 Discriminative Ability of Functional Measures in Knee Osteoarthritis Patients Classified Based on Radiographic Severity

N. Pirayeh¹, M. J. Shaterzadeh Yazdi¹, H. Negahban²,³, M. Mehravar¹, N. Mostafae²

¹ Musculoskeletal Rehabilitation Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
² Department of Physical Therapy, School of Paramedical Sciences, Mashhad University of Medical Sciences, Mashhad, Iran
³ Orthopedic Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

INTRODUCTION

Knee osteoarthritis (OA) is one of the most common causes of functional limitations and disability in older adults. It has been reported that knee pain and progressive disability in performing daily tasks are two primary concerns in patients with knee OA (1, 2). OA severity are often defined and graded based on degenerative changes seen in radiographic findings. The Kellgren and Lawrence (KL) classification of OA is a widely used radiographic classification system, wherein standard anterior-posterior radiographs are defined, ranging from mild (grade I) to severe (grade IV) radiographic OA (3, 4). Accordingly, knee OA patients with a KL grade ≤ 2 are considered as mild, and patients with a KL grade ≥ 3 as moderate-to-severe (5).

Assessment of functional limitations in patients with knee OA is frequently done using patient-reported outcome measures (PROMs) and performance-based ones in knee OA patients classified based on radiographic severity. The discriminative ability of these measures was determined by calculation of sensitivity, specificity, area under the Receiver Operating Characteristic curve, likelihood ratios, and predictive values.

Background. Determining the discriminative ability of functional outcome measures among patients of knee osteoarthritis (OA) with different severity degrees of radiographic signs can be valuable for clinicians and researchers to classify the functional limitations in these patients before starting their rehabilitation.

Objective. To determine the discriminative ability of patient-reported outcome measures (PROMs) and performance-based ones in knee OA patients classified based on radiographic severity.

Methods. Based-on the Kellgren-Lawrence (KL) grading scale, 130 knee OA patients were classified into 65 patients with mild (a KL grade ≤ 2) and 65 patients with moderate-to-severe (a KL grade ≥ 3) radiographic sign. PROMs and performance-based outcome measures in knee OA patients were assessed by Knee Injury and Osteoarthritis Outcome Score (KOOS) questionnaire as well as Timed-up and go test (TUG), Functional Reach Test (FRT), and step test, respectively. The discriminative ability of these measures was determined by calculation of sensitivity, specificity, area under the Receiver Operating Characteristic curve, likelihood ratios, and predictive values.

Results. Our results showed that all subscales of the KOOS except for the sport/recreation and all performance-based outcome measures have good ability to discriminate between the two groups of knee OA. Also, the ADL subscale of KOOS and step test have good ability in accurate identification of mild grade of knee OA patients. The symptoms subscale of KOOS and TUG test have the best ability in correctly identifying moderate-to-severe grade of knee OA.

Conclusions. Our findings provide evidence for the good discriminative ability of functional measures in patients of knee OA classified based on radiographic severity.

KEY WORDS
Knee osteoarthritis; KOOS; performance-based outcome measures; sensitivity, specificity.
The aim of this study was to determine the discriminative ability of functional outcome measures in differentiating knee OA patients with various degrees of severity. Although some studies have reported that the presence and severity of knee OA is associated with functional limitation (14-21), the ability of these functional outcome measures in differentiating knee OA patients with various degrees of radiographic severity has not been investigated. Thus, the aim of this study was to determine the discriminative ability of these patient-reported outcome measures and performance-based ones in knee OA patients classified based on radiographic severity. To analyze the discriminative ability of these measures, sensitivity and specificity, positive and negative likelihood ratios (PLR and NLR), positive and negative predictive value (PPV and NPV), area under the receiver curve (ROC), and best cutoff score for discriminating between knee OA patients with different severity were calculated (12, 22-24).

METHODS

Study population
A total of 130 patients were diagnosed with knee OA by an orthopedist, based on the classification criteria of the American College of Rheumatology (25). These patients were recruited from outpatient physiotherapy and orthopedic clinics in Ahvaz, Iran. The patients included both male and female Persian native speakers aged between 40-70 years with knee pain and radiological signs of knee OA (unilateral or bilateral) who were able to do daily activity independently. Patients diagnosed with secondary knee OA, rheumatoid arthritis or any concurrent systemic inflammatory disease, hip or knee surgery history, total knee replacement, or hip disorder, were excluded (27, 28). Antero-posterior view of knee radiograph taken in position standing with knee 7-10° flexion (29), was considered as reference standard in this study. The radiological severity of knee OA was determined using KL grading on standard anterior-posterior radiographs by an experienced radiologist who was blinded for test results (3). Based on the KL score (range 1-4), severity of knee OA was categorized into two groups: 65 patients with mild (a KL grade ≤ 2) and 65 patients with moderate-to-severe (a KL grade ≥ 3) radiographic signs. The Ethics Committee at Ahvaz Jundishapur University of Medical Sciences approved the study.

Procedure
Anthropometric information of subjects including age, sex, weight, height, body mass index (BMI) was collected. In this study, measures of functional limitations in knee OA patients which were considered as index tests, were assessed by the KOOS questionnaire as a representative of PROM and TUG, FRT and step test as performance-based outcome measures. In bilateral involvement of knee OA, the side with more severe radiographic sign as the target knee was evaluated.
KOOS questionnaire
The KOOS is a 42-items specific patients-reported outcome measure that assesses five domains: Pain, Symptoms, Function in Activities of Daily Living, Function in Sport and Recreation, and knee-related quality of life. All items are scored by a five-point Likert scale from zero (no problems) to four (extreme problems). The scores of each subscale are calculated separately and transformed to a 0-100 scale, where zero indicates severe knee problem and 100 demonstrates no knee problem (30). The KOOS has good reliability and validity in Persian patients of knee OA (28).

Time up and go test
The functional mobility was determined by TUG test. Based on the standard manner, after the verbal command to begin, patients were asked to stand up from a 46-cm- high chair, walk forward 3 m in a straight line at their habitual walking speed, turn 180°, walk back to the chair and sit down again. Time needed for performing these tasks was recorded in seconds using a chronometer. Prior to its commencement, the test was explained to the patients. This test was repeated 3 times and their average was calculated (27).

Functional reach test
FRT is a single-item test developed as a quick screening test for identifying balance problems in older adults. A yard stick was attached to wall at about shoulder height. Patients stood adjacent to, but not touching, the wall with feet placed shoulder width apart and positioned the arm that is closer to the wall at 90 degrees of shoulder flexion with a closed fist. At this time, an initial reading at the 3rd metacarpal head on the yard stick can be taken. Patients were instructed to reach forward along the yardstick as far as possible without stepping or losing their balance. The practitioner talked a reading on the yardstick of the farthest reach attained by the patient without taking a step. The 3rd metacarpal phalangeal joint was used as marker. The initial reading was subtracted from the final to obtain the functional reach score in cm. The average of three trials was noted (9, 31).

Step test
Step test is evaluated dynamic standing balance which has known reliability and validity. Patients were instructed to stand bare feet on the involved leg and maintain balance on the same leg, while stepping the contralateral leg on a 15-cm step and return it to the floor as quickly as possible in 15 seconds without any hand support. For patients with bilateral knee OA, the knee with more severe radiographic signs was deemed as the involved knee. The test was performed only once. Before starting the test, the patients performed two to three practice steps for familiarization with the test. In cases of balance loss during testing, the number of completed steps until this point was recorded and the test was terminated (11, 32).

Statistical analysis
All data were analyzed using SPSS version 21.0. The level of statistical significance was set at p < 0.05. Independent t-test and chi-square test were used for comparison of demographic characteristics between the two groups with differing severity of radiographic signs. Kolmogorov-Smirnov test was used to assess the normality of data distribution. Based on the results of this test, independent t-test was used for comparison of mean values of functional measures between the groups of mild and moderate-to-severe. The ability of each functional measure in distinguishing between the two groups with differing severity of knee OA was examined by determination of sensitivity, specificity, area under the ROC curve (AUC), PLR, NLR, PPV, NPV, and their 95% confidence intervals (22-24). Receiver operating characteristic (ROC) curves were constructed for the analysis of sensitivity and specificity (23). The sensitivity of a test describes its ability in correctly identifying subjects with a condition of interest, whereas specificity is concerned with the ability of a test to recognize the absence of a condition of interest (22, 24).

In this method, the discriminative ability of outcome measures to distinguish between groups was shown using Area under the curve (AUC) for ROC. ROC curve plots sensitivity (true-positive rate) versus 1-specificity (false-positive rate) of the entire range of values of functional measures, which is used to determine AUC and the best cutoff score for discriminating between the two groups (23). Values of AUC range from 0 to 1, where 1 indicates excellent discriminative ability of a measure between two groups and 0 indicates failure of a measure in discrimination between groups. A traditional academic point scale was used to classify the accuracy of the AUC: 0.9 to 1 indicating excellent; 0.8 to 0.89 good; 0.70 to 0.79 acceptable; 0.60 to 0.69 poor; and 0.00 to 0.59 indicating failure. The score with the combination of highest sensitivity and lowest 1-specificity was determined as the best cutoff point which was located in the most “northwest” point on the ROC curve (12).

To determine the clinical meaningfulness of tests and the predictive ability of measures in correctly identifying subjects with a condition of interest, PLR and NLR, PPV and NPV, and their 95% CI were calculated at the best discrimination of highest sensitivity and lowest specificity, where PLR > 10 indicates excellent predictive ability of a measure in correctly identifying subjects with a condition of interest, whereas NLR < 0.1 indicates very good; 10 to 1.0 acceptable; 1.0 to 5.0 poor; and >5.0 indicating failure. A traditional academic point scale was used to classify the accuracy of the PLR: 0.9 to 1 indicating excellent; 0.8 to 0.89 good; 0.70 to 0.79 acceptable; 0.60 to 0.69 poor; and 0.00 to 0.59 indicating failure. The score with the combination of highest sensitivity and lowest specificity was determined as the best cutoff point which was located in the most “northwest” point on the ROC curve (12).
RESULTS
Demographic information of subjects in the two groups of knee OA is provided in table I. The two groups were matched with each other in this regard. Table II reports mean ± SD and P values for each functional measure. The moderate-to-severe group had greater scores than the mild group for all subscales of KOOS and TUG test, but had less values for FRT and step test. As shown, all functional measures were statistically different between the two groups (p < 0.05).

The sensitivity, specificity, AUC values, asymptotic significance, best cutoff score, LRs and PVs and their 95% confidence intervals for each measure are represented in table II. The results of ROC curve analysis showed that, except for sport/recreation subscale, all KOOS subscales had good ability for discrimination between the two groups of knee OA and also three functional tests had good to excellent accuracy in discrimination between them.

Also, the results of other indicators related to the discriminative ability of these measures showed that among the KOOS subscales, the highest specificity (0.92 (95% CI: 0.82-0.97)) and the greatest PLR (7.8 (95% CI: 3.28-18.5)) and PPV (0.86 (95%CI: 0.74-0.95)) were related to the symptoms subscale while the highest sensitivity (0.89 (95% CI: 0.78-0.95)) and the lowest amount of NLR (0.17 (95% CI: 0.08-0.35)) and NPV (0.91 (95% CI: 0.78-0.97)) were related to ADL subscale. In addition, among the functional tests, TUG test had the highest specificity (0.90 (95% CI: 0.80-0.96)) and the greatest PLR (8.66 (95% CI: 4.18-17.5)) and PPV (0.89 (95% CI: 0.78-0.95)), and step test had the highest sensitivity (0.93 (95% CI: 0.84-0.98)) and the lowest amount of NLR (0.09 (95% CI: 0.03-0.24)) and NPV (0.91 (95% CI: 0.78-0.97)).

DISCUSSION
Our results showed that although all subscales of the KOOS except sport/recreation subscale, have good accuracy in discrimination between the two groups of knee OA, the symptoms subscale show the highest specificity and the greatest PLR and PPV, thus it can be useful for confirming the diagnosis of moderate-to-severe grade of knee OA patients. In addition, ADL subscale has the highest sensitivity and the least NLR and NPV and as a result has the best ability for screening for ruling-out moderate-to-severe grade of knee OA patients.

The results also demonstrated that all the performance-based outcome measures have good ability to discriminate between

### Table I. Demographic information of the study population.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mild group (n = 65) Mean (SD)</th>
<th>Moderate to Severe group (n = 65) Mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>56.2 (6.8)</td>
<td>58.3 (7.1)</td>
<td>0.08</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>160.6 (8.2)</td>
<td>159.2 (7.3)</td>
<td>0.27</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>80.3 (12.7)</td>
<td>80.3 (11.1)</td>
<td>1.00</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>31.2 (5.1)</td>
<td>31.8 (4.4)</td>
<td>0.46</td>
</tr>
<tr>
<td>Sex</td>
<td>17 (M), 48 (F)</td>
<td>13 (M), 52 (F)</td>
<td>0.40</td>
</tr>
</tbody>
</table>

M: male; F: female.
Table III. Sensitivity, specificity, area under the curve, P values, cutoff scores, LRs, PVs and 95% CIs for included measures.

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Area under the curve (95% CI)</th>
<th>P value</th>
<th>Cutoff score</th>
<th>Positive LR (95% CI)</th>
<th>Negative LR (95% CI)</th>
<th>Positive PV (95% CI)</th>
<th>Negative PV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOOS pain</td>
<td>0.81 (0.69-0.89)</td>
<td>0.67 (0.54-0.78)</td>
<td>0.80 (0.72-0.88)</td>
<td>0.03</td>
<td>54.5</td>
<td>2.52 (1.74-3.63)</td>
<td>0.27 (0.16-0.46)</td>
<td>0.71 (0.59-0.81)</td>
<td>0.78 (0.65-0.87)</td>
</tr>
<tr>
<td>KOOS symptoms</td>
<td>0.60 (0.47-0.71)</td>
<td>0.92 (0.82-0.97)</td>
<td>0.83 (0.76-0.90)</td>
<td>0.03</td>
<td>59</td>
<td>7.8 (3.28-18.5)</td>
<td>0.43 (0.32-0.58)</td>
<td>0.88 (0.74-0.95)</td>
<td>0.69 (0.58-0.78)</td>
</tr>
<tr>
<td>KOOS ADL</td>
<td>0.89 (0.78-0.95)</td>
<td>0.61 (0.48-0.73)</td>
<td>0.82 (0.79-0.89)</td>
<td>0.08</td>
<td>61.5</td>
<td>2.32 (1.68-3.19)</td>
<td>0.17 (0.08-0.35)</td>
<td>0.69 (0.58-0.79)</td>
<td>0.85</td>
</tr>
<tr>
<td>KOOS sport/rec</td>
<td>0.88 (0.76-0.93)</td>
<td>0.44 (0.32-0.57)</td>
<td>0.66 (0.57-0.76)</td>
<td>0.04</td>
<td>27.5</td>
<td>1.61 (1.27-2.03)</td>
<td>0.24 (0.11-0.50)</td>
<td>0.61 (0.51-0.71)</td>
<td>0.80</td>
</tr>
<tr>
<td>KOOS QOL</td>
<td>0.69 (0.56-0.79)</td>
<td>0.76 (0.64-0.86)</td>
<td>0.80 (0.73-0.88)</td>
<td>0.03</td>
<td>22</td>
<td>3 (1.78-4.81)</td>
<td>0.4 (0.27-0.58)</td>
<td>0.75 (0.61-0.84)</td>
<td>0.71</td>
</tr>
<tr>
<td>TUG test</td>
<td>0.80 (0.67-0.88)</td>
<td>0.90 (0.80-0.96)</td>
<td>0.90 (0.85-0.95)</td>
<td>0.00</td>
<td>11.16</td>
<td>8.66 (4-18.75)</td>
<td>0.22 (0.13-0.35)</td>
<td>0.89 (0.78-0.95)</td>
<td>0.81</td>
</tr>
<tr>
<td>FRT test</td>
<td>0.64 (0.51-0.75)</td>
<td>0.78 (0.66-0.87)</td>
<td>0.80 (0.73-0.87)</td>
<td>0.03</td>
<td>29.5</td>
<td>3 (1.82-4.93)</td>
<td>0.45 (0.32-0.63)</td>
<td>0.75 (0.61-0.85)</td>
<td>0.68</td>
</tr>
<tr>
<td>Step test</td>
<td>0.93 (0.84-0.98)</td>
<td>0.66 (0.53-0.77)</td>
<td>0.89 (0.84-0.94)</td>
<td>0.02</td>
<td>12.5</td>
<td>2.77 (1.96-3.91)</td>
<td>0.09 (0.03-0.24)</td>
<td>0.73 (0.62-0.82)</td>
<td>0.91</td>
</tr>
</tbody>
</table>

CI: confidence interval.
identification of knee OA with moderate-to-severe signs. This result may be justifiable through items within this subscale. Since effusion, morning stiffness, crepitation and limitation of motion are part of hallmarks of patients with knee OA, these symptoms are significantly more important in patients with moderate-to-severe radiographic signs (18). Therefore, the positive result of the symptoms subscale (the score of this subscale was below 59) can be useful for ruling in the moderate-to-severe group of knee OA. Also, the results of the PPV (88%) of the symptoms subscale indicate that if this test is positive, i.e., the score of this subscale is lower than 59 (the optimal cutoff point), with a probability of 88%, the person has the moderate-to-severe grade of knee OA.

On the other hand, among the subscales of the KOOS questionnaire, the subscale of daily activities has the highest sensitivity and the least amount of NLR; as a result, it has the best screening ability in ruling out the moderate-to-severe group of patients with knee OA. Also, the large NPV of this subscale shows that if the test result is negative or the score of this subscale is higher than 61.5, there is an 85% probability that the person will have mild grade of knee OA. The same pattern of findings has also been reported by Collins et al. who conducted a systematic review with meta-analysis of the measurement characteristics of the KOOS in people with knee injuries and/or osteoarthritis, and concluded that the subscale of daily activities has better content validity for older patients (36). Also, Gandek et al. (37) did a comparative study of the validity and responsiveness of the KOOS questionnaire among total knee replacement patients, and reach the conclusion that the subscale of symptoms and quality of life had the best discriminant validity between comorbid condition groups.

The results of functional tests showed that all three tests have good ability for discrimination between mild and moderate-to-severe groups of knee OA. In addition, among these functional tests, the highest sensitivity, the least amount of NLR, and the greatest NPV were related to step test; as a result it has the best ability for ruling out the moderate-to-severe grade of knee OA. Also, due to the highest specificity and the greatest PLR and PPV, the TUG test, can be useful for confirming the diagnosis of moderate-to-severe grade of knee OA patients. Since no study has examined the relationship between the severity of knee OA and clinical tests of standing balance, as noted in Hatfield et al.’s systematic review and meta-analysis of clinical tests of standing balance in knee OA patients in 2015, no comparison could be made between our results and results of other studies (9). However, it can be noted that the results of the present study are in agreement with previous studies that have examined these functional tests in the knee OA group compared with healthy subjects.

Based on the available literature, the most commonly used clinical test for assessing the standing balance in knee OA patients is the step test (9, 11, 32). According to comparative studies of knee OA patients with healthy older adults, due to the large standardized mean differences in assessing the between group differences, it has been reported that this clinical test may have sufficient sensitivity for detecting and monitoring standing balance deficits (9, 11, 32).

On the other hand, based on previous studies, TUG test is one of the simplest and quickest tests of functional mobility evaluation in knee OA patients, which demonstrates excellent reliability (intra-rater and inter-rater reliability were 97% and 96%, respectively) (27). Also, this test is a predictor of risk of falling in the elderly (38).

Based on the results of present study, the optimal cutoff point for TUG test was determined 11.16 with 80% sensitivity and 90% specificity. In agreement with our results, Shumway-cook et al. (38) showed that the TUG test was a sensitive (0.87) and specific (0.87) to identify elder people who are at risk of falling. They stated that the cutoff point of 14-seconds as the optimal cutoff point significantly discriminates between the faller and non-faller groups (38). In 2002, Rose et al. (39) identified cutoff point of 10-second with 71% sensitivity and 86% specificity as the optimal cutoff point for discrimination between non-faller and recurrent faller. Also, Barry et al. (40) demonstrated that this test with a cutoff point ≥ 13.5 seconds with a higher specificity (73%) than sensitivity could be useful for ruling in individuals at higher risk of falling. Thus, depending on the subjects’ characteristics and walking speed, the cutoff score obtained from this test has been reported differently in different studies. Of course, in some of the previous studies, the diagnostic accuracy of this test in correctly classifying individuals as fallers has been reported poor to moderate (41). Since there was no study of discriminative ability of this test in patients with knee OA, we inevitably compared our study results with those of studies on the elderly. Therefore, we recommend future studies on the discriminative ability of functional measures to confirm our findings in patients of knee OA classified based on severity of OA.

Limitations
We only assessed patients with mild and moderate-to-severe grades of OA of tibiofemoral joint and did not investigate patients with other knee conditions (OA of patellofemoral) that may have functional limitations. Thus, future studies may examine the discriminative ability of these measures between patients with OA of tibiofemoral and other knee pathologies. Also, we only assessed the KOOS questionnaire and three functional tests. We recommend examining
 Discriminability of Functional Measures

The discriminative ability of WOMAC and other functional tests with good reliability and validity (6-min walk test and 30-s chair stand test) for discrimination between the two groups of knee OA patients.

CONCLUSIONS
In brief, all subscales of the KOOS except sport/recreation subscale and all of the performance-based outcome measures (TUG, FRT, Step test) have good ability to discriminate between mild and moderate-to-severe groups of knee OA. Thus, these functional measures could be recommended to be evaluated first for clarifying the severity of functional limitations in patients with knee OA. Based on the results obtained, clinical decisions could be made with the aim of appropriate intervention of these patients. Overall, our findings provide evidence-based information for clinicians and researchers about discriminative ability of functional measures in knee OA patients classified based on radiographic signs.

FUNDINGS
This work was supported by the Ahvaz Jundishapur University of Medical Sciences [pht- 9508].

ACKNOWLEDGEMENTS
We would like to acknowledge Dr. Mohammad Momen Gharibvand for his assistance with grading all radiographs according to the Kellgren-Lawrence classification system. This study is part of PhD thesis of Mrs. Pirayeh. Special thanks to Ahvaz Jundishapur University of Medical Sciences for the financial support (PhD thesis grant no: pht- 9508).

CONFLICT OF INTERESTS
The authors declare that they have no conflict of interests.

REFERENCES


