR	Change	ES	SRM	Pre-TKR	Post-TKR	Pre-TKR	Post-TKR
KOOS									
Symptoms	49.2 (19.8)	74.1 (16.9)	24.9 (22.7)	1.25	1.10	0.7	0.0	0.3	3.6
Pain	47.6 (18.1)	80.1 (17.2)	32.5 (21.5)	1.80	1.51	1.3	0.0	0.5	13.7
ADL	54.0 (18.2)	81.8 (16.5)	27.9 (18.8)	1.53	1.49	0.7	0.0	0.3	9.4
Sport	19.0 (19.4)	48.1 (27.2)	29.0 (27.1)	1.49	1.07	28.8	4.1	0.9	3.9
QOL	26.7 (18.5)	63.4 (22.6)	36.8 (25.2)	1.99	1.46	14.5	0.6	0.1	8.2
WOMAC									
Stiffness	44.0 (22.5)	71.0 (20.4)	26.9 (27.4)	1.20	0.98	6.2	0.7	2.2	15.6
Pain	53.1 (18.9)	84.0 (15.9)	30.9 (20.9)	1.63	1.47	1.3	0.0	0.9	21.7
Function	54.0 (18.2)	81.8 (16.5)	27.9 (18.8)	1.53	1.49	0.7	0.0	0.3	9.4
SF-36									
PF	40.1 (22.1)	63.1 (24.3)	23.0 (23.9)	1.04	0.96	1.7	0.7	0.4	2.1
RP	44.4 (27.4)	68.2 (27.2)	23.8 (28.9)	0.87	0.82	7.2	1.6	4.4	21.3
BP	37.0 (18.2)	60.8 (22.9)	23.7 (22.3)	1.31	1.06	4.7	0.9	0.5	8.0
GH	71.8 (18.1)	73.2 (19.5)	1.4 (14.1)	0.08	0.10	0.0	0.0	3.8	5.7
VT	53.4 (20.7)	62.1 (20.0)	8.7 (18.1)	0.42	0.48	1.0	0.7	1.4	2.2
SF	69.7 (27.0)	82.4 (23.1)	12.7 (25.2)	0.47	0.50	1.9	0.2	26.2	51.6
RE	76.1 (27.7)	85.1 (22.3)	9.0 (26.4)	0.32	0.34	2.4	0.4	39.6	56.3
MH	74.7 (18.6)	81.1 (16.2)	6.4 (15.7)	0.34	0.41	0.0	0.1	5.2	9.6
PCS	33.6 (8.5)	42.8 (9.8)	9.2 (9.2)	1.08	1.00	0.1	0.0	0.0	0.1
MCS	52.8 (11.6)	54.5 (9.7)	1.7 (9.9)	0.15	0.17	0.0	0.1	0.1	0.0

* N = 820. ES = Effect size; SRM = Standardized response mean. All measures scored so 0=worst/100=best possible score, except SF-36 PCS/MCS (US general population mean=50, SD=10). ADL = Activities of Daily Living; QOL = Quality of Life; PF = Physical Functioning; RP = Role Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role Emotional; MH = Mental Health; PCS = Physical Component Summary; MCS = Mental Component Summary.

[†] % Floor=% with worst possible (lowest) score; % Ceiling=% with best possible (highest) score.