Comparing the Lower Limb Tasks Questionnaire to the Western Ontario and McMaster Universities Osteoarthritis Index: Agreement, Responsiveness, and Convergence With Physical Performance for Knee Osteoarthritis Patients

Carly McKay, PhD, Harry Prapavessis, PhD, Peter McNair, PhD

From the Exercise & Health Psychology Laboratory, University of Western Ontario, London, Ontario, Canada; and Health & Rehabilitation Research Centre, Auckland University of Technology, Auckland, New Zealand.

Abstract

Objective: To compare the Lower Limb Tasks Questionnaire (LLTQ) with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in terms of agreement, responsiveness, and convergence.

Design: Cross-sectional with an exploratory repeated-measures subsample analysis.

Setting: Community-based seniors’ centers and arthritis clinics.

Participants: Individuals with symptomatic knee osteoarthritis (N = 76) participated, with a subsample of 18 participants contributing to the pre- and postarthroplasty subanalysis.

Intervention: Not applicable.

Main Outcome Measures: Bland and Altman plots of agreement with 95% limits of agreement, statistical responsiveness, and standardized response mean (SRM) were calculated for LLTQ and WOMAC subscales. Both t tests and Wilcoxon rank-sum tests were used to examine changes in pre- and postarthroplasty self-reported function, 50-ft walk speed, stair ascent/descent speed, and isometric quadriceps strength.

Results: The agreement (bias) of the LLTQ activities of daily living (ADL) subscale when compared with the WOMAC physical function (PF) subscale was 1% ± 10% (mean ± SD), and the 95% limits of agreement were −19% to +22%. The statistical responsiveness of the WOMAC-PF and LLTQ ADL was 1.17 and −.63, respectively. The SRMs for these scales were .90 and −.61, respectively. The WOMAC-PF scores showed a notable improvement over the first 6 weeks postarthroplasty, while LLTQ ADL scores were unchanged. The objective measures of function were all significantly worse at 6 weeks.

Conclusions: The LLTQ demonstrated adequate agreement with the WOMAC and acceptable responsiveness for use in place of the WOMAC in nonspecialized clinics. The LLTQ may more accurately represent functional status after total knee arthroplasty, but further study in larger samples is recommended.

Archives of Physical Medicine and Rehabilitation 2013;94:474-9
© 2013 by the American Congress of Rehabilitation Medicine

Osteoarthritis (OA) is one of the most prevalent musculoskeletal conditions among those 65 years and older and is one of the leading causes of disability in this demographic worldwide. It is a degenerative condition characterized by pain, stiffness, and progressive loss of function, but there is an inconsistent relationship between clinical symptoms and radiographic evidence of joint degradation. Patients who report pain or loss of function severe enough to inhibit activities of daily living (ADLs) may have minimal or no associated radiographic findings. Clinicians must therefore rely heavily on self-report measures of symptom severity when determining appropriate treatment courses.

A wide variety of self-report measures are available for use in this population. In an effort to determine how many of these measures are used in practice, Haigh et al conducted a survey of

0003-9993/13/$36 - see front matter © 2013 by the American Congress of Rehabilitation Medicine
http://dx.doi.org/10.1016/j.apmr.2012.09.013
Comparing self-report measures for osteoarthritis

418 European rehabilitation facilities. They found that more than 60 different outcome measures were being used to assess patients with hip and knee OA, with no more than 5 centers using any one instrument. This echoes the findings of similar studies in Canada, Australia, and the United Kingdom, indicating that the use of standard, validated measures for OA assessment is not widespread.11-13 Consequently, there has been a call for the implementation of standardized measures for assessing OA symptoms.8 Yet, many practitioners are not specialists, and maintaining an inventory of questionnaires to assess a variety of conditions is cumbersome. Not only does it require storage space, but clinicians must also be familiar with the administration of each measure and be trained to interpret the scores.14 Furthermore, the use of multiple instruments does not allow the pooling of data, preventing broader analyses of outcomes across clinical populations.15,16 Considering these potential barriers to clinician uptake, it has been suggested that the development of a single instrument to assess outcomes for a number of clinical populations would encourage routine instrument use in everyday practice.16

The Lower Limb Tasks Questionnaire (LLTQ) is a relatively new, function-based, self-report questionnaire that was specifically developed to address issues of clinician uptake. It was formulated based on the recommendations of the World Health Organization to emphasize the delineation between ADLs and sport or recreation activities.16 Furthermore, it was specifically created to be easy to administer and score, and takes patients less than 10 minutes to complete. The LLTQ has been shown to be appropriate for use with several patient groups, including those with sprains, strains, overuse injuries across the lower limb, and low back pain.16 Although the original questionnaire validation study also included patients with OA, a direct comparison of the performance of the LLTQ with that of a standard OA evaluation tool has not yet been undertaken.

The purpose of this study was therefore to compare the LLTQ with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC),17,18 a commonly endorsed measure for hip and knee OA.8,19,20 The LLTQ was assessed in terms of its agreement with the WOMAC, its statistical responsiveness, and its correlation with the pain and function subscales of the WOMAC before and after total knee arthroplasty. Change in LLTQ score was also compared with changes in objective measures of function in the subacute phase after total knee arthroplasty to determine whether it accurately reflects the functional status of the patient.

It was hypothesized that the LLTQ ADL subscale would demonstrate strong convergence with the WOMAC physical function subscale (WOMAC-PF). Additionally, the LLTQ ADL was expected to be equally responsive to changes in functional status after arthroplasty as the WOMAC. Because strong correlations have been found between WOMAC pain and function scores,21-25 it was hypothesized that the LLTQ ADL would correlate with the WOMAC pain subscale, but that this correlation would be lower than that between the WOMAC pain and WOMAC-PF scales. Because of the potential confounder of pain, it was further hypothesized that compared with the WOMAC, the LLTQ ADL would more accurately reflect actual functional status after total knee arthroplasty.

Methods

Participants

Participants were recruited using a convenience sampling strategy through community-based seniors’ centers and arthritis clinics in London, Ontario, Canada. To be eligible, participants had to be 18 years or older, be able to read and write in English, and have been experiencing symptomatic knee OA for a minimum of 6 weeks at the time of questionnaire completion. Participants were not involved in a specific treatment program, but those using oral or topical analgesics and/or a walking aid were included in the sample. Participants were also required to provide informed consent, as per the Office of Research Ethics at the University of Western Ontario.

Procedure

Participation in the study entailed a one-time completion of the WOMAC and LLTQ, which took approximately 20 minutes. After this initial questionnaire administration, a subsample of participants scheduled for total knee arthroplasty also completed the objective function tests (±3d before surgery). This subsample completed both questionnaires and the objective functional tests again 6 weeks (±3d) postoperatively.

Measures

WOMAC

The WOMAC is a 24-item self-administered questionnaire, divided into subscales for pain (5 items), joint stiffness (2 items), and physical function (17 items).18,19 It is rated on a 5-point Likert scale (0–4), with lower scores indicating lower symptom or disability levels. The instrument is scored by summing each subscale to a maximum score of 20, 8, or 68, respectively, or by computing a global score (sum of all 3 subscale scores). Cronbach alpha values for the subscales have reportedly ranged from .86 to .97, and test-retest reliability of the global score ranges from .77 to .83.22,26

LLTQ

The LLTQ is a 20-item self-administered questionnaire, with 10 items forming the ADL subscale and 10 forming the sport/recreation subscale.16 It is scored on a 5-point Likert scale, with lower scores indicating that the respondent has more difficulty performing the given task (eg, walking for 10min, getting in and out of a car). The subscales are summed separately, each to a maximum score of 40, to indicate overall impairment in the 2 functional domains. It is also understood that for some populations, completing the sport/recreation subscale may not be appropriate, and the ADL subscale is sufficient for determining functional disability on its own for these groups. The LLTQ has demonstrated strong internal consistency and concurrent validity, and it is highly reliable (intraclass correlation coefficients [ICCs]) of .96 and .98 for the ADL and sport/recreation subscales.

List of abbreviations:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>activity of daily living</td>
</tr>
<tr>
<td>ICC</td>
<td>intraclass correlation coefficient</td>
</tr>
<tr>
<td>LLTQ</td>
<td>Lower Limb Tasks Questionnaire</td>
</tr>
<tr>
<td>OA</td>
<td>osteoarthritis</td>
</tr>
<tr>
<td>SRM</td>
<td>standardized response mean</td>
</tr>
<tr>
<td>WOMAC</td>
<td>Western Ontario and McMaster Universities Osteoarthritis Index</td>
</tr>
<tr>
<td>WOMAC-PF</td>
<td>WOMAC physical function subscale</td>
</tr>
</tbody>
</table>
respectively). Cronbach alpha values have been reported to be .91 for the ADL subscale and .95 for the sport/recreation subscale, and both domains demonstrate moderate correlations to actual task performance ($r = .62$, $r = .72$).

It was deemed that the sport/recreation subscale was not appropriate for use in the current study population, so only the LLTQ ADL was administered.

**Flat surface walking test**
Participants were asked to walk a distance of 15m, from a standing start, in a straight, quiet corridor. Those who used a walking aid for regular ambulation were permitted to use it during this test. Participants were timed using 2 digital stopwatches (accurate to .01s), and the average of the 2 times was recorded for the trial. Each participant performed 2 trials, separated by 3 minutes. The faster of the 2 average times was used in the analysis.

**Stair ascent/descent**
This test consisted of a stair climb, followed by a stair descent. Participants began from a standing start and were instructed to climb 1 flight (13 steps) of standard stairs, using the railing for balance if necessary. At the top of the stairs, they immediately reversed direction and descended the same staircase. The test was timed using 2 digital stopwatches (accurate to .01s), and the average of the 2 times was recorded for the trial.

**Isometric strength assessment**
After a 5-minute warm-up of walking on a treadmill at a self-selected comfortable pace, participants were seated in an HUR 3530 extension-curl machine, and their thighs were strapped down using inelastic straps with Velcro closures to ensure quadriceps isolation. The lever arm of the machine was positioned at 75°, and the pad was placed just above the foot of the more affected limb. After 2 familiarization bouts at 50% and 75% of maximum effort, respectively, participants were given 5 minutes to rest before 2 test trials were performed. Participants were instructed to contract their quadriceps as forcefully as possible, pushing their leg against the pad of the lever arm. A force meter attached to the lever arm recorded the force output in newtons, and the trial was stopped at the participants’ peak force output. The second trial was performed after a rest period of 3 minutes, and the highest force output from the 2 trials was used in the analysis. Participants received verbal encouragement during both trials.

**Analysis**
Agreement between the LLTQ ADL and the WOMAC-PF was assessed using a Bland and Altman plot of agreement, with associated 95% confidence limits.  This approach uses the variability in individual participant scores, plotting the difference between measurements by the 2 methods against their mean, to show bias between the 2 instruments. Confidence limits are then calculated based on the SD of the mean difference. In the present study, scores on both instruments were standardized to a percentage of the possible total score, and then the LLTQ values were transformed (100 minus percentage score) so that high scores on both instruments indicated greater impairment.

In a subanalysis of those undergoing arthroplasty, statistical responsiveness was calculated as the mean change between preoperative and 6-week postoperative subscale scores, divided by the SD of the preoperative scores. The standardized response mean (SRM) was calculated as the mean score change between the initial and 6-week testing, divided by the SD of the change score. The statistical responsiveness and SRM analyses yielded effect sizes that were interpreted using Cohen’s classifications of small (0.2), medium (0.5), and large (>0.8).

The convergence between the LLTQ and WOMAC before and after total knee arthroplasty was examined using ICCs. Additionally, pre- and postoperative WOMAC, LLTQ, 50-ft timed walk, timed stair ascent/descent, and isometric quadriceps strength scores were compared using paired $t$ tests or Wilcoxon signed-rank tests to determine the direction of change relative to preoperative scores in the subacute postoperative period. The appropriate statistical test was used based on whether these variables violated the assumption of normality.

**Results**
A total of 78 participants were recruited for this study. The overall sample consisted of 44 women (56.4%) and 34 men (43.6%) and had a mean age ± SD of 64.5±16.5y (range, 34–90y). From this sample, 20 individuals underwent total knee arthroplasty. This subsample was composed of 12 women (55.6%) and 8 men (44.4%), with a mean age ± SD of 62.4±6.9y (range, 47–72y).

The agreement between the WOMAC-PF and the LLTQ ADL is presented in figure 1. The bias associated with the LLTQ ADL scale was 1%±10% (mean ± SD), and the 95% limits of agreement were −19% to +22%.

The statistical responsiveness of the WOMAC-PF and LLTQ ADL was 1.17 and −.63, respectively. The SRMs for these scales were .90 and −.61, respectively.

Comparisons between pre- and postoperative WOMAC-PF, LLTQ ADL, 15-m walk, stair ascent/descent, and quadriceps strength scores are presented in table 1. Correlations and ICCs between pre- and postarthroplasty LLTQ ADL and WOMAC subscales are presented in tables 2 and 3.

**Discussion**
The ADL subscale of the LLTQ demonstrated good agreement with the WOMAC-PF, supporting the hypothesis that the 2 scales
would exhibit convergent validity. The small amount of bias indicates that scores on the LLTQ tend to be marginally lower than scores on the WOMAC, but this difference is negligible. The 95% limits of agreement, however, suggest that there is still quite a bit of variability in the differences between the 2 measures. The limits of agreement translate to a raw score difference of −12.92 to +14.96 on the WOMAC-PF. Considering that the minimal clinically important difference for the WOMAC-PF has been reported to be ±10.00, this range of score differences may be meaningful. It must be acknowledged, though, that the minimal clinically important difference is highly context dependent, and the raw score differences may not be large enough to affect treatment decisions.33,34

The LLTQ was expected to be equally responsive to changes in functional status as the WOMAC. Based on the very large effect size associated with the statistical responsiveness of the WOMAC-PF (1.17) and the substantially smaller value corresponding to the LLTQ ADL (−.63), this hypothesis was not supported. Also of note is that the effect sizes attributed to the WOMAC were similar to those previously reported,56 but the small effect sizes associated with the LLTQ are inconsistent with previous research that has demonstrated values ranging from 1.3 to 2.0.16 Although this could be a function of small sample size in the present study, it also may reflect that responsiveness is not an inherent characteristic of a measure, but a product of the sample and context.36,37

There are emerging concerns regarding the structure of the WOMAC subscales, as high correlations between function and pain scores suggest that they are not measuring distinct constructs as intended.21-25 The present findings support this conclusion, since there were high ICCs between the WOMAC-PF and WOMAC pain subscales at both the pre- and postoperative measurement points. As expected, compared with the LLTQ ADL, the WOMAC-PF demonstrated a stronger correlation with the WOMAC pain subscale. Although these correlations appear similar in size, the net response variance difference is between 5% to 8%. In short, these findings suggest that the WOMAC-PF is confounded by pain to a greater extent than the LLTQ ADL. This confounding generally appears to be greater after total knee arthroplasty, which may be important to note when using these self-report instruments to evaluate surgical outcomes.

Because the LLTQ may be less vulnerable to the confounder of pain, it may more accurately reflect functional status after surgery as well. Our findings show that the WOMAC-PF had a notable improvement over the first 6 weeks after surgery, while LLTQ ADL scores were unchanged. Since the objective measures of function were all significantly worse at 6 weeks, this suggests that the WOMAC-PF may overestimate functional improvement in the subacute postoperative phase. This may be because of rapid reductions in pain after surgery, which, considering the strong relationship between the WOMAC subscales, would be reflected in WOMAC-PF scores. Although the LLTQ ADL did not mirror the declines seen in objectively measured function, its lack of improvement indicates that the LLTQ may provide a more accurate assessment of postoperative function. This also suggests that the superior responsiveness found for the WOMAC-PF may be artificially inflated because of the scale’s sensitivity to pain. The low responsiveness of the LLTQ is therefore not likely to be a weakness of the instrument, but merely a product of the patients’ lack of postoperative improvement.

### Study limitations

This study is limited by its use of a convenience sample. Potential underrepresentation of those with severe symptoms does not allow us to generalize the performance of the LLTQ to those with more advanced OA. Moreover, we were unable to assess the effect of potential selection bias; however, this is unlikely to have influenced the relationship between the LLTQ and WOMAC. Because the instruments address the same outcome variables and are constructed using similar language and scales, the correlation between the two is likely a product of the questionnaires themselves, not the underlying characteristics of the target population.

This study is also limited by a relatively small sample size, particularly for the responsiveness analysis. It is possible that the variability seen in score differences between the WOMAC and

---

### Table 1

<table>
<thead>
<tr>
<th>Measures</th>
<th>Preoperative</th>
<th>Postoperative</th>
<th>Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC-PF</td>
<td>30.17±11.91</td>
<td>16.83±11.76</td>
<td>t=−3.83 (P=.001)</td>
</tr>
<tr>
<td>LLTQ ADL</td>
<td>21.25±6.38</td>
<td>21.78±6.86</td>
<td>z=−0.26 (P=.79)</td>
</tr>
<tr>
<td>15-m walk (s)</td>
<td>11.79±4.65</td>
<td>13.54±5.91</td>
<td>z=−2.11 (P=.04)</td>
</tr>
<tr>
<td>Stair ascent/descent (s)</td>
<td>20.44±9.58</td>
<td>25.42±11.73</td>
<td>t=2.34 (P=.04)</td>
</tr>
<tr>
<td>Quadriceps strength (Nm/kg)</td>
<td>0.92±0.56</td>
<td>0.54±0.33</td>
<td>z=3.33 (P&lt;.001)</td>
</tr>
</tbody>
</table>

NOTE. Values are mean ± SD or as otherwise indicated.

### Table 2

<table>
<thead>
<tr>
<th>Measures</th>
<th>WOMAC Pain</th>
<th>WOMAC-PF</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC-PF</td>
<td>.69 (.52—.87)</td>
<td>NA</td>
</tr>
<tr>
<td>LLTQ ADL</td>
<td>.63 (.43—.83)</td>
<td>.74 (.58—.90)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.

### Table 3

<table>
<thead>
<tr>
<th>Measures</th>
<th>WOMAC Pain</th>
<th>WOMAC-PF</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC-PF</td>
<td>.80 (.51—1.10)</td>
<td>NA</td>
</tr>
<tr>
<td>LLTQ ADL</td>
<td>.77 (.48—1.07)</td>
<td>.83 (.62—1.05)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not applicable.
LLTQ in the present sample does not represent the population value, and a larger sample would more accurately estimate the bias or limits of agreement for these instruments. The generalizability of the results is also limited because only 1 treatment type (arthroplasty) was assessed.

Future directions

Additional research is recommended to address the sample size limitations of the present study. Furthermore, examining both the WOMAC and LLTQ in terms of clinically important differences and responsiveness to other treatments is necessary. It would also be useful to get clinician perspectives on the use of standardized instruments in practice to determine the relative ease of administration and interpretation of both questionnaires, with the purpose of identifying features that may be improved to encourage use in clinical settings.

Additional investigation into the relationship between perceived function and pain in this population is also recommended. Determining whether these outcomes can be reliably measured as independent constructs would help to refine the interpretation of clinical results and may assist in treatment decision-making.

Conclusions

The need for many specialized instruments in a clinical setting has been acknowledged as a barrier to practitioner uptake. Based on the results of the present study, using the ADL subscale of the LLTQ may present a reasonable alternative. It demonstrated adequate performance in this sample, and for clinicians who are not currently using a patient-reported outcome measure or who would like to streamline their questionnaire inventory in a non-specialized clinic, the LLTQ ADL is a viable option. The LLTQ may also more accurately represent functional status after total knee arthroplasty and therefore should be considered for use in this population.

Supplier

a. Ab Hur Oy, Patamaentie 4, 67100 Kokkola, Finland.

Keywords

Activities of daily living; Osteoarthritis; Outcome assessment; Rehabilitation

Corresponding author

Carly McKay, PhD, Sport Injury Prevention Research Centre, University of Calgary, 2500 University Dr NW, Calgary, Alberta, Canada, T2N 1N4. E-mail address: cdmckay@ucalgary.ca.

References

24. Terwee C, van der Sluijs R, van Lummel R, Benink R, Meijers W, de Vet H. Self reported physical functioning was more influenced by pain...