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,*† Stephanie C. Petterson, PhD, PT,* Katie E. Clements, PT, DPT,* Joseph A. Zeni Jr., PhD, PT,* James J. Irrgang, PhD, PT, ATC,‡ and Lynn Snyder-Mackler, ScD, PT, ATC*

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 and at 1 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. Keywords: knee arthroplasty, outcome measures, self-reported function, performance measures, disability.

© 2010 Elsevier Inc. All rights reserved.

Osteoarthritis (OA) is a primary source of disability in the United States, and the knee joint is the most common joint to develop OA [1,2]. Total knee arthroplasty (TKA) is the most frequently performed joint arthroplasty [3], engendering considerable interest in determining how TKA surgery affects patients’ functional ability and impacts the disability resulting from knee 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 because they are less expensive, are less time intensive, and reduce the number of patients lost at follow-up because they do not require a clinic visit. The Outcome Measures in Rheumatology conference stated that patient-reported measures of functional outcomes are recommended for all randomized controlled trials of interventions for patients with knee OA, whereas 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 that 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 preoperative and postoperative 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 that 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 (1 month) and long-term (1 year) recovery following TKA for knee OA. We hypothesized that there would be greater improvement in the region-specific Knee Outcome Survey–Activities of Daily Living Scale (KOS-ADLS) and generic Short Form–36 (SF-36) health status measure compared with 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 with 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 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 (ROM), muscle strength, and knee swelling and that 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 ± 0.10 m; 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 orthopedic 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; if they were diagnosed with uncontrolled blood pressure, neoplasms, diabetes mellitus, or neurologic disorders such as multiple sclerosis or Parkinson disease; or if they had 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 before 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 to 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 ROM, and increase lower extremity strength, and training to improve functional ability [19]. Each patient was treated based on his or her individual impairments according to the guidelines for intervention.

**Testing Procedures**

Participants underwent testing at 3 different periods in this prospective cohort investigation: (1) approximately 2 weeks preoperatively, (2) 1 month postoperatively, 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 6-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 preoperative ability and postoperative functional recovery after TKA [20].

**6-minute walk test (6MW)**

The 6-minute walk test (6MW) is a measure of aerobic endurance, typically performed on a flat, level surface. Participants are instructed to walk as far as possible in 6 minutes, and the distance is recorded. It is commonly used to assess functional exercise capacity in various populations, including those with knee OA undergoing TKA. This test evaluates a patient’s ability to perform daily activities that involve sustained walking, such as shopping, climbing stairs, or commuting. It is a simple, quick, and reliable measure that can be performed in any setting with a minimal requirement for equipment.
**Patient-Reported Measures**

**Short Form–36 Health Questionnaire**

The Medical Outcomes Survey SF-36 was used in this study because 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) that 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 because 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**

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 describing functional activities are graded from the best score of “I have no difficulty performing the task” to the worst score of “I am unable to perform the task.” Items are scored on a 6-point scale from 0 to 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 complete disability, and a score of 100 represents a level of knee function before the patient had any symptoms.

**Performance-Based Measures of Activity**

**Timed Up and Go Test**

The TUG measures the time it takes a patient to rise from an armed chair (seat height of 46 cm), walk 3 m, 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 (Intraclass Correlation Coefficient [ICC] = 0.97) [27].

**Stair Climbing Test**

The SCT was designed to measure mobility in older adults with excellent test-retest reliability (ICC = 0.93) [28]. This questionnaire measured time to complete the TUG within the nearest one hundredth of a second. The TUG was used to measure mobility in older adults with excellent test-retest reliability (Intraclass Correlation Coefficient [ICC] = 0.97) [27].

**Six-Minute Walk Test**

The 6MW was a timed test that measured how far patients could walk on a level surface in 6 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 6MW has been favored as a performance measure because of 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 a 0- 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 midpatellar height in a supine position using a flexible plastic measuring tape. Girth measurement has shown acceptable reliability for determination of gross changes in knee swelling for individuals after knee surgery [32].

Knee ROM
Knee ROM was measured using a standard long-arm goniometer. Active knee extension ROM was measured while in supine position 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 toward the buttocks; and the 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°, and the position of the dynamometer was recorded at the initial testing of each patient; an identical setup was used for each subsequent strength assessment. Patients performed a warm-up and completed up to 3 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. 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 rank 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 measures. An α level of .05 was chosen for determination of significance.

Table 1. Functional Outcome Measures Over Time (Means, SD, and Effect Size)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Preop 1 mo</th>
<th>Pre–1 mo Effect Size</th>
<th>12 mo</th>
<th>Pre–12 mo Effect Size</th>
<th>1 mo–12 mo Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36 PCS</td>
<td>31.8 ± 8.2</td>
<td>29.2 ± 6.7</td>
<td>-0.28</td>
<td>45.6 ± 8.4</td>
<td>1.72</td>
</tr>
<tr>
<td>KOS % (0-100)</td>
<td>49.9 ± 15.6</td>
<td>55.3 ± 12.4</td>
<td>0.35</td>
<td>84.5 ± 10.4</td>
<td>0.22</td>
</tr>
<tr>
<td>GRS % (0-100)</td>
<td>55.3 ± 20.5</td>
<td>53.8 ± 18.3</td>
<td>-0.07</td>
<td>90.9 ± 8.7</td>
<td>0.74</td>
</tr>
<tr>
<td>TUG (s)</td>
<td>10.1 ± 2.8</td>
<td>11.3 ± 2.9</td>
<td>-0.43</td>
<td>7.9 ± 1.8</td>
<td>0.79</td>
</tr>
<tr>
<td>SCT (s)</td>
<td>19.7 ± 8.3</td>
<td>25.6 ± 10.2</td>
<td>-0.71</td>
<td>12.7 ± 4.5</td>
<td>0.84</td>
</tr>
<tr>
<td>6MW (m)</td>
<td>460.4 ± 123.3</td>
<td>399.7 ± 104.1</td>
<td>-0.49</td>
<td>541.3 ± 118.8</td>
<td>0.66</td>
</tr>
</tbody>
</table>

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

*Measures that are statistically significantly changed between each time point (corrected P < .05).
†Measures that are statistically significantly different only between the preoperative and 12-month tests and between 1-month and 12-month tests (corrected P < .05).
Results

All patient-reported and performance-based measures of activity (Table 1) and impairment measures (Table 2) changed over time ($P < .05$). From the preoperative test to the 1-month test, the SF-36 PCS, all performance-based measures of activity, and all impairment measurements worsened ($P < .05$). The Global Rating Scale (GRS) did not significantly change ($P > .05$), whereas the KOS-ADLS significantly improved from the preoperative measure ($P < .05$). 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 with the 1-month test ($P < .05$). 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 the 12-month test for the entire performance-based test array ($P < .05$) (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 1 month after surgery compared with preoperatively.
Knee Bodily pain (SF-36 subscale 0-100) 0.27 *
Knee swelling (girth difference in cm between limbs) −0.06
Involved knee extension ROM (°) −0.06
Involved knee flexion ROM (°) 0.28 *
Normalized uninvolved quadriceps strength (N/BMI) 0.27 *
Normalized involved quadriceps strength (N/BMI) −0.02
Normalized involved quadriceps strength (N/BMI) 0.28 *

1-mo test
Knee swelling (girth difference in cm between limbs) 0.53 *
Involved knee extension ROM (°) −0.03
Involved knee flexion ROM (°) 0.27 *
Normalized uninvolved quadriceps strength (N/BMI) −0.16
Normalized involved quadriceps strength (N/BMI) 0.09

12-mo test
Knee swelling (girth difference in cm between limbs) −0.1
Involved knee extension ROM (°) 0.01
Involved knee flexion ROM (°) 0.02
Normalized uninvolved quadriceps strength (N/BMI) 0.19
Normalized involved quadriceps strength (N/BMI) 0.26 *

N indicates newtons.
*Correlations are statistically significant (P < .05).

Table 4. Pearson Correlation Coefficients for Impairments Related to Functional Outcome Measures

<table>
<thead>
<tr>
<th>Preoperative test</th>
<th>KOS-ADLS</th>
<th>TUG</th>
<th>SCT</th>
<th>6MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodily pain (SF-36 subscale 0-100)</td>
<td>0.57 *</td>
<td>−0.25 *</td>
<td>−0.17</td>
<td>0.20</td>
</tr>
<tr>
<td>Knee swelling (girth difference in cm between limbs)</td>
<td>−0.06</td>
<td>−0.02</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>Involved knee extension ROM (°)</td>
<td>−0.06</td>
<td>0.07</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Involved knee flexion ROM (°)</td>
<td>0.28 *</td>
<td>−0.22 *</td>
<td>−0.25 *</td>
<td>0.05</td>
</tr>
<tr>
<td>Normalized uninvolved quadriceps strength (N/BMI)</td>
<td>−0.02</td>
<td>−0.31 *</td>
<td>−0.34 *</td>
<td>0.33 *</td>
</tr>
<tr>
<td>Normalized involved quadriceps strength (N/BMI)</td>
<td>0.28 *</td>
<td>−0.45 *</td>
<td>−0.46 *</td>
<td>0.44 *</td>
</tr>
</tbody>
</table>

1-mo 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 (°) | −0.30 * | 0.09 | 0.01 | 0.00 |
| Involved knee flexion ROM (°) | 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 * |

12-mo 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 (°) | 0.01 | 0.33 * | 0.20 * | −0.21 * |
| Involved knee flexion ROM (°) | 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 * |

(P < .05) (Fig. 1). Perceived limitations due to knee stiffness and limitations from muscle weakness were not significantly changed from preoperation to 1 month postoperation (P = .36 and .077, respectively), whereas limitation due to swelling was the only symptom that worsened during this acute period (P < .05). Improvement occurred in the entire set of symptom item scores from the 1-month test to the 12-month test (P < .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 = .714) (Fig. 2). The remaining tasks (walking, up and down stairs, and rising from a chair) all improved from the preoperative test to the 1-month assessment, except the limitations in squatting and kneeling that worsened over this time frame (P < .05). All the items improved from the 1-month assessment to the 12-month assessment (P < .05). The largest perceived limitations in ADL tasks at the long-term assessment were 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 preoperatively to 1 month postoperatively (P = .315 and .055, respectively), but both were significantly improved by the 12-month tests (P < .05). The remaining items of physical functioning, role-physical, social functioning, role-emotional, and vitality all initially worsened, but subsequently significantly improved by the 1-year mark (P < .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 < .05) and weakly to the TUG and SCT at 12 months (r = −0.20 and −0.29, P < .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 < .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 < .05). Differences in knee girth between the limbs had no significant relationship with any outcome measure at any time (P > .05).

There were statistically significant (P < .05) but small correlations between the 2 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 preoperatively, 1 month after TKA, and 1 year after TKA, respectively) compared with the SF-36 PCS (r = 0.34, 0.16, and 0.41 preoperatively, 1 month after TKA, and 1 year after TKA, respectively). The greatest difference in strength of correlation with GRS and the 2 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 1 year. The patient-report measures of activity limitations did not reflect the acute worsening of performance-based activity and physical impairments. The performance-based measures had greater responsiveness during the acute stages after surgery than the 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 with 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 tends to overestimate the actual short-term 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 (1-cm girth increase), loss of motion (25% reduction), and an increased use of assistive devices during performance-based tests from before to 1 month after surgery. All performance-based measures of activity limitations also underwent a considerable decline by 1 month. Scores on the patient-reported outcome measures, however, failed to show worsening and in 1 case actually showed improvement; and patients tended to overestimate their outcome (Fig. 3). This is in accord with the previous findings of Parent and Moffet [36], who reported significant improvements in self-reported functional ability 2 months after TKA, whereas objective functional measures of gait speed, stair climbing time, and 6-minute walking distance tended to be worse or no better than preoperative values. 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 1-month time point because 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 1-month test when compared with preoperative scores. Nevertheless, the KOS-ADLS had twice the strength of correlation than the SF-36 PCS with the GRS at the 1-month test. The GRS did not significantly change from the preoperative measure. So although 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 2 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 those in the patient report. Despite having the greatest variance, the SCT test was the most responsive performance-based measure in short-term recovery. The large changes in the SCT 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 has 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 that found the 6MW to have superior responsiveness to the TUG [36]. Even the rather low demand task of the TUG was more responsive from the preoperative condition compared with 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 [38].

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 (Fig. 4). In some cases, the differences in patients' report of limitations and their actual clinical changes are considerable. Although patients perceive that their ability to negotiate stairs is better from the preoperative to the 1-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 with 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 2 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 [39,40].

Discrepancies in patients' perceptions of short-term change vs 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 [17,20,29]. Patients also do not perceive any changes in the limitations in function due to increased stiffness at 1 month despite losing 20° of average knee flexion and about a degree of knee extension from before to 1 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 1 month after surgery. One could speculate that the acute improvement may reflect that patients' pain at 1 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 Kennedy [9] have previously reported that pain was the principle determinant of Western Ontario and McMaster Universities Arthritis Index (WOMAC) scores in patients with TKA and change in pain substantially influences patients' perceptions of change in functional ability. The data from the current study support this premise and expand 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. Although the KOS-ADLS includes

---

**Fig. 4.** 6MW, SCT, and TUG scores compared to patients' perceptions of functional skills of walking, ascending and descending stairs, and rising from a chair, respectively. (A) Left vertical axis is 6MW distance in meters (m). Right vertical axis is the KOS score. (B) Left vertical axis is SCT time in seconds (s). Right vertical axis is the KOS score. (C) Left vertical axis is the TUG score in seconds (s). Right vertical axis is the KOS score. KOS, Knee Outcome Survey - Activities of Daily Living Scale; 6MW, Six Minute Walk; SCT, Stair Climb Test; TUG, Timed Up & Go.
only 1 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 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 [23,31,41], and there is a potential that the responsiveness of the outcome measures are slightly overstated compared with 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 ROM, 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 ask patients to perform at their fastest possible performance, which may not be considered part of usual daily activity. Fear of pain or reinjury, 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 [42].

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
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


