

Melbourne, Australia; <sup>‡</sup>Dept. of Physiotherapy, Sch. of Hlth. Sci., The Univ. of Melbourne, Melbourne, Australia

**Purpose:** Measurement of treatment outcomes is critical to both research and clinical practice. The Osteoarthritis Research Society International (OARSI) recommended set of performance-based tests of physical function represents the tests of typical activities relevant to individuals diagnosed with hip or knee OA and following joint replacements. The five tests recommended were the 30s Chair Stand Test (CST), 40m Fast-paced Walk Test (40mFPWT), Stair Climb Test (SCT), the Timed up and Go Test (TUG) and the 6 Minute Walk Test (6MWT). Although the tests were recommended on global opinion, feasibility and the available measurement evidence, none of the tests were able to fulfil all desirable measurement criteria. This study aimed to estimate the reliability of the OARSI recommended set and estimate the amount of measurement error associated with each test.

**Methods:** This was a prospective study with repeated measures between two independent raters within a single session and within one rater over a one-week interval. To be eligible, participants were required to fulfil the inclusion criteria based on clinical diagnostic criteria for knee or hip OA established by the American College of Rheumatology. At the first test session participants performed all five performance tests with two independent raters to examine the reliability between raters within a test session. After a one-week interval at the second test session participants repeated the performance tests again with one rater, to examine reliability within a rater between sessions. Participants completed a self-reported global rating of change questionnaire at the second session to determine if any substantial change in their hip or knee condition had occurred between test sessions. The testing order of both raters and the performance tests was randomized. Reliability was estimated using intra-class correlation coefficients (ICC). Measurement error was expressed as standard error of measurement and minimal detectable change. Interpretation of ICC values was based on inspection of the lower one-sided 95% confidence interval to ensure that it met a recommended *minimum acceptable level*, which was set at 0.70. Point estimates of values more than 0.75 indicated sufficient reliability and values more than 0.90 indicated optimal reliability.

**Results:** Fifty-one people (mean age 64.5 years, SD 6.21 years, range 51–81 years, 24 females (47%)) participated. Fifteen people had primary hip OA (mean age 66.3 years, SD 5.80 years, 9 females (60%)), whilst 36 had primary knee OA (mean age 63.6 years, SD 6.28 years, 15 females (42%)). There was no missing data and no adverse events occurred at any testing occasion. The 40mFPWT, 11-step SCT and 6MWT achieved optimal levels of between rater and within rater reliability (ICC >0.90, lower 1-sided 95%CI >0.70), with acceptable measurement error (<10%). The CST was sufficiently reliable between sessions and between raters (ICC 0.85–0.86, lower 1-sided 95%CI: 0.70–0.79), with acceptable measurement error (<10%). The TUG did not meet minimal acceptable levels of reliability for either test condition (ICC 0.75, lower 1-sided 95% CI: 0.58–0.63). The tests demonstrated a minimal detectable change (MDC<sub>90</sub>) of between 7.6–18% over a week period.

**Conclusions:** This study provides estimates of reliability and measurement error of the OARSI recommended performance-based test of physical function in people with hip and/or knee OA. Except for the TUG, the tests were all sufficiently reliable with acceptable measurement error. The 40mFPWT, 11-step SCT and 6MWT demonstrated the most optimal levels of reliability and lowest measurement error. To be confident of real change in physical performance when applying these tests in individuals with hip and/or knee OA, clinicians and researchers should aim to see changes of at least: 2 stands for the CST; 0.16 m/s for the 40mFPWT; 2.3 seconds for the 11-step SCT; 1.47 seconds for the TUG; and 42.3 meters for the 6MWT.

## 559

### A PROXY TOTAL KNEE ARTHROPLASTY ALGORITHM IS PREDICTIVE OF TOTAL KNEE ARTHROPLASTY: DATA FROM THE OSTEOARTHRITIS INITIATIVE

J.B. Driban <sup>†</sup>, L.L. Price <sup>†</sup>, M. Nevitt <sup>‡</sup>, J. Lynch <sup>‡</sup>, G.H. Lo <sup>§</sup>||, C.B. Eaton <sup>¶</sup>, T.E. McAlindon <sup>†</sup>. <sup>†</sup>Tufts Med. Ctr., Boston, MA, USA; <sup>‡</sup>Univ. of California - San Francisco, San Francisco, CA, USA; <sup>§</sup>Med. Care Line and Res. Care Line, Houston Hlth.Services Res. and Dev. (HSR&D) Ctr. of Excellence Michael E.

DeBakey VAMC, Houston, TX, USA; <sup>||</sup>Section of Immunology, Allergy, and Rheumatology, Baylor Coll. of Med., Houston, TX, USA; <sup>¶</sup>Ctr. for Primary Care and Prevention, Alpert Med. Sch. of Brown Univ., Pawtucket, RI, USA

**Purpose:** A proxy for total knee arthroplasty (TKA) may have potential as an endpoint in osteoarthritis research. Escobar et al. (2003) developed an algorithm in Spain and Riddle et al. (2014) adapted it for application in the Osteoarthritis Initiative (OAI), a large cohort study in the United States. This newly adapted algorithm has only been used among individuals who had a TKA. We postulated that this algorithm might be predictive of total knee replacements with potential use as a proxy for TKA in epidemiologic studies. Therefore, we explored whether the algorithm is associated with a TKA in the next 12 months within the entire OAI cohort. We also explored reasons why knees with a TKA were not classified as “appropriate”.

**Methods:** For these knee-visit based analyses, we included both native knees from all 4,377 participants with complete data in the OAI, an observational study that included those without any risk factors for osteoarthritis, those at high risk for osteoarthritis, and those with pre-existing symptomatic radiographic knee osteoarthritis. We replicated the algorithm described by Riddle et al., which included 16 decision rules based on knee pain and function (WOMAC pain and function in the past 7 days), radiographic severity (Kellgren-Lawrence [KL] grade), number of compartments affected (based on OARSI joint space narrowing scores), joint stability and range of motion (flexion contracture, collateral laxity), and age. Joint stability and range of motion were carried forward because they were only available at certain visits. Based on the algorithm, knees were characterized as “inappropriate or inconclusive” vs “appropriate” for TKA at baseline and the first 4 annual follow-up visits. For our analyses, the focus was on the “appropriate” group - since this would be the proxy endpoint. Among knees that required confirmation of patellofemoral osteoarthritis to classify a knee as “appropriate” one reader reviewed the magnetic resonance images for the presence of both a definite osteophyte and a full or partial thickness cartilage lesion (Hunter et al., 2011). Actual TKAs during the first 5 years of the OAI were self-reported by participants and then confirmed by the OAI coordinating center. To determine if the surgical appropriateness algorithm was related to TKA in the next 12 months we performed a logistic regression with generalized estimating equation to account for correlations within participant over time. We also ran descriptive statistics to understand why some participants were deemed “inappropriate” despite receiving a TKA. Knees were censored after a TKA.

**Results:** Overall, 183 (2.1%) knees had a TKA. The surgical appropriateness algorithm deemed 1067 (2.9%) knee-visits as “appropriate” for surgery. Being deemed appropriate for a TKA was associated with a participant receiving a TKA in the next 12 months (odds ratio = 27.5, 95% confidence interval = 20.4 to 37.1). Specifically, 85 (46.5%) knees with a TKA were deemed appropriate at the visit prior to surgery. At 982 knee-visits (in 501 knees), knees were defined as “appropriate” but did not receive a TKA and 98 knees received a TKA despite being not being classified as “appropriate” at the visit just prior to receiving TKA. The 3 most common decision rules that that resulted in a knee not being classified as “appropriate” despite getting a TKA were 1) KL ≤ 3 with slight or moderate pain (n = 27, 28%), 2) KL = 3 with normal range of motion and stability in adults ≥ 55 years of age with intense or severe symptoms (n = 23, 23%), and 3) unicompartamental tibiofemoral osteoarthritis (KL = 4) in adults ≥ 55 years of age with moderate symptoms (n = 17, 17%). The two most common reasons for being defined as appropriate despite not receiving a TKA were 1) KL = 4 with intense or severe pain and being ≥ 55 years of age (n = 554, 56%) and 2) KL = 3 with intense or severe pain and limited range of motion and poor stability among adults ≥ 55 years of age (n = 322, 33%).

**Conclusions:** The proxy algorithm was strongly associated with individuals having a TKA within 12 months. There were cases of misclassifications. For example, many individuals who received a TKA but were classified as inconclusive or inappropriate had severe radiographic osteoarthritis, but did not meet the threshold definition of symptoms required by the algorithm to be “appropriate” for TKA. This highlights the possibility that the algorithm may benefit from modifications that update the definition of “appropriate TKA” in the United States. Despite the need for further development, such algorithms have great potential as endpoints in osteoarthritis research.