

# Reliability and Clinically Important Improvement Thresholds for Osteoarthritis Pain and Function Scales: A Multicenter Study

Jasvinder A. Singh, Ruili Luo, Glenn C. Landon, and Maria Suarez-Almazor

**ABSTRACT. Objective.** To assess the reliability and clinically meaningful thresholds of intermittent and constant osteoarthritis pain (ICOAP) score, the Knee injury and Osteoarthritis Outcome Score Physical function Short-form (KOOS-PS), the Hip disability and Osteoarthritis Outcome Score Physical function Short-form (HOOS-PS), and the Quality of life subscales of HOOS/KOOS (HOOS-QOL/KOOS-QOL) in patients with knee or hip arthritis.

**Methods.** One hundred and ninety-five patients (141 knee, 54 hip) seen at 2 orthopedic outpatient clinics with a diagnosis of knee or hip OA completed patient-reported questionnaires (ICOAP pain scale, KOOS-PS, HOOS-PS, KOOS-QOL, HOOS-QOL) at baseline and 2-week followup. Reliability was assessed using intraclass correlation coefficients (ICC). We calculated minimum clinically important difference (MCID) and moderate improvement in the subgroup that reported change in the status of their affected joint.

**Results.** The reliability as assessed by ICC was as follows: ICOAP pain scale, 0.63 (0.48, 0.74) in patients with knee arthritis, and 0.86 (0.73, 0.93) for hip arthritis; KOOS-PS, 0.66 (0.52, 0.77); HOOS-PS, 0.82 (0.66, 0.91); KOOS-QOL, 0.79 (0.69, 0.86); and HOOS-QOL, 0.67 (0.42, 0.83). MCID and moderate improvement estimates in patients with knee arthritis were ICOAP pain scale, 18.5 and 26.7; KOOS-PS, 2.2 and 15.0; and KOOS-QOL, 8.0 and 15.6. A smaller sample in patients with hip arthritis precluded MCID and moderate improvement estimates.

**Conclusion.** We found that ICOAP pain and KOOS-PS/HOOS-PS scales were reasonably reliable in patients with hip OA. Reliability of these scales was much lower in patients with knee arthritis. Thresholds for clinically meaningful change in pain or function on these scales were estimated for patients with knee arthritis. (First Release Jan 15 2014; J Rheumatol 2014;41:509–15; doi:10.3899/jrheum.130609)

## Key Indexing Terms:

RELIABILITY      SENSITIVITY TO CHANGE      KNEE      HIP      ARTHRITIS      PAIN

Recent efforts by 2 leading organizations, the Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology Clinical Trials (OMERACT)<sup>1,2</sup>, have led to the development of new pain and function assessments for osteoarthritis (OA). These include the inter-

mittent and constant osteoarthritis pain (ICOAP) score<sup>3</sup> and short forms of 2 validated function scales — the Hip disability and Osteoarthritis Outcome Score Physical function Short-form (HOOS-PS) and the Knee injury and Osteoarthritis Outcome Score Physical function Short-form

*From the Medicine Service, Birmingham VA Medical Center and Department of Medicine, University of Alabama; Center for Surgical Medical Acute Care Research and Transitions, Birmingham VA Medical Center; Division of Epidemiology, School of Public Health, University of Alabama, Birmingham, Alabama; Departments of Health Sciences Research and Orthopedic Surgery, Mayo Clinic School of Medicine, Rochester, Minnesota; Section of Rheumatology and Clinical Immunology, University of Texas, M.D. Anderson Cancer Center; St. Luke's Episcopal Health System, Houston, Texas, USA.*

*Supported by US National Institutes of Health Clinical Translational Science Award 1 KL2 RR024151-01 (Mayo Clinic Center for Clinical and Translational Research), and the resources and facilities at the Birmingham VA Medical Center, Alabama, USA. The original study was supported by unrestricted grants to scientific societies (Osteoarthritis Research Society International and Outcome Measures in Rheumatology) from pharmaceutical companies: Pfizer, Expansciences, Novartis, Negma Lerads, Rottapharm, Fidia, and Pierre Fabre Santé laboratories, and a research grant from the European League Against Rheumatism. Dr. Suarez-Almazor is the recipient of a K24 award from the National Institute for Arthritis, Musculoskeletal, and Skin Disorders (AR053593). Dr. Singh has received research and travel grants from Takeda and*

*Savient, and consultant fees from Savient, Regeneron, Takeda, and Allergan.*

*The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.*

*J.A. Singh, MBBS, MPH, Medicine Service, Center for Surgical Medical Acute Care Research and Transitions, Birmingham VA Medical Center, and Department of Medicine, University of Alabama, and Division of Epidemiology, School of Public Health, University of Alabama, and Departments of Health Sciences Research and Orthopedic Surgery, Mayo Clinic School of Medicine; R. Luo, PhD, Section of Rheumatology and Clinical Immunology, University of Texas, M.D. Anderson Cancer Center; G.C. Landon, MD, St. Luke's Episcopal Health System; M. Suarez-Almazor, MD, PhD, Section of Rheumatology and Clinical Immunology, University of Texas, M.D. Anderson Cancer Center.*

*Address correspondence to Dr. Singh, University of Alabama, Faculty Office Tower 805B, 510 20th Street S, Birmingham, Alabama 35294, USA. E-mail: Jasvinder.md@gmail.com*

*Accepted for publication October 8, 2013.*

(KOOS-PS)<sup>4,5,6</sup>. These assessments are somewhat similar to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)<sup>7</sup> and are being used increasingly in outcome studies on patients with OA.

Published studies have provided initial validation data for these instruments<sup>4,5,6</sup>. In addition, reliability and sensitivity to change data have recently been published. ICOAP was reliable with an intraclass correlation coefficient (ICC) of 0.85 in patients with knee/hip arthritis<sup>3</sup> and ICC ranging 0.65–0.81 in patients who underwent knee/hip replacement surgery<sup>8</sup>. In 2 studies of patients who underwent knee/hip replacement surgery, ICOAP was responsive to change with standardized response means (SRM) ranging from 0.54–1.82<sup>8</sup> and 1.02–2.29<sup>9</sup>, similar to other measures such as WOMAC<sup>9</sup>, higher for hip than knee replacement. The SRM for KOOS-PS and HOOS-PS ranged from 0.54–1.82<sup>8</sup> in patients who underwent knee or hip replacement surgery. However, none of the prior studies estimated clinically important change thresholds for these instruments. In addition, validation in US cohorts has not been done.

The primary objective of our study was to examine the test-retest reliability of ICOAP pain, HOOS-PS, and KOOS-PS questionnaires and the effect of age, race/ethnicity, and sex on reliability in a multicenter US study. We also assessed the thresholds for clinically important differences for these questionnaires in patients with knee or hip arthritis.

## MATERIALS AND METHODS

**Study population.** Our study included patients recruited in 2 large medical centers (Veterans Affairs Medical Center, Minneapolis, Minnesota, and M.D. Anderson Cancer Center, Houston, Texas). Cohorts consisted of consecutive patients with a diagnosis of knee or hip OA who had radiographic evidence of hip or knee OA and were referred to orthopedic surgeons for consideration of joint replacement surgery. Patients were excluded if they had no knee/hip OA, prior knee or hip replacement, or concomitant inflammatory arthritis, or were unable to complete the questionnaire. These patients were recruited as part of an international multicenter study of patients with knee or hip pain. Details of the original study are described elsewhere<sup>10</sup>. As part of the original study, patients completed pain, function, and quality of life (QOL) assessments at the initial visit only. For our study, each patient who completed the baseline survey at the 2 US centers also received the same survey by mail at 2 weeks to test reliability and clinically important improvement thresholds. The study was approved by the institutional review boards at the Minneapolis VA medical center and the M.D. Anderson Cancer Center.

**Validation and statistical analyses.** All patients received a repeat survey 2 weeks after the first survey with the same questions as the first plus 2 additional questions. The first additional question was whether they had undergone a joint replacement in the joint for which they were evaluated. The second additional question was “Since the last time you completed the survey 2 weeks ago, would you say your hip (or knee) arthritis is: A great deal better, somewhat better, about the same, somewhat worse or a great deal worse?” Patients were asked to choose one of the response options. Patients were included in the analyses if they had not undergone joint replacement surgery, had answered the second question, and had returned their survey. Sensitivity analyses were performed in a subset that responded within 20 days of the first survey (an extra 6-day window was allowed for delay because of mailing time).

Each patient completed the following self-reported validated questionnaires: (1) knee or hip function assessment – either the HOOS-PS for hip or KOOS-PS for knee for function<sup>4,5,6</sup>, developed as short forms for assessments of physical function that has been translated into multiple languages<sup>11,12</sup>; (2) pain assessment with ICOAP score<sup>3,9</sup> translated into other languages<sup>13,14,15</sup>; and (3) KOOS knee-related QOL, KOOS-QOL, and HOOS-QOL subscales of the original KOOS and HOOS questionnaires<sup>16,17</sup>. The score range for ICOAP pain, KOOS-PS, and HOOS-PS is 0–100, 100 being the worst. The score range for KOOS-QOL and HOOS-QOL is 0–100, 100 being the best. ICOAP pain questionnaire has 11 questions that are used to calculate the overall ICOAP pain; 5 questions related to constant pain and 6 questions to intermittent pain, which are used to calculate ICOAP intermittent pain and ICOAP constant pain scores. All 3 ICOAP pain scores ranged 0–100. ICOAP, KOOS-PS, and HOOS-PS were administered as complete instruments; KOOS-QOL and HOOS-QOL are in the 5 subscales of the original KOOS and HOOS questionnaires<sup>17,18,19</sup> that were administered as part of the original study<sup>10</sup>, while the other 4 subscales were not administered to reduce patient burden and because of relevance of the original study.

Those 110 patients who reported their hip (or knee) arthritis being “about the same” were included for the reliability/reproducibility analyses. We used ICC to assess the correlation between baseline and followup assessments. Ninety-five percent CI and p values were presented. ICC was also calculated for patient subgroups by age, sex, and race/ethnicity as follows: (1) age group: < 65 versus ≥ 65 years old; (2) sex; (3) race/ethnicity: white versus nonwhite. We used 1-way ANOVA to compute the ICC and determine the between-subject variation and within-subject variation as a measure of test-retest reliability.

Patients who responded that their knee (or hip) arthritis was a great deal better or somewhat better constituted the datasets used for estimating minimum clinically important difference (MCID) for improvement (somewhat better), as recommended<sup>20,21,22</sup>, similar to previous studies<sup>23,24</sup> and moderate improvement (a great deal better), respectively. This was calculated as the mean of the difference between baseline and followup score for each patient reporting that their arthritis was better since the last survey. We also calculated Minimal Detectable Change (MDC) using a statistical anchor to estimate meaningful change<sup>25</sup>. All analyses were done using SAS, version 9.3. Statistical significance was set at  $p < 0.05$ .

## RESULTS

**Study population.** Clinical and demographic characteristics of the source and study populations are summarized in Table 1. Of the 107 patients in Minneapolis and 176 patients in Houston recruited for the original study that included the baseline survey<sup>10</sup>, 83 patients from Minneapolis and 112 patients from Houston (total 195 patients) returned the followup mailed surveys and constituted the analytic dataset (Table 1). Of those, 79 patients from Minneapolis and 71 patients from Houston returned their 2-week surveys within 20 days of the first survey and constituted the dataset for sensitivity analyses. Thus, our study included 195 patients: 141 had knee OA and 54 had hip OA. Four patients did not answer the patient global question, so were not eligible for reliability or sensitivity to change analyses.

The mean age of the patients was 61 years, 43% were female, 74% were white, 66% were married, and the mean body mass index was 33.9 kg/m<sup>2</sup>. The mean (SD) time to second survey completion and receipt was 17.6 days ( $\pm$  6.8).

Compared to responders, nonresponders to the followup survey were younger (57 vs 61 yrs;  $p = 0.0065$ ) but had no

Table 1. Clinical and demographic features of the source population and overall study cohort that responded to both baseline and followup survey.

	All, n = 195	Houston, n = 112	Minneapolis, n = 83
Age, yrs, mean (SD)	60.8 (± 11.4)	60.1 (± 8.9)	61.8 (± 14)
BMI, kg/m <sup>2</sup> , mean (SD)	33.9 (± 7.2)	34.9 (± 7.7)	32.6 (± 6.4)
Duration of symptoms, mos, mean (SD)	111.1 (± 129.9)	92.4 (± 114.2)	138.4 (± 146.3)
Sex, n (%)			
Female	83 (42.6)	77 (92.8)	6 (7.2)
Male	112 (57.4)	35 (31.3)	77 (68.8)
Marital status, n (%)			
Married	129 (66.2)	78 (60.5)	51 (39.5)
Not married	66 (33.8)	34 (51.5)	32 (48.5)
Education, n (%)			
Above college	67 (34.4)	44 (65.7)	23 (34.3)
Some college	100 (51.3)	54 (54)	46 (46)
High school or below	27 (13.8)	14 (51.9)	13 (48.1)
Ethnicity, n (%)			
Black	37 (19)	34 (91.9)	3 (8.1)
Hispanic	6 (3.1)	5 (83.3)	1 (16.7)
White	144 (73.8)	68 (47.2)	76 (52.8)
Other	8 (4.1)	5 (62.5)	3 (37.5)
Employment status, n (%)			
Employed	85 (43.6)	60 (70.6)	25 (29.4)
Other	109 (55.9)	52 (47.7)	57 (52.3)
Lives alone, n (%)			
No	140 (71.8)	84 (60)	56 (40)
Yes	54 (27.7)	28 (51.9)	26 (48.1)
ICOAP pain, mean (SD)	53.9 (± 22.3)	54.8 (± 22.5)	52.6 (± 22.2)
ICOAP constant pain, mean (SD)	51.1 (± 25.6)	51.7 (± 25.6)	50.1 (± 25.6)
ICOAP intermittent pain, mean (SD)	56.3 (± 22.8)	57.4 (± 23)	54.7 (± 22.6)
HOOS-PS/KOOS-PS, mean (SD)	55 (± 20)	55.7 (± 18.9)	54 (± 21.4)
HOOS-QOL/KOOS-QOL, mean (SD)	26.9 (± 20.9)	24 (± 20.3)	30.7 (± 21.1)

ICOAP: intermittent and constant osteoarthritis (OA) pain score; HOOS-PS: Hip disability and OA Outcome Score Physical function Short-form; KOOS-PS: Knee injury and OA Outcome Score Physical function Short-form; QOL: quality of life; BMI: body mass index.

significant differences in marital status, education level, ethnicity, and employment status.

*Clinically important change for improvement.* For patients reporting change in knee or hip arthritis transition question, we calculated estimates for MCID and moderate improvement. The distribution of patients in these categories is presented in Table 2. The baseline and followup scores in patients who improved somewhat or a great deal, or were about the same are shown in Table 3. The MCID estimates for improvement in ICOAP pain, ICOAP

constant pain, and ICOAP intermittent pain were 18.5, 18.7, and 18.4, respectively; respective moderate improvement estimates were 26.7, 29.6, and 24.3. For KOOS-PS, MCID and moderate improvements were 2.2 and 15.0 (Table 4). MCID and moderate improvement estimates for KOOS-QOL were 8.0 and 15.6, respectively (Table 4). Sensitivity analyses with 150 patients showed minimal changes (data available from authors on request).

*Reproducibility and reliability.* One hundred ten patients (81 with knee arthritis; 29 with hip arthritis) reported that their

Table 2. Patients by response to the patient global question at followup survey.

Since the last time you completed the survey 2 weeks ago, would you say your hip (or knee) arthritis is:	Combined, n = 191 <sup>a</sup> n (%)	Houston, n = 112, n (%)	Minneapolis, n = 79 <sup>a</sup> n (%)
A great deal better	14 (7)	8 (7)	6 (8)
Somewhat better	32 (17)	21 (19)	11 (14)
About the same	110 (58)	58 (52)	52 (66)
Somewhat worse	27 (14)	22 (20)	5 (6)
Great deal worse	8 (4)	3 (3)	5 (6)

<sup>a</sup> Four patients from Minneapolis site did not respond to this transition question at the followup survey; p = 0.05, comparing Houston to Minneapolis site.

Table 3. Mean scores on instrument subscales at baseline and followup in patients with symptoms of knee osteoarthritis (OA)<sup>a</sup>.

Knee	Baseline, n = 141	Great Deal Better, n = 12	Patient Global on Followup	
			Somewhat Better, n = 26	About the Same, n = 81
ICOAP pain	53.3 (± 23.0)	17.4 (± 12.2)	41.3 (± 14.1)	47.4 (± 22.0)
ICOAP constant pain	50.7 (± 26.0)	13.8 (± 13.3)	38.5 (± 19.7)	43.0 (± 24.3)
ICOAP intermittent pain	55.5 (± 23.3)	20.5 (± 14.9)	43.6 (± 15.5)	51.1 (± 22.1)
HOOS-PS/KOOS-PS	54.5 (± 19.4)	31.4 (± 12.2)	52.1 (± 17.4)	51.0 (± 18.5)
HOOS-QOL/KOOS-QOL <sup>b</sup>	27.4 (± 21.1)	59.4 (± 19.9)	34.8 (± 16.5)	29.5 (± 20.2)

Total N = 141. <sup>a</sup> Only 1 and 5 patients reported feeling a great deal better or somewhat better among those with symptoms of hip OA, thereby not allowing us to perform meaningful analyses for patients with symptoms of hip OA. <sup>b</sup> N = 20 for “somewhat better” and N = 69 for the “about the same” categories for HOOS-QOL/KOOS-QOL scores; 11 patients reported worse status on patient global question. ICOAP: intermittent and constant osteoarthritis (OA) pain score; HOOS-PS: Hip disability and OA Outcome Score Physical function Short-form; KOOS-PS: Knee injury and OA Outcome Score Physical function Short-form; QOL: quality of life.

Table 4. Minimum clinically important difference (MCID) and moderate improvement thresholds in patients with symptoms of knee osteoarthritis (OA).

	Improvement in Score, Mean (SD)	Relative % Improvement in Score	MDC90
ICOAP pain (scale 0–100)			
Moderate improvement (“great deal better”), n = 12	–26.7 (± 14.8)	–60.5	49.6
MCID improvement (“somewhat better”), n = 26	–18.5 (± 22.0)	–31.0	46.6
ICOAP constant pain			
Moderate improvement (“great deal better”), n = 12	–29.6 (± 14.8)	–68.3	53.8
MCID improvement (“somewhat better”), n = 26	–18.7 (± 24.4)	–32.7	49.6
ICOAP intermittent pain			
Moderate improvement (“great deal better”), n = 12	–24.3 (± 17.9)	–54.3	48.7
MCID improvement (“somewhat better”), n = 26	–18.4 (± 25.4)	–29.7	50.8
KOOS-PS (scale 0–100)			
Moderate improvement (“great deal better”), n = 12	–15.0 (± 16.4)	–32.3	35.5
MCID improvement (“somewhat better”), n = 26	–2.2 (± 17.5)	–4.0	28.3
KOOS-QOL (scale 0–100)			
Moderate improvement (“great deal better”), n = 12	15.6 (± 18.8)	35.7	39.0
MCID improvement (“somewhat better”), n = 25	8.0 (± 16.1)	25.7	29.0

Relative % improvement in score was calculated as = 100 (followup score mean–baseline score mean)/baseline score mean. ICOAP: intermittent and constant OA pain score; KOOS-PS: Knee injury and OA Outcome Score Physical function Short-form; QOL: quality of life; MDC: minimal detectable change.

arthritis was about the same as at the time of the baseline survey. They constituted the analytic dataset for the assessment of reproducibility/reliability (Table 5). The ICC was 0.63 for ICOAP pain in patients with knee arthritis and 0.86 for hip arthritis. Respective ICC for ICOAP constant pain were 0.57 and 0.81 and for ICOAP intermittent pain were 0.64 and 0.83. ICC was 0.66 for KOOS-PS and 0.82 for HOOS-PS (Table 5). ICC for KOOS-QOL was 0.79 and for HOOS-QOL was 0.67.

Data on variation in reproducibility of ICOAP pain, HOOS/KOOS PS and QOL by age, sex, and race/ethnicity are available on request from authors. We noted variation in reproducibility in KOOS-PS by sex and race/ethnicity and by race/ethnicity in KOOS-QOL. In the hip cohort, variations in reproducibility for HOOS-PS and HOOS-QOL were

Table 5. Reproducibility assessed with intraclass correlation coefficients (ICC) in patients with knee and hip osteoarthritis (OA).

	ICC (95% CI)	
	Knee, n = 81	Hip, n = 29
ICOAP pain	0.63 (0.48, 0.74)	0.86 (0.73, 0.93)
ICOAP constant pain	0.57 (0.40, 0.70)	0.81 (0.64, 0.90)
ICOAP intermittent pain	0.64 (0.49, 0.75)	0.83 (0.68, 0.91)
KOOS-PS	0.66 (0.52, 0.77)	NA
HOOS-PS	NA	0.82 (0.66, 0.91)
KOOS-QOL	0.79 (0.69, 0.86)	NA
HOOS-QOL	NA	0.67 (0.42, 0.83)

ICOAP: intermittent and constant OA pain score; HOOS-PS: Hip disability and OA Outcome Score Physical function Short-form; KOOS-PS: Knee injury and OA Outcome Score Physical function Short-form; QOL: quality of life.



noted by race/ethnicity (data available from authors). Sociodemographic and clinical characteristics of sensitivity cohort data are available from authors on request. Sensitivity analyses with 150 patients showed minimal changes compared to the main analyses of the cohort of 195 patients (data available from authors).

## DISCUSSION

In this 2-center study of ethnically diverse US cohorts, we found that 3 assessments of pain and function, i.e., ICOAP, KOOS-PS, and HOOS-PS, were reproducible in patients with knee and hip arthritis. Reproducibility/reliability of ICOAP pain, HOOS-PS, and KOOS-PS was good in hip OA (0.82–0.88) and moderate in patients with knee OA (0.52–0.66). Reliability varied somewhat with age, sex, and race/ethnicity, as expected. Our study also provided reliability statistics for KOOS-QOL and HOOS-QOL scales. We also present estimates for MCID and moderate improvement for these scales in patients with knee arthritis.

The main finding from our study was that ICOAP, KOOS-PS, and HOOS-PS had moderate to good test-retest reproducibility. In a recent single center study in Europe that assessed test-retest reliability at 2 weeks in patients with OA who later underwent joint arthroplasty, ICC for ICOAP, HOOS-PS, and KOOS-PS scales ranged from 0.80–0.84 in patients with hip OA and 0.65–0.85 in knee OA, similar to WOMAC<sup>8</sup>. Our ICC were within this previously reported range; thus our multicenter study confirms this earlier finding in a more ethnically diverse population. The study cohorts were assembled similarly in the 2 studies. However, we had a racially/ethnically diverse population compared to the previous study, which might partially explain an ICC toward the lower end of the range in our knee cohort. Another potential reason for lower ICC in knee cohort may be a week-to-week variation in knee pain that is not determined in the global question asking about the worsening of arthritis. The ICC may also differ because of differences in prevalence of disease (which has been shown to affect ICC<sup>26</sup>) or because of random variation, given a small sample size for the hip cohort. The studies differed in that we analyzed only patients who reported no change in the status of their knee/hip arthritis between 2 visits (56.4% of the cohort) versus analyses of all patients in the previous study (because the transition question was not asked) for reliability assessment, a more conservative approach in our study that takes into account any change between assessments. In our study, reliability for KOOS-QOL and HOOS-QOL were 0.79 and 0.67, respectively, lower than the ICC of 0.89 reported for the Persian version of KOOS-QOL<sup>27</sup>. This may be due to minor content differences from translation or a difference in study populations.

We noted minimal variation in reproducibility for pain and minimal to moderate variation in reproducibility for QOL assessments by various patient characteristics

including age, sex, and race/ethnicity. Most 95% CI were overlapping, signifying that these differences were not statistically significant in this small sample. This indicated the lack of difference or the lack of power to detect a small difference. These findings highlight the effect of patient characteristics on patient-reported outcomes. This was not unexpected and has been reported previously with Medical Outcomes Study Short-Form 36, with reliability ranging from 0.65 to 0.94 across patient groups<sup>28</sup>. However, to our knowledge, reliability variation has not been studied in the previous validation studies of most other instruments. One must also not overinterpret these findings, and they need to be reproduced. These findings also suggest that future studies reporting on validity of instruments should consider controlling the statistics by sex, race, and age, which could all provide useful guidance to the users of the instruments. This also raises the question of whether various validation characteristics of instruments, such as MCID and validity statistics that are usually reported for overall populations and not for individual subgroups, are more accurate and applicable to some patient subgroups than others. This is likely the case, because the overall average incorporates a range among respondents. This is a broad research agenda, not limited to these instruments, which needs more attention.

Our study provides estimates of clinically important improvement by estimating MCID and moderate improvement for these instruments in patients with knee OA, thus adding to the current knowledge; the hip OA sample was not large enough to perform analyses. ICOAP<sup>8,9</sup>, HOOS-PS, and KOOS-PS<sup>8</sup> have been shown to be sensitive to change in previously published studies in patients who underwent knee or hip arthroplasty, surgeries demonstrated to be associated with significant improvements in pain and function after knee/hip joint replacement, similar to WOMAC. Thus, these measures have desirable psychometric properties.

Estimation of clinically meaningful changes is critical for validated instruments, because it provides guidance for calculating sample sizes for studies aimed at examining patient-relevant outcomes and comparing different interventions in patients with knee/hip OA or those undergoing arthroplasty, such as comparing arthroplasty implants or treatment pathways. Moderate improvement represents important change for the patients. For the knee cohort, moderate improvement was estimated at 27 units for ICOAP pain scale, 30 units for ICOAP constant pain, and 24 units for ICOAP intermittent pain. To our knowledge, this is the first study to provide MCID and moderate improvement estimates for these validated scales. We also provided estimates for MCID for KOOS-PS and KOOS-QOL. These thresholds represent meaningful changes that can be appreciated by patients as above and beyond the daily variation. Future studies should estimate MCID and moderate

improvement for patients with hip OA; because only a few patients provided this information, we were unable to estimate those.

Our study findings must be interpreted considering the limitations. Patients were recruited from orthopedic offices where they were assessed for joint replacement surgery, and these estimates may be different for populations with milder arthritis or other causes of knee pain, and in younger patients. Our study was not powered to examine differences in reliability by patient characteristics (age, sex, and race) and therefore we may have missed small differences. Despite a reasonable study sample size, we had a small sample for estimating clinically important differences, especially for moderate improvement estimations (standard errors were large), and for assessing reliability in the hip cohort. Although similar sample sizes have been used to derive reliability and validation statistics<sup>29,30,31</sup>, our confidence in these estimates is not high. Also, a small sample did not allow us to examine MCID and moderate improvement thresholds by categories of baseline scores. More studies are needed to confirm our estimates in larger groups of patients. Study strengths include that this was a multicenter study, we recruited a significant number of minority patients and women, and we assessed instruments that are relevant to patients with OA.

We found that in the hip cohort, ICOAP pain, HOOS-PS, and HOOS-QOL had moderate to high reliability. Reliability was moderate for ICOAP pain and KOOS-PS and high for KOOS-QOL in the knee cohort. Our study provided estimates for clinically meaningful changes for improvement for these assessments, which can be used to calculate sample sizes for future randomized and cohort studies and allow comparison of different treatment modalities/implants. Future studies should assess whether our estimates for clinically meaningful changes in patients with knee or hip arthritis are stable across other populations.

## ACKNOWLEDGMENT

The authors thank the patients for participation in the original study.

## REFERENCES

1. Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am* 2001;83-A:1856-64.
2. Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R, et al. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. *J Rheumatol* 2007;34:1432-5.
3. Hawker GA, Davis AM, French MR, Cibere J, Jordan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure—an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:409-14.
4. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert JF, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthritis Cartilage* 2009;17:843-7.
5. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:551-9.
6. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008;16:542-50.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
8. Ruysen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. *Clin Exp Rheumatol* 2011;29:231-7.
9. Davis AM, Lohmander LS, Wong R, Venkataramanan V, Hawker GA. Evaluating the responsiveness of the ICOAP following hip or knee replacement. *Osteoarthritis Cartilage* 2010;18:1043-5.
10. Gossec L, Paternotte S, Maillefert JF, Combescur C, Conaghan PG, Davis AM, et al. The role of pain and functional impairment in the decision to recommend total joint replacement in hip and knee osteoarthritis: an international cross-sectional study of 1909 patients. Report of the OARSI-OMERACT Task Force on total joint replacement. *Osteoarthritis Cartilage* 2011;19:147-54.
11. Ornetti P, Perruccio AV, Roos EM, Lohmander LS, Davis AM, Maillefert JF. Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). *Osteoarthritis Cartilage* 2009;17:1604-8.
12. Goncalves RS, Cabri J, Pinheiro JP, Ferreira PL, Gil J. Reliability, validity and responsiveness of the Portuguese version of the Knee injury and Osteoarthritis Outcome Score—Physical Function Short-form (KOOS-PS). *Osteoarthritis Cartilage* 2010;18:372-6.
13. Goncalves RS, Cabri J, Pinheiro JP, Ferreira PL, Gil J. Cross-cultural adaptation and validation of the Portuguese version of the intermittent and constant osteoarthritis pain (ICOAP) measure for the knee. *Osteoarthritis Cartilage* 2010;18:1058-61.
14. Maillefert JF, Kloppenburg M, Fernandes L, Punzi L, Gunther KP, Martin Mola E, et al. Multi-language translation and cross-cultural adaptation of the OARSI/OMERACT measure of intermittent and constant osteoarthritis pain (ICOAP). *Osteoarthritis Cartilage* 2009;17:1293-6.
15. Kessler S, Grammozis A, Gunther KP, Kirschner S. Die intermittierende und konstanter Schmerz-Score (ICOAP) - ein Fragebogen, um Schmerzen bei Patienten mit Gonarthrose bewerten (German). The intermittent and constant pain score (ICOAP) - a questionnaire to assess pain in patients with gonarthrosis. *Z Orthop Unfall* 2011;149:22-6.
16. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998;28:88-96.
17. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol* 2003;32:46-51.
18. Nilsson AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)—validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10.

19. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes* 2003;1:17.
20. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
21. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991;4 Suppl:142S-58S.
22. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.
23. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain* 2004;8:283-91.
24. Fortin PR, Stucki G, Katz JN. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum* 1995;38:1027-30.
25. Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273-80.
26. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005;58:246-51.
27. Salavati M, Akhbari B, Mohammadi F, Mazaheri M, Khorrami M. Knee injury and Osteoarthritis Outcome Score (KOOS); reliability and validity in competitive athletes after anterior cruciate ligament reconstruction. *Osteoarthritis Cartilage* 2011;19:406-10.
28. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40-66.
29. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987;67:206-7.
30. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry* 1989;46:1006-11.
31. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 1997;36:980-8.