Published in final edited form as:
J Arthroplasty. 2011 August ; 26(5): 728–737. doi:10.1016/j.arth.2010.06.004.

Measuring Functional Improvement after Total Knee Arthroplasty Requires both Performance-Based and Patient-Report Assessments: A Longitudinal Analysis of Outcomes

Ryan L. Mizner, PhD, PT[1,2], **Stephanie C. Petterson, PhD, PT**[1], **Katie E. Clements, PT, DPT**[1], **Joseph A Zeni Jr., PhD, PT**[1], **James Irrgang, PhD, PT, ATC**[3], and **Lynn Snyder-Mackler, ScD, PT, ATC**[1]

[1]Department of Physical Therapy, University of Delaware, 301 McKinly Laboratory, Newark, DE, USA 19716.

[2]School of Physical Therapy and Rehabilitation Science, The University of Montana, 32 Campus Drive, Missoula, MT, USA 59812.

[3]Department of Orthopaedic Surgery, University of Pittsburg School of Medicine, Suite 911 Kaufmann Medical Building, 3471 Fifth Avenue, Pittsburgh, PA, USA 15213.

Abstract

The purpose was to explore the responsiveness of both patient-report and performance-based outcome measures to determine functional changes during the acute and long-term postoperative recovery after total knee arthroplasty (TKA). One hundred patients scheduled for unilateral TKA underwent testing preoperatively, 1 month, and 12 months postoperatively using the Delaware Osteoarthritis Profile. All physical performance measures decreased initially after surgery then increased in the long term, however the perceived function did not follow the same trend and some showed an increase immediately after surgery. Patient-report measures were variable with no to small response early, but had excellent long-term responsiveness that was twice as large as performance measures. Patient perception fails to capture the acute functional declines after TKA and may overstate the long-term functional improvement with surgery.

INTRODUCTION

Osteoarthritis is a primary source of disability in the United States and the knee joint is the most common joint to develop osteoarthritis [1,2]. Total knee arthroplasty (TKA) is the most frequently performed joint replacement surgery [3] engendering considerable interest in determining how TKA surgery affects patients' functional ability and impacts the disability resulting from knee osteoarthritis (OA).

Both patient-reported and performance-based measures of physical function are used to evaluate outcomes after TKA. Patient report measures are the most commonly used as they are less expensive, less time intensive and reduce the number of patients lost at follow-up

Address for correspondence: Lynn Snyder-Mackler, P.T., A.T.C., Sc.D. Department of Physical Therapy 301 McKinly Laboratory University of Delaware Newark, DE 19716 smack@udel.edu Ph: (302) 831-3613; Fax: (302) 831-4234.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

because they do not require a clinic visit. The Outcome Measures in Rheumatology conference (OMERACT 3) stated that patient-reported measures of functional outcomes are recommended for all randomized controlled trials of interventions for patients with knee OA, while performance-based measurements of physical function are optional [4].

Patient-report measures of physical function provide useful information related to patients' perceptions of physical function, but there is a burgeoning body of evidence that suggests patient-reports fail to capture the actual change in functional performance after TKA [5-8]. Scores on patient-report can be substantially influenced by patients' pain [6,8-10] as well as their level of exertion during function tasks [9,11]. Improvements in patient-report often correspond strongly with improvements in patient's report of pain [8,9,12]. Patients who have advanced knee OA and subsequent TKA have difficulty discriminating pain from their ability to perform functional tasks.

Few studies have included both pre-operative and post-operative assessments of physical function using both patient-report and performance-based measures of physical function. Even fewer include acute assessment coupled with adequate follow-up after TKA [6,7,13]. Change relative to the preoperative measure is particularly important as patients' preoperative health status has a strong relationship with postoperative outcomes [14-16]. The assessment of physical function beyond the acute recovery phase is key as patients' outcomes normally do not stabilize until at least 6 months after surgery [17,18]. In addition, no study to our knowledge has concurrently measured impairments of body function, such as mobility of the knee and muscle power to determine how these impairments relate to patient-reported and performance-based measures of activity and participation. There are some data that suggest patient-report measures of activity and participation have poor concurrent validity with performance-based measures, but what specific functional tasks most manifest this disconnect are not known.

The purpose of our investigation was to explore the responsiveness of both patient-report and performance-based outcome measures during the acute (one month) and long-term (one year) recovery following TKA for knee OA. We hypothesized that there would be greater improvement in the region-specific KOS-ADLS and generic SF-36 health status measure compared to the level of improvement in the performance-based measures of activity, especially in the early stages after TKA.

Furthermore, we compared the validity of patients' perception of activity limitations compared to their actual activity limitations as determined by an array of performance-based measures of activity. We hypothesized that there would only be moderate relationships between patient-report and the performance-based outcome measures. In addition, we explored how impairments in body function related to the patient's perceptions of activity limitations as well as to performance-based measures of activity limitation. We predicted that performance-based measures of activity would more closely relate to physical impairments of knee range of motion, muscle strength, knee swelling and the patient-report measures would more closely relate to measures of pain.

MATERIALS AND METHODS

Participants

One hundred patients (52 men and 48 women, mean ± standard deviation age: 65 ± 9 years, height: 1.71 ± 0.10m, mass: 89.9 ± 15.1 kg, body mass index (BMI): 30.8 ± 4.5) who were scheduled for unilateral TKA for end-stage knee OA were recruited from an experienced group of local orthopaedic surgeons. All participants underwent a TKA with a tricompartmental, cemented endoprosthesis with a medial parapatellar surgical approach.

Patients were included in the study on a consecutive basis as part of a clinical trial of rehabilitation after TKA (NCT00224913). Potential research participants were excluded from the study if they had musculoskeletal involvement other than unilateral TKA limiting their function or if they were diagnosed with uncontrolled blood pressure, neoplasms, diabetes mellitus, neurological disorders such as multiple sclerosis or Parkinson's disease, or a BMI greater than 40 (morbidly obese). The uninvolved knee was not screened for radiographic arthritic changes, but if patients had an average uninvolved knee pain greater than 4 out of 10 on a verbal analog scale, or if they planned to have surgery on their uninvolved knee within a year (staged bilateral TKA), they were excluded from the study. All participants agreed to participate in the study and signed written informed consent forms approved by the human subjects review board at the University of Delaware prior to enrollment in the study.

Postoperative care

Following the TKA, participants underwent a standardized 3 day inpatient hospital care protocol, followed by home physical therapy for 2 to 3 weeks after discharge from the hospital. One month after surgery, patients began 6 weeks of outpatient rehabilitation 2-3 times per week, at the University of Delaware Physical Therapy Clinic. Outpatient physical therapy included a previously published set of interventions designed to control pain and swelling, improve knee range of motion (ROM), increase lower extremity strength, and training to improve functional ability[19]. Each patient was treated based on their individual impairments according to the guidelines for intervention.

Testing procedures

Participants underwent testing at three different time periods in this prospective cohort investigation: 1) approximately 2 weeks preoperatively; 2) 1 month; and 3) 12 months postoperatively. The order of testing was completion of patient-reported measures of activity followed by assessment of active knee ROM and circumferential knee girth. Participants then completed the timed up and go test (TUG) and stair climbing test (SCT) before undergoing isometric quadriceps strength assessment in the Muscle Performance Laboratory. The six minute walk test (6MW) was the final test. Collectively, these tests are part of the Delaware Osteoarthritis Profile and have been effectively used to measure pre-operative ability and post-operative functional recovery after TKA [20].

Patient-Reported Measures

Short Form-36 Health Questionnaire (SF-36)—The Medical Outcomes Survey Short Form-36 (SF-36) was utilized in this study as it is reliable, internally consistent, and easy to administer [21,22] and has been repeatedly used as a generic health measure in patients with TKA and knee OA [22-24]. This questionnaire includes 8 scales of differing domains of health: physical functioning, bodily pain, role-physical, general health, vitality, role-emotional, social functioning, and mental health. Each scale is scored on a 0 to 100 scale with a 100 representing the best score possible. The SF-36 can also provide a physical component summary (PCS) which represents a composite score for the respective physical scales of the questionnaire. The PCS scores are standardized normative scores based on the general population's score with an average of 50 and a standard deviation of 10 [24]. The SF-36 PCS was the focus of this investigation as it is a composite of all scales to represent the physical aspects of health. The individual domains are reported to reveal which portions of the measure are changed with TKA intervention.

Knee Outcome Survey – Activities of Daily Living Scale (KOS-ADLS)—The KOS-ADLS is a 14 item questionnaire with items designated to assess how patients perceive

commonly described knee symptoms restrict their daily life (pain, stiffness, swelling, giving way, weakness, and limping) as well as their perception of the level of functional limitations during activities of daily living (walk, up stairs, down stairs, standing, kneeling, squatting, sitting, and kneeling) [25]. Potential item responses for the symptoms question are graded from the best score of “I don’t have the symptom” to the worst score of “the symptom prevents me from all activities of daily living”. Items are scored on a 6 point scale from 0 to a maximum of 5 points allotted to each item. Scores are presented as a percentage of the maximal score with 100% representing full perceived knee function for activities of daily living. The KOS-ADLS questionnaire has excellent reliability and good responsiveness to treatment for patients with disorders of the knee [25,26]. The KOS-ADLS score will be the focus of analysis as an outcome measure, but each individual item score will also be provided for comparison with clinical measures of impairments and the performance test array.

Global Rating Score of Knee Function—Research participants were asked to rate their perception of knee functional ability on a scale of 0 to 100%. A score of 0 represented a complete disability and a score of 100 represents a level of knee function prior to the patient having any symptoms.

Performance-Based Measures of Activity

Timed Up and Go Test (TUG)—The TUG measures the time it takes a patient to rise from an armed chair (seat height of 46 cm), walk 3 meters, turn, and return to sitting in the same chair. Patients were instructed to walk as quickly as they felt safe and comfortable. The use of the arms of the chair was permitted to stand up and sit down. A stopwatch was used to measure the time to complete the TUG within the nearest one hundredth of a second. The TUG is widely used to measure mobility in older adults with excellent test-retest reliability (ICC=0.97) [27].

Stair Climbing Test (SCT)—The SCT measure was a physical performance measure that assessed the time it took a research participant to ascend and descend a flight of twelve, 18-cm high steps with a depth of 28 cm. Patients were asked to complete the test as quickly as they felt safe and comfortable and 1 handrail was allowed if required although participants were encouraged to use just their legs for stair negotiation. Time to negotiate the stairs was measured to the nearest one hundredth of a second with a stopwatch. Rejeski et al. [28] found that a similar stair climb task had an excellent test–retest reliability coefficient of 0.93 [28].

For both the TUG and SCT, a practice trial was completed and the mean of two subsequent trials was used for analysis. Assistive devices were allowed only if the patient was unsafe or could not complete the test without a cane or walker. The incidence of use for all performance tests was recorded. The use of handrail during the SCT was also recorded.

Six Minute Walk Test (6MW)—The six minute walk test was a timed test that measured how far patients could walk on a level surface in six minutes. Patients were allowed to use an assistive device if necessary and they were asked to cover as much distance as they could with rests as needed. They were provided with standardized feedback during each test as described previously [29]. The six minute walk test has been favored as a performance measure due to its strong responsiveness to change over time in patients with TKA [7,30,31].

Physical Impairment Measures

Pain—The influence of patients' pain on functional tasks was measured using the bodily pain scale from the SF-36 Health Survey. The score of this scale is derived from 2 items (11-level scale) from the generic questionnaire: 1) asks the patient's intensity of bodily pain or discomfort and 2) measures the extent to which pain interfered with the patient's normal work . Scores are then transformed to zero to 100 point scale where a higher score represents a better outcome. The bodily pain scale is commonly used in studies involving patients with TKA [23], it has potential to provide a wide range of pain scores, and the score represents both the intensity of pain and how pain influences daily activities. The scale has shown adequate test-retest reliability with a coefficient of 0.85 within the general population [24].

Knee Girth—Knee girth was determined by measurement of the transverse plane circumference of the knee at mid-patellar height in a supine position using a flexible plastic measuring tape. Girth measurement have shown acceptable reliability for determination of gross changes in knee swelling for individuals after knee surgery [32].

Knee Range of Motion—Knee range of motion (ROM) was measured using a standard long-arm goniometer. Active knee extension ROM was measured while in supine with the participant's foot propped off the treatment table. Positive values were used to indicate a position of flexion at maximal knee straightening and negative numbers were used to represent positions in knee hyperextension. To determine knee flexion ROM, patients maintained a supine position and were asked to actively slide the heel towards the buttocks and maximal active knee flexion was measured. The examination of knee ROM in patients with knee OA has adequate reliability with a coefficient of 0.96 for flexion and 0.81 for extension [33].

Quadriceps Strength—Quadriceps strength was measured isometrically using a burst superimposition technique that has been previously described in detail [34]. Patients were seated on a dynamometer with the knee flexed to 75 degrees and the position of the dynamometer was recorded at the initial testing of each patient and an identical setup was used for each subsequent strength assessment. Patients performed a warm-up and completed up to three maximal isometric contractions lasting approximately 3 seconds. Approximately 2 seconds into the contraction, an electrical stimulator delivered a supramaximal electrical stimulus to the quadriceps muscle to quantify activation failure. The trial with the highest volitional force production was used for analysis. Maximal isometric force during quadriceps strength testing was normalized to BMI. The determination of quadriceps strength using a burst superimposition technique is highly reliable (ICC = 0.98) [35].

Statistical Methods

Differences in the physical impairments, the scales of the SF-36, the performance-based assessments, and the patient-report questionnaires over time were analyzed using repeated measures analysis of variance (ANOVA). If significance was achieved in the repeated measures assessment, then pairwise comparisons with a Bonferroni correction were performed to assess if there were differences in means between testing sessions. Responsiveness of each outcome measure was assessed by determining the effect size of the differences in means (effect size = difference in means between initial measurement and the subsequent measurement / standard deviation of the initial measurement). A negative effect size represents worsening over time and positive effect sizes represent improvement. Changes in the individual items of the KOS-ADLS score over time were assessed using a Friedman Repeated Measures test. If significance was reached, post hoc analysis using a Wilcoxon Signed Ranks Test was performed to compare changes in an item over time using a Bonferroni approach to correct for the multiple comparisons. Differences in the frequency

of use of assistive devices during each performance test were evaluated using the McNemar test for dependent proportions. Pearson product correlation coefficients were calculated to determine the associations between the patient-reported and performance-based outcome measures as well as to determine the relationships between the functional outcome measures with the physical impairment measurements. An alpha level of 0.05 was chosen for determination of significance.

RESULTS

All patient-reported and performance-based measures of activity (Table 1) and impairment measures (Table 2) changed over time ($p<0.05$). From the pre-operative test to the one month test, the SF-36 PCS, all performance-based measures of activity and all impairment measurements worsened ($p<0.05$). GRS did not significantly change ($p>0.05$), while the KOS-ADLS significantly improved from the preoperative measure ($p<0.05$). The order of responsiveness from largest to smallest effect size of change (either improvement or worsening) was SCT, 6MW, TUG, KOS-ADLS, SF-36 PCS, and GRS.

At 12 months, all of the patient-reported and performance-based measures of activity and all impairment measures improved compared to the one month test ($p<0.05$). The rank order of responsiveness in the long-term change in outcome measures from the preoperative to 12 month test from greatest to least change was KOS-ADLS, GRS, SF-36 PCS, SCT, TUG, and 6MW. The frequency of use of an assistive device significantly increased from the preoperative to 1 month assessment during the TUG and 6MW performance tests and decreased from the 1 month to 12 month test for the entire performance-based test array ($p<0.05$) (Table 3).

There were substantial differences in which of the individual KOS-ADLS items changed over time after surgery. Patient-reported functional limitations due to pain, giving way, and limping were significantly less one month after surgery compared to preoperatively ($p<0.05$) (Figure 1). Perceived limitations due to knee stiffness and limitations from muscle weakness were not significantly changed from preoperatively to 1 months ($p=0.36$ and 0.077 respectively), while limitations due to swelling was the only symptom that worsened during this acute time period ($p<0.05$). Improvement occurred in the entire set of symptom item scores from 1 month test to the 12 month test ($p<0.05$).

The patient's perception of sitting was the only activity of daily living item of the KOS-ADLS that did not change from the preoperative to 1 month assessment ($p=0.714$) (Figure 2). The remaining tasks (walking, up and down stairs, sitting, and rising from a chair) all improved from the preoperative test to 1 month, except the limitations in squatting and kneeling which worsened over this time frame ($p<0.05$). All the items improved from the 1 month assessment to the 12 month assessment ($p<0.05$). The largest perceived limitations in ADL tasks at the long-term assessment was in kneeling and squatting and residual perceived limitations in both of these tasks were substantial.

The separate scales of the SF-36 had a varied response to TKA. There was no change in the general health scale made across both the acute and long-term assessments (both p-values equal close to 1.0). The bodily pain and mental health items had no significant change from preoperatively to 1 month postoperatively ($p=0.315$ and 0.055, respectively), but both were significantly improved by the 12 month tests ($p<0.05$). The remaining items of physical functioning, role-physical, social functioning, role-emotional, and vitality all initially worsened, but subsequently significantly improved by the year mark ($p<0.05$ for all tests).

Significant relationships existed between the physical impairments and the patient-reported and performance-based measures of activity (Table 4). Pain, as measured by the SF-36

bodily pain scale, related only weakly to the TUG and 6MW preoperatively (r = −0.25 and 0.20, $p<0.05$) and weakly to the TUG and SCT at 12 months (r=−0.20 and −0.29, $p<0.05$). In contrast, the pain measure had the strongest relationship of all the impairments at every test period with the patient-reported KOS-ADLS score. In addition, the KOS-ADLS had a weak relationship with involved quadriceps strength at the preoperative and 12 month assessments, knee flexion ROM at the preoperative and 1 month test, and knee extension at the 1 month assessment ($p<0.05$). Quadriceps strength of the involved and uninvolved limbs had the strongest relation to times on the SCT and TUG as well as distance on the 6MW at all time points ($p<0.05$). Differences in knee girth between the limbs had no significant relationship with any outcome measure at any time ($p>0.05$).

There were statistically significant ($p<0.05$), but small correlations between the two patient-reported outcome measures and the GRS at each time point, except with the GRS and SF-36 PCS at the 1 month test. The KOS-ADLS was more strongly related to the GRS for patient perceived physical function (r = 0.43, 0.38 and 0.48 pre-operatively, 1 month and 1 year after TKA, respectively) compared to the SF-36 PCS (r = 0.34, 0.16, 0.41 pre-operatively, 1 month and 1 year after TKA, respectively). The greatest difference in strength of correlation with GRS and the two patient-reported outcome measures occurred during the acute 1 month measurement when the correlation between the GRS and the KOS-ADLS was more than twice as strong as the relationship of the GRS with the SF-36 PCS score.

DISCUSSION

Patients' performance-based activity limitations and physical impairments worsened early after surgery with a subsequent substantial improvement at one year. The patient-report measures of activity limitations did not reflect the acute worsening of performance-based based activity and physical impairments. The performance-based measures had greater responsiveness during the acute stages after surgery than the patient patient-report questionnaires. The ranking of responsiveness reversed order from the acute stage to the long term measurement with a larger effect observed in patient-reported outcome measures compared to the performance-based assessments. Thus, there is poor concurrent validity between patient-reported and performance-based measures of activity limitations after TKA. Solely using patients' perceptions of change tend to overestimate the actual short- and long-term changes in physical function after TKA. Evaluations that include performance-based and patient perception measures, such as the Delaware Osteoarthritis Profile, provide a more comprehensive perspective of recovery after TKA.

Acute changes in patient-reported outcome after TKA suggest that patients dramatically overestimate their functional ability early after surgery. The change in physical impairment was dramatic with muscle weakness (50% loss from preoperative measures), knee swelling (1cm girth increase), loss of motion (25% reduction), and an increased use of assistive devices during performance-based tests from before to one month after surgery. All performance-based measures of activity limitations also underwent a considerable decline by one month. Scores on the patient-reported outcome measures, however, failed to show worsening, in one case actually showed improvement, and patients tended to overestimate their outcome (Figure 3). This is in accord with the previous findings of Parent and Moffet, who reported significant improvements in self-reported functional ability two months after TKA, while objective functional measures of gait speed, stair climbing time and six-minute walking distance tended to be worse or no better than pre-operative values [36]. There was some drop in the SF-36 PCS measure acutely, but the effect size of the change was relatively small.

We expected to see more worsening and a larger effect size in the change for the region specific KOS-ADLS questionnaire than the SF-36 PCS between the preoperative and one month time point as it was designed to capture signs and symptoms as well as activity limitations that are relevant to the knee joint. Unexpectedly, the KOS-ADLS actually showed a slight improvement at the one month test when compared to preoperative scores. Nevertheless, the KOS-ADLS had twice the strength of correlation than the SF-36 PCS with the GRS at the one month test. The GRS did not significantly change from the preoperative measure. So while an improvement in the KOS-ADLS may be unexpected, the measure is still an appropriate tool to capture the important outcome of patients' perceived functional ability.

In contrast to the discord found in the two patient-report questionnaires, all of the performance-based measures of activity limitations showed a worsening in physical function with the absolute values of the effect sizes much larger than patient-report. Despite having the greatest variance, the SCT test was most responsive performance-based measure in short term recovery. The large changes in the stair climbing test may be a product of the substantial physical demands required for the test. As the tasks demands of the activity become more physically demanding there is a larger proportion of the TKA patient population who have difficulty with that activity [37]. The SCT was followed closely by the 6MW which is consistent with the work of Parent and Moffet on short-term (preoperative to 2 month) responsiveness who found the 6MW to have superior responsiveness to the TUG [38]. Even the rather low demand task of the TUG was more responsive from the preoperative condition compared to patient-report. The TUG also has a history of reaching a ceiling effect after TKA with improvements reaching a plateau in improvement earlier than the SCT and 6MW [39].

The current study offers an opportunity to highlight which tasks and impairments are especially disparate between patient-reported and performance-based measures of activity limitations during the short term recovery via the KOS-ADLS (Figure 4). In some cases, the differences in patients' report of limitations and their actual clinical changes are considerable. While patients perceive that their ability to negotiate stairs is better from the preoperative to the one month tests on the KOS-ADLS, their times on the SCT deteriorated nearly 30% during that interval. Patients also reported less limitation in their ability to walk and rise from a chair compared to their preoperative assessment, yet their times on the TUG slowed 12% and the distance traveled during the 6MW was 13% less. The only activity tasks on the KOS-ADLS that worsened from the preoperative score was patients' perceptions of limitations in the two most physically demanding tasks on the scale, namely kneeling and squatting. Kneeling and squatting are very prevalent daily tasks and the vast majority of patients rank these tasks as some of the most difficult to perform after surgery [40,41].

Discrepancies in patients' perceptions of short-term change versus clinical measurements of change are not isolated to activity limitation tasks. There was a substantial disparity in the change in patients' perception of limitations from muscle weakness and the measured change in muscle weakness impairment. Patients perceive that their limitations due to muscle weakness subside slightly in the acute stage of recovery despite losing half their quadriceps strength. Quadriceps weakness has been shown repeatedly to be one of the prime impairments before and after TKA that has been significantly related to functional performance and has been related to limb avoidance loading patterns acutely after TKA [17,29,42]. Patients also do not perceive any changes in the limitations in function due to increased stiffness at one month despite losing 20 degrees of average knee flexion and about a degree of knee extension from before to one month after surgery.

Patients did have congruence between the perception of swelling and a clinical measure of knee swelling as the category of swelling was the only impairment measure in the KOS-ADLS that showed worsening in the acute stage. The measure of limitations due to pain was improved at 1 month from preoperative scores even though many patients were still taking narcotic pain relievers one month after surgery. One could speculate that the acute improvement may reflect that patients' pain at one month is a different type of pain from their preoperative pain. The surgical pain also tends to be transient throughout the day and not as constant in nature as advanced arthritic pain. Finally, the relief from the arthritis ache may overshadow the postoperative surgical discomfort.

Not only are performance-based activity measures more responsive acutely, they also tend to have stronger relationships with impairments other than pain. Stratford and colleagues have previously reported that pain was the principle determinant of WOMAC scores in patients with TKA and change in pain substantially influences patients' perceptions of change in functional ability [9]. The data from the current study supports this premise and expands on their work in that the performance-based measures more closely matched the patients' impairments of quadriceps muscle weakness and restricted knee ROM than patient-reported outcome measures. While the KOS-ADLS includes only one question that inquires about patients' activity limitations related to pain, preoperative and 1 month scores were still most strongly related to patients' pain. The relationship between pain and KOS-ADLS score weakened substantially by the long-term assessment and was very nearly the same strength as the correlation between involved quadriceps strength and KOS-ADLS score. These relationships may be a reflection of the resolution of most of the patients' knee pain by the 12 month test coupled with a no change in involved quadriceps strength from the preoperative condition.

The study has some remaining limitations in its design that merit attention when interpreting the results. The patients in the study had excellent outcomes in comparison to previous findings [23,31,43] and there is a potential that the responsiveness of the outcome measures are slightly overstated compared to most populations. The discrepancy is especially true of the performance measures. The weakness measure in the KOS-ADLS does not specifically ask about the quadriceps although this muscle is especially impacted by surgery. Similarly, the stiffness measure in the KOS-ADLS may not be interpreted by patients as limitations in range of motion, but rather the sensation of tightness or difficulty to move the joint. There are differences in the level of effort described between the patient-report and performance-based measures of activity. The KOS-ADLS measures activity limitations experienced during usually daily activity and the performance assessments are asking patients to perform at their fastest possible performance which may not be considered part of usual daily activity. Fear of pain or re-injury, expectations of the outcome of surgery, satisfaction with the care provided, education, BMI and smoking status may also influence how an individual completes a patient-reported outcome measure [44].

In summary, there is poor concurrent validity between patient-reported and performance-based measures of physical functioning in patients who undergo unilateral TKA. Performance-based tests are necessary to fully characterize the change in physical function of patients after TKA as they provide objective information of how the patients actually function that are not captured by patient-reported measures. These conclusions could have important implications in clinical decision making. If the need for intervention is based solely on the patient's perception of activity limitation, then important limitations in function of the knee and performance of the individual may go unrecognized and untreated, limiting the individual's potential for achieving optimal outcomes. The Delaware Osteoarthritis Profile or similar comprehensive examinations that include subjective and

objective measures should become the standard for evaluating function and recovery after surgery.

Acknowledgments

Funding for the project was provided by the National Institutes of Health (NIH) grant, NIH R01-HD041055. Additional funding was provided to support the investigators by an NIH training grant (T32-HD07490) and the American Physical Therapy Association's Foundation for Physical Therapy Grants (Mary McMillian and PODS 1 Scholarships). We wish to recognize the expertise and support provided by the many physical therapists working in the University of Delaware Physical Therapy Clinic as well as Drs. Michael Axe, Alex Bodenstab, Leo Raisis, and William Newcomb at First State Orthopaedics. Finally, we acknowledge the efforts and support of the volunteers who participated in this research.

REFERENCES

1. Felson DT, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med. 2000; 133:8.
2. Felson DT, et al. Risk factors for incident radiographic knee osteoarthritis in the elderly: the Framingham Study. Arthritis Rheum. 1997; 40:4.
3. Orthopaedic Patients and Conditions. [cited 2005]. 2005 Available from: http://www.aaos.org/wordhtml/research/stats/patientstats.htm.
4. Bellamy NKJ, Boers M. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol. 1997; 24:4.
5. Jacobs CA, Christensen CP. Correlations between knee society function scores and functional force measures. Clin Orthop Relat Res. 2009; 467:9.
6. Stratford PW, Kennedy DM, Hanna SE. Condition-specific Western Ontario McMaster Osteoarthritis Index was not superior to region-specific Lower Extremity Functional Scale at detecting change. J Clin Epidemiol. 2004; 57:10.
7. Ouellet D, Moffet H. Locomotor deficits before and two months after knee arthroplasty. Arthritis Rheum. 2002; 47:5. [PubMed: 11932871]
8. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. Phys Ther. 2006; 86:11.
9. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. J Clin Epidemiol. 2006; 59:2.
10. Maly MR, Costigan PA, Olney SJ. Determinants of self-report outcome measures in people with knee osteoarthritis. Arch Phys Med Rehabil. 2006; 87:1. [PubMed: 16401430]
11. Stratford PW, et al. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. Arthritis Rheum. 2003; 49:4.
12. Konig A, et al. Balance sheets of knee and functional scores 5 years after total knee arthroplasty for osteoarthritis: a source for patient information. J Arthroplasty. 2000; 15:3.
13. Kennedy DM, et al. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. BMC Musculoskelet Disord. 2005; 6
14. Jones CA, Voaklander DC, Suarez-Alma ME. Determinants of function after total knee arthroplasty. Phys Ther. 2003; 83:8. [PubMed: 12495408]
15. Mizner RL, et al. Preoperative quadriceps strength predicts functional ability one year after total knee arthroplasty. J Rheumatol. 2005; 32:8.
16. Fortin PR, et al. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. Arthritis Rheum. 1999; 42:8.
17. Mizner RL, Petterson SC, Snyder-Mackler L. Quadriceps Strength and the Time Course of Functional Recovery after Total Knee Arthroplasty. J Orthop Sports Phys Ther. 2005; 35:7.
18. Fortin PRPJ, Clarke AE. Timing of total joint replacement affects clinical outcomes among patients with osteoarthritis of the hip or knee. Arthritis Rheum. 2002; 46:12.

19. Stevens JE, Mizner RL, Snyder-Mackler L. Neuromuscular electrical stimulation for quadriceps muscle strengthening after bilateral total knee arthroplasty: a case series. J Orthop Sports Phys Ther. 2004; 34:1. [PubMed: 15179987]
20. Petterson SC, et al. Improved function from progressive strengthening interventions after total knee arthroplasty: a randomized clinical trial with an imbedded prospective cohort. Arthritis Rheum. 2009; 61:2.
21. Ware JE Jr. et al. 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(Suppl):4.
22. Brazier JE, et al. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology (Oxford). 1999; 38:9.
23. Ethgen O, et al. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. J Bone Joint Surg Am. 2004; 86-A:5.
24. Ware, J.; Kosinski, M.; Gandek, B. SF-36 Health Survey: Manual and Interpretation Guide. QualityMetric Incorporated; Lincoln, RI: 2002. 1993
25. Irrgang JJ, et al. Development of a patient-reported measure of function of the knee. J Bone Joint Surg Am. 1998; 80:8.
26. Marx RG, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. J Bone Joint Surg Am. 2001; 83-A:10.
27. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. Phys Ther. 2002; 82:2.
28. Rejeski WJ, et al. Assessing performance-related disability in patients with knee osteoarthritis. Osteoarthritis Cartilage. 1995; 3:3. [PubMed: 8581752]
29. Mizner RL, Snyder-Mackler L. Altered loading during walking and sit-to-stand is affected by quadriceps weakness after total knee arthroplasty. J Orthop Res. 2005; 23:5.
30. Kennedy DM, et al. Assessing recovery and establishing prognosis following total knee arthroplasty. Phys Ther. 2008; 88:1.
31. Moffet H, et al. Effectiveness of intensive rehabilitation on functional ability and quality of life after first total knee arthroplasty: A single-blind randomized controlled trial. Arch Phys Med Rehabil. 2004; 85:4.
32. Soderberg GL, Ballantyne BT, Kestel LL. Reliability of lower extremity girth measurements after anterior cruciate ligament reconstruction. Physiother Res Int. 1996; 1:1. [PubMed: 9238718]
33. Cibere J, et al. Reliability of the knee examination in osteoarthritis: effect of standardization. Arthritis Rheum. 2004; 50:2.
34. Stevens JE, Mizner RL, Snyder-Mackler L. Quadriceps strength and volitional activation before and after total knee arthroplasty for osteoarthritis. J Orthop Res. 2003; 21:5.
35. Snyder-Mackler L, et al. Reflex inhibition of the quadriceps femoris muscle after injury or reconstruction of the anterior cruciate ligament. J Bone Joint Surg Am. 1994; 76:4.
36. Parent E, Moffet H. Comparative responsiveness of locomotor tests and questionnaires used to follow early recovery after total knee arthroplasty. Arch Phys Med Rehabil. 2002; 83:1. [PubMed: 11782824]
37. Noble PC, et al. Does total knee replacement restore normal knee function? Clin Orthop Relat Res. 2005; 431
38. Parent EMH. Comparitive Responsiveness of Locomotor Tests and Questionnaires Used to Follow Early Recovery After Total Knee Arthroplasty. Arch Phys Med Rehabil. 2002; 83:1. [PubMed: 11782824]
39. Kennedy DM, et al. Modeling early recovery of physical function following hip and knee arthroplasty. BMC Musculoskelet Disord. 2006; 7
40. Weiss JM, et al. What functional activities are important to patients with knee replacements? Clin Orthop. 2002; 404
41. Schai PA, Gibbon AJ, Scott RD. Kneeling ability after total knee arthroplasty. Perception and reality. Clin Orthop. 1999; 367

NIH-PA Author Manuscript
NIH-PA Author Manuscript

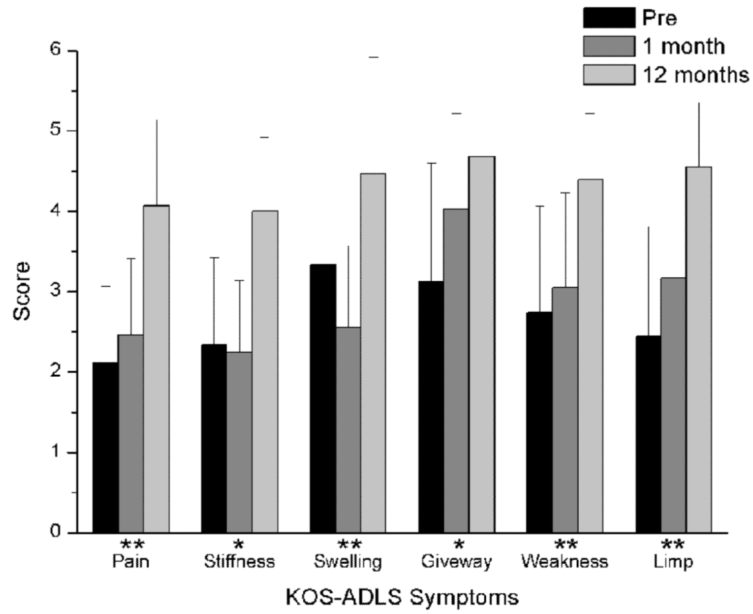
42. Petterson SC, et al. Progressive strengthening interventions improve function after total knee arthroplasty: a randomized clinical trial with imbedded prospective cohort. Arthritis Care Res. 2009 In Press.

43. Walsh M, et al. Physical impairments and functional limitations: a comparison of individuals 1 year after total knee arthroplasty with control subjects. Phys Ther. 1998; 78:3.

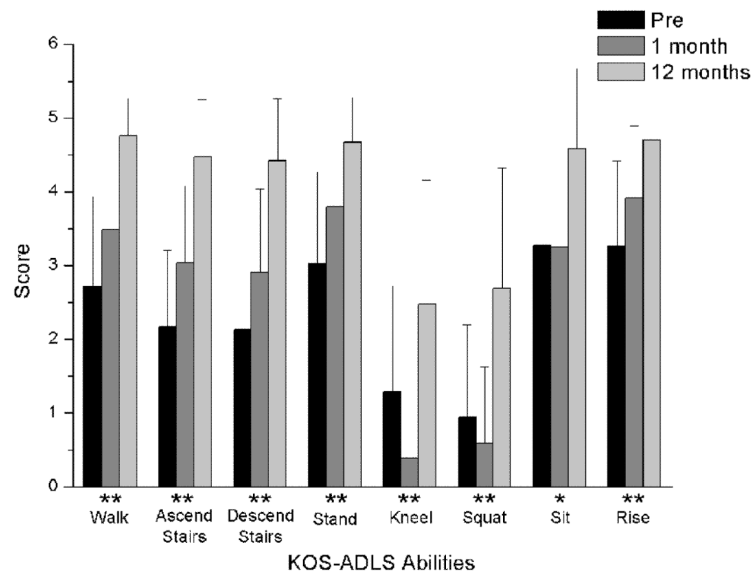
44. Kowalchuk DA, et al. Prediction of patient-reported outcome after single-bundle anterior cruciate ligament reconstruction. Arthroscopy. 2009; 25:5.

NIH-PA Author Manuscript

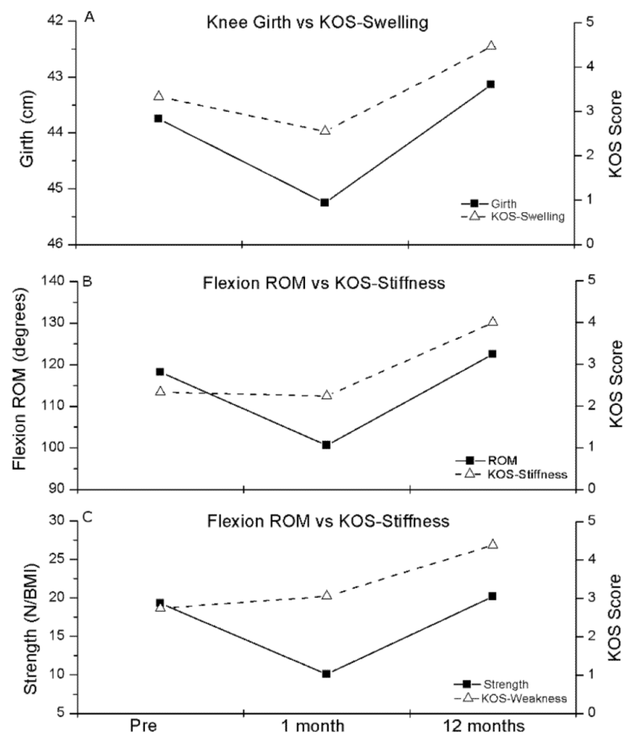
NIH-PA Author Manuscript



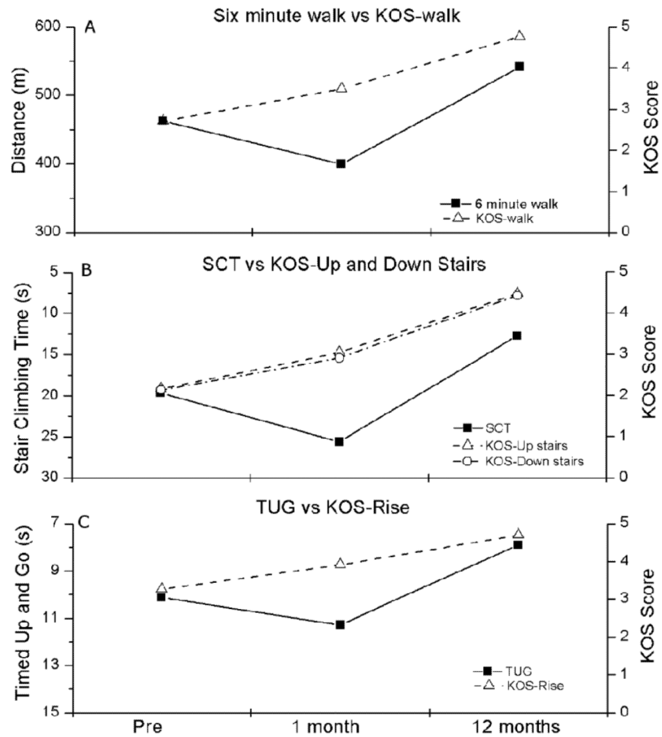
1. .


Pre
1 month
12 months
Score
0
1
2
3
4
5
6
**
Walk
**
Ascend
Stairs
**
Descend
Stairs
**
Stand
**
Kneel
**
Squat
*
Sit
**
Rise
KOS-ADLS Abilities

2. .


A
Knee Girth vs KOS-Swelling
Girth (cm)
KOS Score
Girth
KOS-Swelling
B
Flexion ROM vs KOS-Stiffness
Flexion ROM (degrees)
KOS Score
ROM
KOS-Stiffness
C
Flexion ROM vs KOS-Stiffness
Strength (N/BMI)
KOS Score
Strength
KOS-Weakness
Pre
1 month
12 months

3. .


Six minute walk vs KOS-walk
A
Distance (m)
KOS Score
6 minute walk
KOS-walk
SCT vs KOS-Up and Down Stairs
B
Stair Climbing Time (s)
SCT
KOS-Up stairs
KOS-Down stairs
TUG vs KOS-Rise
C
Timed Up and Go (s)
TUG
KOS-Rise
Pre
1 month
12 months

4. .

Table 1

Functional Outcome Measures Over Time. Means, SD, and Effect Size.

	Preop	1mo	Pre-1mo Effect Size	12mo	Pre-12mo Effect Size	1mo-12mo Effect Size
SF-36 PCS (General Adult Population = 50 ± 10)***	31.8 ± 8.2	29.2 ± 6.7	**−0.28**	45.6 ± 8.4	**1.72**	**2.45**
KOS% (0-100) ***	49.9 ± 15.6	55.3 ± 12.4	**0.35**	84.5 ± 10.4	**2.22**	**2.35**
GRS% (0-100) **	55.3 ± 20.5	53.8 ± 18.3	**−0.07**	90.9 ± 8.7	**1.74**	**2.03**
TUG (sec) ***	10.1 ± 2.8	11.3 ± 2.9	**−0.43**	7.9 ± 1.8	**0.79**	**1.17**
SCT (sec) ***	19.7 ± 8.3	25.6 ± 10.2	**−0.71**	12.7 ± 4.5	**0.84**	**1.26**
6MW (m) ***	460.4 ± 123.3	399.7 ± 104.1	**−0.49**	541.3 ± 118.8	**0.66**	**1.36**

Negative effect sizes represent worsening of score and positive effect sizes represent improvements in score.

*** measures that are statistically significantly changes between each time point (corrected p<0.05)

** measures that are statistically significantly different only between the preop to 12 mo and 1mo tests to 12 mo tests (corrected p<0.05) sec = seconds, m = meters, SF-36 PCS = Short Form-36 physical component summary, KOS = Knee Outcome Survey - Activities of Daily Living Scale, GRS = Global Rating Scale, TUG = Timed Up and Go, SCT = Stair Climb Test, 6MW = Six Minute Walk Test.

Table 2

Physical Impairment Measures Over Time. Means & SD.

	Preop	1mo	12mo
NMVIC Uninvolved Quadriceps *	23.2 ± 8.8	22.5 ± 8.0	21.3 ± 8.8
NMVIC Involved Quadriceps ††	19.3 ± 7.5	10.1 ± 3.9	20.2 ± 8.2
Knee Flexion ROM ***	118.2 ± 15.0	100.6 ± 14.4	122.5 ± 9.2
Knee Extension ROM *	4.5 ± 5.3	5.6 ± 4.1	0.55 ± 2.7
Girth at Mid-Patella Uninvolved (cm)	42.5 ± 4.0		
Girth at Mid-Patella Involved (cm) ***	43.7 ± 4.2	45.3 ± 4.4	43.1 ± 4.0

*** measures that are statistically significantly changes between each time point (correct p<0.05)

†† measures that are statistically significantly changes between preop to 1mo and 1mo to 12 mo (corrected p<0.05)

* measures that are statistically significantly changes only between preop to 12 mo (corrected p<0.05) NMVIC = Normalized maximal voluntary isometric contraction, ROM = range of motion, cm = centimeters

Table 3

Count of Patients Who Required Use of an Assistive Devices for Functional Performance Measures

	Preop	1 mo	12mo
TUG †	4	12	0
SCT †	53	61	50
6MW †	5	12	1

† measures that are significantly different only between 1mo and 12 mo tests $p<0.05$ TUG = Timed Up and Go. SCT = Stair Climb Test. 6MW = Six Minute Walk Test.

Table 4a

Preoperative Test. Pearson Correlation Coefficients for impairments related to functional outcome measures.

	Knee Outcome Survey - Activities of Daily Living Scale	Timed Up & Go Test	Stair Climbing Test	Six Minute Walk Test
Bodily Pain (SF-36 subscale 0-100)	**0.57** *	**−0.25** *	−0.17	**0.20** *
Knee Swelling (girth difference in cm between limbs)	−0.06	−0.02	0.05	0.00
Involved Knee Extension ROM (deg)	−0.06	0.07	0.12	0.01
Involved Knee Flexion ROM (deg)	**0.28** *	**−0.22** *	**−0.25** *	0.05
Normalized Uninvolved Quadriceps Strength (N/BMI)	−0.02	**−0.31** *	**−0.34** *	**0.33** *
Normalized Involved Quadriceps Strength (N/BMI)	**0.28** *	**−0.45** *	**−0.46** *	**0.44** *

Table 4b

One Month Test. Pearson Correlation Coefficients for impairments related to functional outcome measures.

	Knee Outcome Survey - Activities of Daily Living Scale	Timed Up & Go Test	Stair Climbing Test	Six Minute Walk Test
Bodily Pain (SF-36 subscale 0-100)	**0.53** *	−0.05	−0.03	−0.07
Knee Swelling (girth difference in cm between limbs)	−0.03	−0.12	−0.13	0.1
Involved Knee Extension ROM (deg)	**−0.30** *	0.09	0.01	0.00
Involved Knee Flexion ROM (deg)	**0.27** *	−0.19	−0.18	0.17
Normalized Uninvolved Quadriceps Strength (N/BMI)	−0.16	**−0.40** *	**−0.48** *	**0.49** *
Normalized Involved Quadriceps Strength (N/BMI)	0.09	**−0.40** *	**−0.36** *	**0.44** *

Table 4c

Twelve Month Test. Pearson Correlation Coefficients for impairments related to functional outcome measures.

	Knee Outcome Survey - Activities of Daily Living Scale	**Timed Up & Go Test**	**Stair Climbing Test**	**Six Minute Walk Test**
Bodily Pain (SF-36 subscale 0-100)	**0.27** *	**−0.20** *	**−0.29** *	0.10
Knee Swelling (girth difference in cm between limbs)	−0.1	−0.15	−0.08	0.12
Involved Knee Extension ROM (deg)	0.01	**0.33** *	**0.20** *	**−0.21** *
Involved Knee Flexion ROM (deg)	0.02	**−0.21** *	−0.03	0.07
Normalized Uninvolved Quadriceps Strength (N/BMI)	0.19	**−0.45** *	**−0.51** *	**0.55** *
Normalized Involved Quadriceps Strength (N/BMI)	**0.26** *	**−0.48** *	**−0.45** *	**0.58** *

N = newtons, BMI = body mass index, deg = degrees

*correlations are statistically significant ($p<0.05$)