**Clinical Orthopaedics
and Related Research®**
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

CLINICAL RESEARCH

Validation of the HOOS, JR: A Short-form Hip Replacement Survey

Stephen Lyman PhD, Yuo-Yu Lee MS, Patricia D. Franklin MD,
MBA, MPH, Wenjun Li PhD, David J. Mayman MD, Douglas E. Padgett MD

Abstract

Background   Patient-reported outcome measures (PROMs) are increasingly in demand for outcomes evaluation by hospitals, administrators, and policymakers. However, assessing total hip arthroplasty (THA) through such instruments is challenging because most existing measures of hip health are lengthy and/or proprietary.

Questions/purposes   The objective of this study was to derive a patient-relevant short-form survey based on the Hip disability and Osteoarthritis Outcome Score (HOOS), focusing specifically on outcomes after THA.

Methods   We retrospectively evaluated patients with hip osteoarthritis who underwent primary unilateral THA and who had completed preoperative and 2-year postoperative PROMs using our hospital's hip replacement registry. The 2-year followup in this population was 81% (4308 of 5351 patients). Of these, 2371 completed every item on the HOOS before surgery and at 2 years, making them eligible for the formal item reduction analysis. Through semistructured interviews with 30 patients, we identified items in the HOOS deemed qualitatively most important to patients with hip osteoarthritis. The original HOOS has 40 items, the four quality-of-life items were excluded a priori, five were excluded for being redundant, and one was excluded based on patient-relevance surveys. The remaining 30 items were evaluated using Rasch modeling to yield a final six-item

S. Lyman (✉), Y.-Y. Lee
Healthcare Research Institute, Hospital for Special Surgery, 530 East 70th Street, New York, NY 10021, USA
e-mail: lymans@hss.edu

P. D. Franklin, W. Li
Department of Orthopedics and Physical Rehabilitation, University of Massachusetts Medical School, Worcester, MA, USA

D. J. Mayman, D. E. Padgett
Adult Reconstruction and Joint Replacement Service, Hospital for Special Surgery, New York, NY, USA

HOOS, Joint Replacement (HOOS, JR), representing a single construct of "hip health." We calculated HOOS, JR scores for the Hospital for Special Surgery (HSS) cohort and validated this new score for internal consistency, external validity (versus HOOS and WOMAC domains), responsiveness to THA, and floor and ceiling effects. Additional external validation was performed using calculated HOOS, JR scores in collaboration with the Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) nationally representative joint replacement registry (n = 910).

Results The resulting six-item PROM (HOOS, JR) retained items only from the pain and activities of daily living domains. It showed high internal consistency (Person Separation Index, 0.86 [HSS]; 0.87 [FORCE]), moderate to excellent external validity against other hip surveys (Spearman's correlation coefficient, 0.60–0.94), very high responsiveness (standardized response means, 2.03 [95% CI, 1.84–2.22] [FORCE]; and 2.38 [95% CI, 2.27–2.49] [HSS]), and favorable floor (0.6%–1.9%) and ceiling (37%–46%) effects. External validity was highest for the HOOS pain (Spearman's correlation coefficient, 0.87 [95% CI, 0.86–0.89] [HSS]; and 0.87 [95% CI, 0.84–0.90] [FORCE]) and HOOS activities of daily living (Spearman's correlation coefficient, 0.94 [95% CI, 0.93–0.95] [HSS]; and 0.94 [95% CI, 0.93–0.96] [FORCE]) domains in the HSS validation cohort and the FORCE-TJR cohort.

Conclusions The HOOS, JR provides a valid, reliable, and responsive measure of hip health for patients undergoing THA. This short-form PROM is patient relevant and efficient.

Level of Evidence   Level III, diagnostic study.

Introduction

In this era of focus on patient-centered outcomes, there is an increasing demand for patient-reported outcome measures (PROMs) to assess the effectiveness of elective orthopaedic surgical procedures [4, 7, 24]. The total hip arthroplasty (THA), which is one of the most successful interventions for improving patients' quality of life [36], is no exception. The Harris hip score, Western Ontario & McMaster Universities Arthritis Index (WOMAC), Oxford Hip Score, and Hip disability and Osteoarthritis Outcome Score (HOOS) are all used in various contexts to measure the outcomes after THA [3, 16, 17, 23, 25–27, 29], hip resurfacing arthroplasty, and other end-stage hip osteoarthritis treatments. However, because most of these instruments are lengthy, administration can disrupt clinic flow, whereas incomplete survey responses and other inefficiencies and limitations are not infrequent. Although well designed for research purposes, they have not been universally adopted because they have

not proven suitably efficient as tools for large patient registries and other outcomes reporting needs [2, 14].

The Centers for Medicare & Medicaid Services (CMS) recently released their proposed knee and hip arthroplasty PROMs to meet their pay-for-performance measures [6], with the expectation that these surveys be patient-centered and nonproprietary. The Harris hip score is partially surgeon-derived, the WOMAC is proprietary, and the Oxford Hip Score is partially proprietary (free licensing under certain circumstances, but supporting documentation requires payment), leaving the 40-question HOOS as the only CMS-recommended hip-specific measure. The HOOS physical function survey (HOOS-PS) is a short-form survey developed for patients with hip limitations, but validation of this instrument was not limited to patients with advanced osteoarthritis and purposely excluded the HOOS pain domain questions, which is the dominant reason for which patients undergo THA [8]. We therefore endeavored to develop a nonproprietary short-form hip-specific PROM that meets the CMS requirements for outcome measurement and is an efficient method of capturing these outcomes, as the 40-question HOOS may be burdensome for patients and may disrupt clinic flow.

The objective of our study was to derive a short-form survey based on the HOOS focusing specifically on outcomes after THA. Specifically, we sought to develop and validate a new tool in a population of patients undergoing THA, with particular attention to internal consistency, external validity, responsiveness to THA, and floor and ceiling effects.

Patients and Methods

In designing this study we considered the outcomes measure criteria recommend by Fitzpatrick, et al. [12] of appropriateness, reliability, validity, responsiveness, precision, interpretability, acceptability, and feasibility. Because we endeavored to derive a short-form PROM rather than develop a new instrument, we relied on the framework proposed by Rothman et al. [31] in the 2009 International Society For Pharmacoeconomics and Outcomes Research report on use and modification of existing PROMs.

Subjects

Derivation and validation of the HOOS, Joint Replacement (HOOS, JR.) was performed at the Hospital for Special Surgery (HSS) using data from our existing institutional review board-approved total hip replacement registry, which enrolled patients between May 1, 2007, and January 31, 2012. Our institutional registry prospectively collected patient demographics and PROMs for THA, including the

HOOS and WOMAC, but this validation effort was performed retrospectively.

Approximately 85% of all patients undergoing primary unilateral THA for osteoarthritis consented for registry participation, at which time they were administered a preoperative HOOS Survey. Approximately 84% of these patients returned a baseline HOOS survey. Of these, approximately 81% also returned a 2-year HOOS survey. Patients who returned the HOOS survey preoperatively and at 2 years postoperatively were eligible for inclusion in the HOOS, JR item eligibility assessment (n = 4308), whereas only patients who completed every item in the preoperative HOOS were included in the item reduction and validation process (n = 2371). The large decrease in eligible patients was attributable to several frequently skipped items being deemed eligible for inclusion (eg, "difficulty squatting" was eligible, but 14.4% of patients skipped the item, which would eliminate all patients who did not complete the squatting item). For the final inclusion assessment and validation, we randomly divided the full cohort into learning (n = 1186) and validation (n = 1185) cohorts for the purpose of building (learning) and validating (validation) the new PROM. The development and validation process was performed using full HOOS surveys rather than administering the HOOS, JR to new patients. External validation was performed using full HOOS surveys from 910 patients, who had unilateral THAs, from the nationally representative Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) registry who completed preoperative and 2-year postoperative HOOS surveys [1, 13].

Item Eligibility Assessment

The HOOS consists of 40 self-administered items (Table 1) in five domains: pain (10 items); symptoms (five items); activities of daily living ([ADL]; 17 items); sports and recreation (four items); and hip-related quality of life ([QOL]; four items). A priori, we excluded the four questions from the QOL domain because unlike other HOOS items, they do not address specific hip movements or activities. A preliminary analysis of the feasibility of using Rasch analysis for development of this short-form also excluded all items from the QOL domain.

Before initiation of the validation, 30 consecutive patients from four surgeons scheduled for primary THA were asked to rate the importance of each item in the HOOS survey on a scale from 1 to 3 (1 = unimportant, 2 = somewhat important, 3 = very important). These patients were not different from the full cohort used for item reduction and validation (Table 2). Mean relevance scores were calculated for each item. Items with a mean relevance score of 2.0 or greater in which a minimum of ⅔ (66.7%) of patients rated the item as at least "somewhat relevant" were eligible for inclusion in the HOOS, JR validation. These thresholds were used in a previous validation of the Foot and Ankle Outcome Score [30]. One item, "light domestic duties", was excluded due to a lack of relevance.

Once the relevance survey was completed, we excluded redundant items that measured the same activity in the pain and ADL (or sports and recreation) domains of the HOOS: going up or down stairs, walking on a flat surface, standing upright, and walking on an uneven surface. We assessed the importance of these items using the relevance survey responses and the difficulty of these items using the preoperative responses from the full cohort to determine whether the ADL (or sports and recreation) or pain domain items were dominant. For all items, the four pain items were deemed more relevant and more difficult by patients (Table 1), so the ADL (four) and sports and recreation (one) items were excluded from the HOOS, JR validation, leaving 30 items for assessment.

Statistical Analysis

Item Reduction Process

Before applying the Rasch model, a principal component factor analysis was used to assess the unidimensionality of the 30 items, which means that all items forming the questionnaire measure a single construct or a single dimension. To evaluate the internal validity of the HOOS, JR, Rasch analysis was performed using a partial-credit model [22]. The most basic form of the Rasch model is based on a binary-response scale. The partial-credit model is an extension to the basic Rasch model and is devised for responses in which one has two or more ordered categories. It permits each item to have its own unique number of categories and modeled distance between adjacent categories. Overall fit of the data to the Rasch model was evaluated in three ways: (1) information-weighted and outlier-sensitive mean-square statistics for each item were calculated to test whether there were items that did not fit with the model expectancies. Mean squares greater than 0.8 and less than 1.2 were considered acceptable fit. Items outside this range were considered underfit ($\geq$ 1.2) or overfit or redundant ($\leq$ 0.8) [21]; (2) for the chi-square tests, p values less than 0.05 indicated poor fit of the item to the model; and (3) information-weighted and outlier-sensitive standardized residuals (t-statistics) $\pm$ 2.5 indicate adequate fit [28]. Items outside this range were considered underfit ($>$ 2.5) or overfit ($<$ 2.5). Based on the established item fit parameters, items were removed sequentially and not retained in the subsequent iterative

Table 1. HOOS items with results from importance survey, baseline and 2-year survey, and bootstrapping retention

HOOS items and subdomains	Importance (n = 30) (mean ± SD)	Percent missing at baseline (n = 2371)	Mean difficulty (baseline) (mean ± SD)	SRM (2 years)	Percent retained
Symptoms					
S1. Do you feel grinding, hear clicking, or any other type of noise from your hip?	2.1 ± 0.8	1.5	1.2 ± 1.3	0.7	0.3
S2. Difficulties spreading your legs wide apart	2.7 ± 0.6	2.1	2.1 ± 1.2	1.5	0.0
S3. Difficulties to stride out when walking	2.9 ± 0.3	1.9	2.3 ± 1.0	1.8	1.4
S4. How severe is your hip joint stiffness after first wakening in the morning?	2.6 ± 0.6	0.0	2.2 ± 1.0	1.6	5.1
S5. How severe is your hip stiffness after sitting, lying, or resting later in the day?	2.7 ± 0.5	0.1	2.2 ± 0.9	1.6	1.9
Pain					
P1. How often do you experience hip pain?	2.9 ± 0.3	2.7	3.2 ± 0.7	2.6	3.0
P2. Straightening hip fully	2.5 ± 0.7	1.8	1.9 ± 1.0	1.6	17.9
P3. Bending hip fully	2.7 ± 0.6	3.2	2.3 ± 1.0	1.8	56.6
P4. Walking on a flat surface	2.5 ± 0.5	1.5	2.0 ± 0.9	1.9	64.3
P5. Going up or down stairs	2.7 ± 0.4	2.0	2.3 ± 0.9	1.8	99.1*
P6. At night while in bed	2.7 ± 0.5	1.1	1.8 ± 1.0	1.5	70.2*
P7. Sitting or lying	2.6 ± 0.6	1.4	1.5 ± 0.9	1.3	2.3
P8. Standing upright	2.5 ± 0.7	1.5	1.7 ± 0.9	1.5	31.5
P9. Walking on a hard surface (asphalt, concrete, etc)	2.3 ± 0.6	2.3	2.1 ± 0.9	1.8	94.6*
P10. Walking on an uneven surface	2.6 ± 0.6	2.7	2.4 ± 0.9	1.9	95.6*
Activities of daily living					
A1. Descending stairs	2.6 ± 0.7	1.5	1.8 ± 1.0	1.5	N/A
A2. Ascending stairs	2.7 ± 0.5	1.5	2.2 ± 1.0	1.6	N/A
A3. Rising from sitting	2.8 ± 0.6	0.5	2.1 ± 1.0	1.6	90.2*
A4. Standing	2.7 ± 0.6	0.9	1.7 ± 1.0	1.5	N/A
A5. Bending to floor/pick up an object	2.8 ± 0.5	0.7	2.3 ± 1.0	1.6	89.3*
A6. Walking on a flat surface	2.3 ± 0.7	0.4	1.9 ± 0.9	1.7	N/A
A7. Getting in/out of car	2.6 ± 0.5	0.2	2.3 ± 0.9	1.8	70.6*
A8. Going shopping	2.0 ± 0.8	2.8	2.1 ± 0.9	1.8	21.7
A9. Putting on socks/stockings	2.8 ± 0.4	1.2	2.5 ± 1.0	1.5	56.2
A10. Rising from bed	2.4 ± 0.6	0.2	1.9 ± 0.9	1.6	66.9*
A11. Taking off socks/stockings	2.6 ± 0.6	1.6	2.3 ± 1.1	1.5	28.9
A12. Lying in bed (turning over, maintaining hip position)	2.7 ± 0.6	0.3	2.1 ± 1.0	1.6	95.8*
A13. Getting in/out of bath	2.3 ± 0.8	9.5	1.7 ± 1.0	1.3	97.0*
A14. Sitting	2.3 ± 0.7	0.2	1.4 ± 0.9	1.2	93.2*
A15. Getting on/off toilet	2.4 ± 0.7	0.0	1.7 ± 1.0	1.4	75.5*
A16. Heavy domestic duties (moving heavy boxes, scrubbing floors, etc)	2.2 ± 0.7	11.6	2.6 ± 1.0	1.6	46.6
A17. Light domestic duties (cooking, dusting, etc)	2.0 ± 0.7	3.7	1.6 ± 0.9	1.5	N/A
Sports and recreation					
SP1. Squatting	2.4 ± 0.7	14.4	2.8 ± 1.1	1.5	51.7
SP2. Running	2.3 ± 0.8	23.4	3.2 ± 1.1	1.5	0.0
SP3. Twisting/pivoting on loaded leg	2.7 ± 0.6	11.3	2.9 ± 1.0	1.8	0.0
SP4. Walking on uneven surface	2.5 ± 0.7	6.6	2.5 ± 0.9	1.8	N/A
Quality of life					
Q1. How often are you aware of your hip problem?	2.8 ± 0.5	2.3	3.6 ± 0.6	2.1	N/A

Table 1. continued

HOOS items and subdomains	Importance (n = 30) (mean ± SD)	Percent missing at baseline (n = 2371)	Mean difficulty (baseline) (mean ± SD)	SRM (2 years)	Percent retained
Q2. Have you modified your lifestyle to avoid activities potentially damaging activities to your hip?	2.7 ± 0.5	1.7	2.9 ± 1.0	1.4	N/A
Q3. How much are you troubled with lack of confidence in your hip?	2.6 ± 0.7	2.0	2.7 ± 1.1	1.8	N/A
Q4. In general, how much difficulty do you have with your hip?	2.8 ± 0.5	1.6	2.8 ± 0.8	2.3	N/A

* Greater than 66.7%; HOOS = Hip disability and Osteoarthritis Outcome Score; SRM = standardized response means; N/A = not applicable for items that were not formally assessed using the bootstrapped Rasch analysis.

Table 2. Demographics of relevancy cohort and full cohort

Variable	Relevancy cohort (n = 30)*	Full cohort (n = 2371)¶	p value
Age (years)	59 ± 14	64 ± 11	0.042
BMI (kg/m^2)	27 ± 4	28 ± 5	0.440
Sex	38% female	51% female	0.309

* For HOOS items eligibility assessment 30 consecutive patients scheduled for primary total hip replacement from four surgeons were interviewed; ¶patients who returned the HOOS survey preoperatively, returned a 2-year survey, and completed all items from the HOOS baseline survey were eligible for item reduction and validation.

analysis. Standardized residuals are highly sensitive to sample size and therefore were used only to guide decision-making [32].

To refine the most likely candidate items for removal, we performed bootstrapping of 500 samples of 1800 patients using the full cohort (n = 2371). Bootstrapping is a resampling technique and allows us to estimate the accuracy of our approximation of all patients using only available patients. This was performed with replacement so patients could be selected in each sample more than once. Each bootstrapped sample was run through an automated Rasch modeling algorithm. Items retained in the final Rasch model using the automated exclusion criteria in more than ⅔ of the 500 models were considered in the final Rasch analysis process.

Item response categories also were examined to determine if they produced sequentially ordered thresholds [19]. Differential item functioning is a form of item bias that can occur when different groups in the sample give different responses to an individual item despite equal levels of the underlying trait [15]. Differential item functioning was assessed using the classified differential item functioning categories based on the Mantel-Haenszel statistic [9, 10]. We evaluated differential item functioning by sex, age ($<$ 65, $\geq$ 65 years), BMI ($<$ 30, $\geq$ 30 kg/m^2), and Deyo-Charlson comorbidity index (0, 1–2, 3+).

Final inclusion assessment using the learning cohort consisted of a manual-reduction process using the Rasch modeling and assessment statistics.

Scoring

HOOS, JR scoring was scaled to 100 points just as the original HOOS domains, with 0 representing total hip disability and 100 representing perfect hip health. As with the previous HOOS-PS validation [8], scores for the HOOS, JR were determined using a Rasch-based person score from the validation cohort. A crosswalk table converting raw sum score to the interval level measure scaled from 0 to 100 was provided to facilitate the use and scoring of HOOS, JR (Appendices 1 and 2. Supplemental material is available with the online version of CORR®). The HOOS, JR scores were derived from the responses to full HOOS surveys from both registries.

Validation Process

The final survey underwent a formal validation process in the HSS validation cohort and the FORCE-TJR registry. The internal consistency is a measure of how well the items in the instrument measure the same construct. The internal consistency reliability of the HOOS, JR instrument was evaluated by a Person Separation Index (PSI) [38] that is similar to reliability indices such as Cronbach's alpha. A higher PSI value indicates a stronger ability of the scale to differentiate between patients with various degrees of ability, providing evidence of good internal consistency. A

Person-item Map
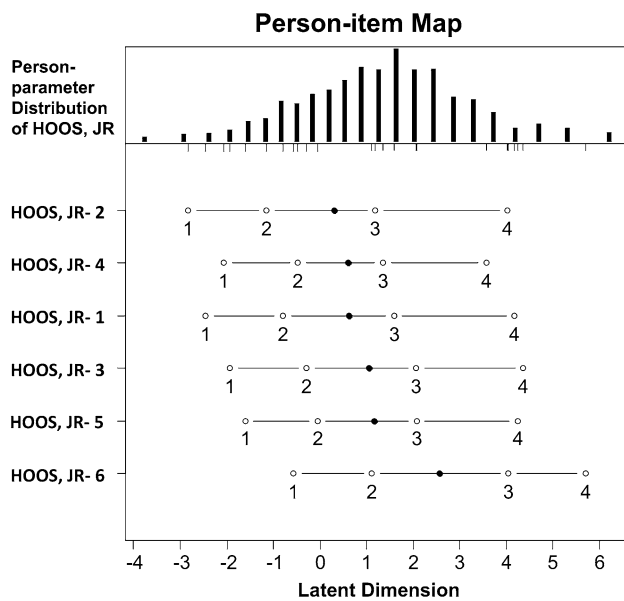
Fig. 1 A map shows person-ability and difficulty for the six items of the HOOS, JR. The horizontal line represents the measure of the variable in linear log units. The bar graph at the top of the figure shows each patient's ability, with ability increasing from right to left. The bottom graph shows each item's relative difficulty for this validation sample, with difficulty increasing from right to left. The numbers represent the thresholds between response categories. For data to adhere to the Rasch model, threshold points are correctly ordered, indicating patients have no difficulty consistently discriminating between response categories. HOOS, JR- 1: (Pain) Going up or down stairs; HOOS, JR- 2: (Pain) Walking on an uneven surface; HOOS, JR- 3: (activities of daily living [ADL]) Rising from sitting; HOOS, JR- 4: (ADL) Bending to floor/pick up an object; HOOS, JR- 5: (ADL) Lying in bed (turning over, maintaining hip position); HOOS, JR- 6: (ADL) Sitting.

PSI value greater than 0.7 was considered acceptable [11]. Residual item correlations were used to assess local independence of the items, that there was no appreciable correlation between the items included in the survey. Items with residual correlations greater than 0.3 are considered to be locally dependent [35]. After the final items were selected, a principal component analysis on the standardized residual was used to verify whether the remaining, selected items measure a one-dimensional construct. In a successful Rasch analysis, the residuals should be uncorrelated and there will be no presence of subdimensions. An eigenvalue of the first residual factor greater than three and an eigenvalue of each item greater than 1.4 suggest that additional subdimensions are likely to be present [20, 33].

Responsiveness of the instrument to changes after total hip replacement was assessed using standardized response means [18] and compared with other validated PROMs (HOOS domains, WOMAC domains) in the HSS validation cohort and FORCE-TJR registry at 2 years after THA. A standardized response mean greater than 0.8 is considered

large [34]. Floor (percent at worst possible score preoperatively) and ceiling (percent at best possible score postoperatively) effects were calculated and compared with other validated instruments. Finally, external construct validity was assessed by comparing the Spearman's correlations between HOOS, JR and the previously validated PROMs. A Spearman's correlation coefficient of 0.8 or greater is considered very high external validity [37]. We used a scatterplot overlying a contour plot based on bivariate kernel density estimation between HOOS, JR and other HOOS domains to visually assess the external correlations. A bandwidth multiplier of one was used for each kernel density estimate. Areas of high density correspond to areas where there are many overlapping points.

This validation assessment was repeated to consider further reduction without information loss (ie, validation measures remain robust in all dimensions even after exclusion of additional items) because two eligible items were measuring similar activities (walking on a hard surface and walking on an uneven surface), and one additional item may not represent a universal activity (getting in/out of bath) because it was the most often skipped question (10% missing in the full HSS cohort).

Factor analyses were performed using SAS® 9.3 (SAS Institute Inc, Cary, NC, USA) and Rasch analysis using the eRm R Package (R Foundation, Vienna, Austria).

The HSS cohort included 2371 patients with hip osteoarthritis, from 31 surgeons, who underwent primary, unilateral THA at HSS between May 2007 and January 2012. These patients had a mean age of 64 ± 11 years, 51% were female, and they had a mean BMI of 28 ± 5 kg/m². The learning and validation cohorts had similar age, sex, and BMI distributions. The FORCE-TJR registry consisted of 910 patients with hip osteoarthritis, from 108 surgeons across 36 practices from 22 US states, undergoing primary unilateral THA between June 2011 and January 2013. These patients had a mean age of 65 ± 11 years, 57% were female, and they had a mean BMI of 29 ± 6 kg/m².

Results

Item Reduction

Item reduction yielded a six-item PROM (HOOS, JR), which retained items only from the pain and ADL domains. Of the 40 items in the full HOOS, four were excluded a priori as part of the HOOS QOL domain; four ADL items and one sports and recreation item were excluded as being redundant, with pain items measuring similar activities. The relevance survey results excluded one additional question before formal item reduction modeling. "Light domestic duties" were not considered relevant by a
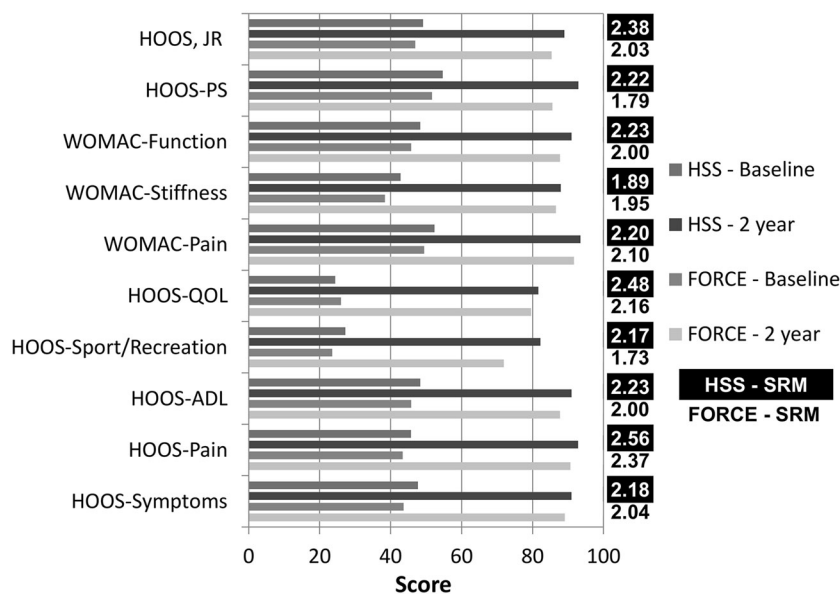
Fig. 2 The standardized response means (SRM) of hip arthroplasty outcomes measures at preoperative baseline and 2 years after surgery are shown. HSS = Hospital for Special Surgery; FORCE = Function and Outcomes Research for Comparative Effectiveness; QOL = quality of lfe; ADL = activities of daily living; HOOS-PS = HOOS Physical Function Short-Form.

majority of respondents (Table 1), leaving 30 items for modeling.

Bootstrapped Rasch models reduced these 30 items to 12 before an iterative manual Rasch modeling process was performed. Excluded items were retained in 0% to 64% of bootstrapped models with only four excluded items exceeding 50% retention (Table 1). Despite our a priori exclusion threshold of 66.7% retention, no item with less than 90% retention was included in the final model. Iterative manual Rasch modeling using the learning cohort resulted in a one-dimensional survey consisting of eight items that were well fit.

Three of these remaining eight items were identified as having questionable properties after further evaluation. Walking on a hard surface and walking on an uneven surface had a residual item correlation of 0.44, suggesting item dependency independent of a person's functional ability. Walking on an uneven surface was considered more relevant and more difficult by patients preoperatively, and therefore was retained in favor of walking on a hard surface. Finally, getting in or out of bath was missing in 10% of the full cohort's surveys, exceeding the combined missingness of the other seven items combined. Getting in or out of bath was also one of the least relevant (34th of 40 HOOS items) and least difficult (34th of 40) activities preoperatively. Therefore, to reiterate, we settled on a final HOOS, JR of six items (Fig. 1). These six items had appropriate and acceptable person-ability and item-difficulty properties with responses correctly ordered for each item in a person's personal hip functional ability (Fig. 1).

There was also consistent spread across responses and distances between responses based on person-ability.

Validation

The HOOS, JR had acceptable internal consistency (PSI, 0.86 [HSS]; and 0.87 [FORCE]). Principal component analysis on the standardized residuals determined that the items all existed in a single dimension. All validation analyses were performed using the HSS validation cohort and FORCE-TJR registry [1, 13].

Responsiveness of the HOOS, JR exceeded the theoretical 0.8 standardized response means threshold and was comparable or favorable against all other hip PROM domains evaluated with standardized response means of 2.03 (95% CI, 1.84–2.22) (FORCE) and 2.38 (95% CI, 2.27–2.49) (HSS) (Fig. 2). Only HOOS-pain (standardized response mean, 2.37 [95% CI, 2.16–2.58] [FORCE]; and 2.56 [95% CI, 2.42–2.70] [HSS]) and HOOS-QOL (standardized response mean, 2.16 [95% CI, 1.97–2.35] [FORCE]; and 2.48 [95% CI, 2.32 – 2.64] [HSS]) had higher standardized response means of scores considered. The floor (0.6%–1.6%) and ceiling (41%–46%) properties of the HOOS, JR were similar to or better than other domains of the HOOS and WOMAC (Fig. 3). External validity was high with the HOOS, JR having very high correlations with HOOS-Pain (0.87, [95% CI, 0.86–0.89] [HSS]; 0.87, [95% CI, 0.84–0.90] [FORCE]), HOOS-ADL/ WOMAC-function (0.94, [95% CI, 0.93–0.95] [HSS]; 0.94

Floor Effect

A

Ceiling Effect

B
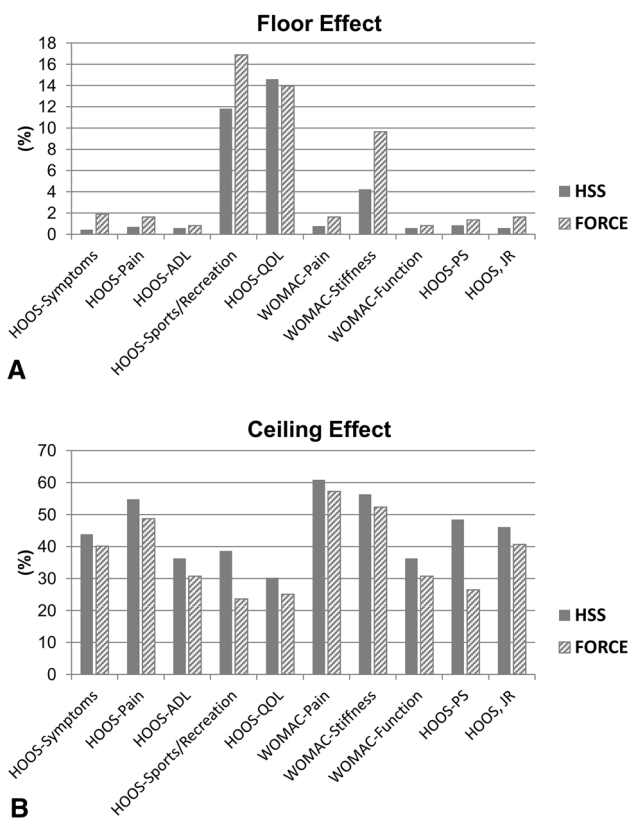
Fig. 3A–B This graph shows the (**A**) floor and (**B**) ceiling effects for 10 patient-reported outcome measures; HOOS-PS = HOOS Physical Function Short-Form; ADL = activities of daily living; QOL = quality of life; HSS = Hospital for Special Surgery; FORCE = Function and Outcomes Research for Comparative Effectiveness.

External Validity of HOOS, JR
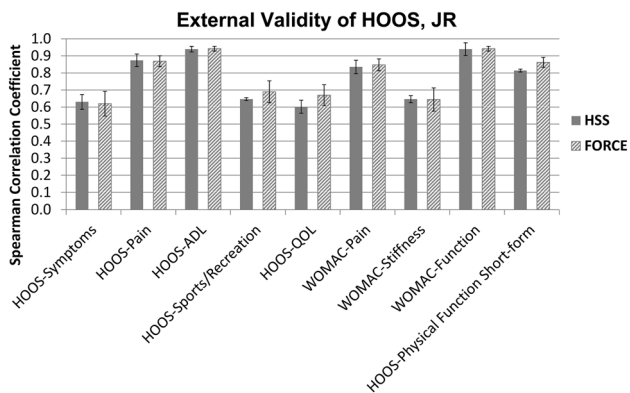
Fig. 4 A comparison of the external validity of the HOOS, JR against nine other patient-reported outcome measures using Spearman's correlation coefficient is shown. HSS = Hospital for Special Surgery; FORCE = Function and Outcomes Research for Comparative Effectiveness; QOL = quality of lfe; ADL = activities of daily living; HOOS-PS = HOOS Physical Function Short-Form.

[95% CI, 0.93–0.96] [FORCE]), WOMAC-pain (0.84, [95% CI, 0.81–0.86] [HSS]; 0.85, [95% CI, 0.81–0.88] [FORCE]), and HOOS-PS (0.81, [95% CI, 0.79–0.84] [HSS]; 0.86, [95% CI, 0.83–0.89] [FORCE]) (Fig. 4). The HOOS, JR also showed high correlations with HOOS-symptoms (0.62, [95% CI, 0.55–0.69] [HSS]; 0.63, [95% CI, 0.59–0.67] [FORCE]), HOOS-sports and recreation (0.65, [95% CI, 0.61–0.68] [HSS]; 0.69, [95% CI, 0.63–0.75] [FORCE]), HOOS-QOL (0.60, [95% CI, 0.56–0.64] [HSS]; 0.67, [95% CI, 0.61–0.73] [FORCE]), and WOMAC-stiffness (0.64, [95% CI, 0.58–0.71] [HSS]; 0.65, [95% CI, 0.61–0.68] [FORCE]). A scatterplot confirmed the very high correlations with pain (Fig. 5) and ADL (Fig. 6) at baseline and 2 years.

Discussion

With a rapid movement toward using PROMs for THA outcomes assessment by the CMS, there was a need for a nonproprietary, reliable, and responsive hip assessment PROM that also was efficient. Therefore, we endeavored to develop a short-form version of the HOOS that was directly relevant to patients undergoing THA. The HOOS, JR is a six-question short-form alternative to the longer HOOS and WOMAC surveys for PROM assessment for patients undergoing THA. We anticipate the HOOS, JR will be self-administered on paper or electronically as that is how patients in the HSS total hip replacement registry and FORCE-TJR completed their HOOS surveys.

Limitations

This study has numerous limitations. Development of the HOOS, JR was done at one tertiary care musculoskeletal specialty hospital in a large urban area. Although the patient population is diverse in socioeconomic status and residential environment (including patients from urban, suburban, and rural regions), most are from urban areas, therefore there may be a bias in the item responses for these patients. The FORCE-TJR cohort was older, more likely to be female, and had higher BMI than the HSS cohort. However, external validation of the HOOS, JR was successful using the FORCE-TJR cohort with geographically diverse patients and surgeon practices, which suggests the HOOS, JR remains robust outside the specialty-care setting. The development and validation were performed in the United States, which may limit the international utility,
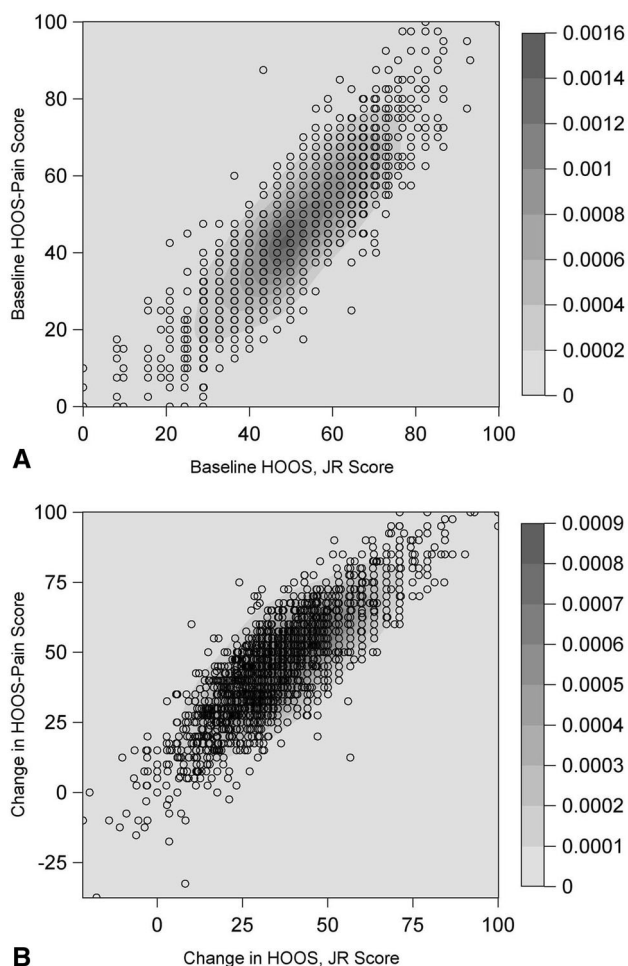
Fig. 5A–B The contour map shows the HOOS-pain domain versus (**A**) HOOS, JR at baseline and (**B**) the change in score from baseline to 2 years after THA. A scatterplot overlays a contour plot based on bivariate kernel density estimation. A bandwidth multiplier of one was used for each kernel density estimate. Areas of high density correspond to areas where there are many overlapping points. The scatterplot shows the positive correlation between the HOOS, JR (x-axis) and the HOOS-pain domain (y-axis) at baseline and the change between baseline and 2-year followup.

although the resulting items are universal movements or hip positions.

Although we know that 81% of patients who underwent THAs were accounted for at 2 years, and patients who are lost to followup may have had inferior health status compared with those with complete followup, this may have limited the number of patients in this study with lower HOOS, JR scores, and so may have to some degree limited our ability to assess the performance of this outcomes instrument in the lower ranges of patient function. However, we believe this is not a serious limitation, because the original HOOS was developed for assessment of the full range of hip conditions and many of the items eliminated in

our reduction process were those most often skipped by patients with lower function who did return surveys, leaving only activites or movements that patients should be expected to be able to perform after THA.

Unfortunately, given the pragmatic nature of the validation, we were unable to compare it with the Oxford Hip Score or other validated hip-specific PROMs not originally collected in the HSS or FORCE-TJR registries. Given the popularity of the Oxford Hip Score, cross-validation and development of a crosswalk between these two short-form PROMs should be done. We also validated the survey only in patients with a diagnosis of osteoarthritis who had unilateral primary THA. We plan to perform future validation for other surgical indications (such as rheumatoid arthritis, femoral neck fracture), bilateral THAs, and alternative hip replacement surgery (hip resurfacing, partial hip replacement).

Another limitation pertained to the retrospective study design. Our study was done as a pragmatic validation process using existing full HOOS surveys to complete a new short-form survey rather than validating the survey in a new cohort of patients. Patients were not administered the full and the short-form surveys for comparison. Rather, the HOOS, JR was derived from the full HOOS. Because item order theoretically is related to responses, it is possible that responses to HOOS, JR items were influenced by HOOS items not included in the HOOS, JR. However, this pragmatic process does lend itself one theoretical advantage. Because the validated HOOS, JR was derived from the full HOOS, it can be calculated for other HOOS respondents for direct comparison. Because five of the six final HOOS, JR items are included in the WOMAC, the HOOS, JR possibly could be calculated from patients with existing WOMAC scores, allowing for direct comparisons between HOOS, JR and WOMAC patient responses in different cohorts. We plan to develop a crosswalk between these surveys as part of our future work. Current work at our institution includes validating administration of the HOOS, JR at more frequent times through mobile devices to gain a clearer understanding of how these instruments work at the individual patient level and during the early postoperative recovery period. This flexibility should allow hospitals or clinics to administer the surveys in their preferred fashion.

A final limitation is that the full HOOS and WOMAC generate domain-specific subscores for pain, function (ADL), and hip symptoms, but the HOOS, JR does not; so although these long PROMs can be transformed to the overall hip health measure, the scores are not directly comparable to those of the HOOS, JR. With time, adoption of the HOOS, JR will allow for "calibration" of what the hip health (HOOS, JR) score represents for pain and hip disease. In aggregate across all patients of a specific
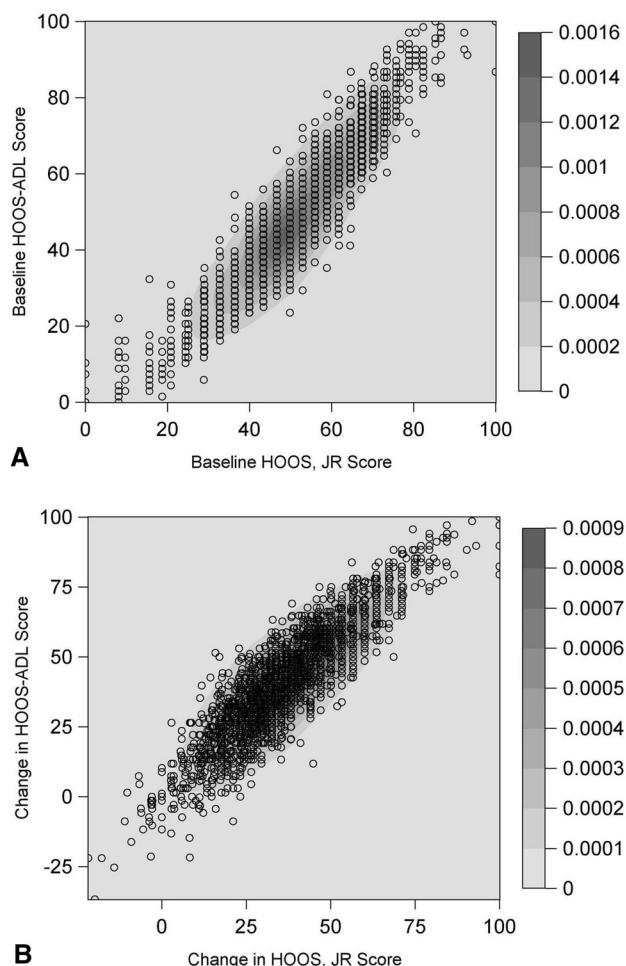
Fig. 6A–B The contour map shows the HOOS-ADL domain versus (**A**) HOOS, JR at baseline and (**B**) change in score from baseline to 2 years after THA. A scatterplot overlays a contour plot based on bivariate kernel density estimation. A bandwidth multiplier of one was used for each kernel density estimate. Areas of high density correspond to areas where there are many overlapping points. The scatterplot shows the positive correlation between the HOOS, JR (x-axis) and the HOOS-ADL domain (y-axis) at baseline and the change between baseline and 2-year followup. ADL = activities of daily living;

surgeon or hospital, the before and after changes in HOOS, JR scores after THA will capture improved hip health.

Validity

Content validity and test-retest reliability was assessed previously for these items through the work of the original HOOS development team [26]. Nevertheless, we assessed relevance to patients undergoing THA specifically and found one question that was not considered relevant by our patients: light domestic duties. This may reflect the daily

activities of older American adults, who spend more than three times as much of their waking hours in leisure activities than in doing household responsibilities (7 hours versus 2 hours, Bureau of Labor Statistics, 2013 American Time Use Survey) [5]. The HOOS, JR held together as a single construct, which we define as "hip health" because it combines aspects of pain and ADL (no HOOS symptoms or HOOS sports and recreation items were retained) movements or activities that are directly relevant and difficult for patients with advanced hip osteoarthritis.

External construct validity also was seen, with the HOOS, JR having high correlations with the pain and ADL domains of the HOOS and the pain and function domains of the WOMAC, whereas moderate correlations were seen for other HOOS and WOMAC domains. This was true in an internal HSS validation cohort and in a nationally representative THA registry comprised of 108 surgeons in 37 practice settings (73% in community-based practices) across 22 US states.

Responsiveness

The items included on the HOOS, JR are relevant to patients with hip osteoarthritis and difficult for these patients to perform before undergoing THA. We found that responsiveness for this instrument is high relative to hip PROMs that were developed for individuals with less-severe hip disability. This was true for the HSS validation cohort and the FORCE-TJR registry. The theoretical and practical advantages of higher responsiveness are that fewer subjects are needed to adequately power outcomes studies using highly responsive instruments [26].

Conclusions

Given the rapid move toward pay-for-performance outcomes reporting for the CMS, the HOOS, JR could be an efficient and responsive alternative patient-relevant survey for hospitals and surgeons to comply with coming regulations. The HOOS, JR is an efficient alternative to traditional outcomes surveys and could be used for clinical outcomes assessment or as a research tool to assess group-level outcomes. There is still a place for the full surveys to examine the various facets (domains) of hip health in more detailed research projects or to assess individual patient symptoms. However, given the increasing demand for comparative outcomes data, the HOOS, JR offers an efficient, pragmatic solution that has been validated in a large tertiary care specialty hospital and more broadly in a nationally representative sample of US community-based practices.

References

1. Ayers DC, Franklin PD. Joint replacement registries in the United States: a new paradigm. *J Bone Joint Surg Am.* 2014;96:1567–1569.

2. Ayers DC, Zheng H, Franklin PD. Integrating patient-reported outcomes into orthopaedic clinical practice: proof of concept from FORCE-TJR *Clin Orthop Relat Res.* 2013;471:3419–3425.

3. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. *J. Rheumatol.* 2002;29:2473–2476.

4. Black N. Patient reported outcome measures could help transform healthcare. *BMJ.* 2013;346:f167.

5. Bureau of Labor Statistics. American Time Survey. Available at: http://www.bls.gov/tus/. Accessed January 12, 2014.

6. CMS.gov. Centers for Medicare & Medicaid Services. Measure Methodology. Available at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed October 6, 2015.

7. Concannon TW. Can patient centered outcomes research improve healthcare? *BMJ.* 2015;351:h3859.

8. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Lohmander LS. The development of a short measure of physical function for hip OA: HOOS-Physical Function Short-form (HOOS-PS) – an OARSI/OMERACT initiative. *Osteoarthritis Cartilage.* 2008;16:551–559.

9. Dorans NJ, Schmitt AP. *Constructed Response and Differential Item Functioning: A Pragmatic Approach.* ETS Research Report RR-91-47. Princeton, NJ, USA: Educational Testing Service; 1991.

10. Dorans, NJ, Schmitt, AP, Bleistein CA. The standardization approach to assessing comprehensive differential item functioning. *J Educ Meas.* 1992;29:309–319.

11. Fisher WP Jr. Reliability, separation, strata statistics. *Rasch Measurement Transactions.* 1992;6:238.

12. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998;2:i-iv, 1–74.

13. Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: a patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA.* 2012;308:1217–1218.

14. Franklin PD, Harrold L, Ayers DC. Incorporating patient-reported outcomes in total joint arthroplasty registries: challenges and opportunities. *Clin Orthop Relat Res.* 2013;471:3482–3488.

15. Holland PW, Wainer H, eds. *Differential Item Functioning.* Hillsdale, NJ, USA: Lawrence Erlbaum Associates; 1993.

16. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties: a comparison between the Harris score and the Oxford hip score. *J Arthroplasty.* 2005;20:1037–1041.

17. Kalantar JS, Talley NJ. The effects of lottery incentive and length of questionnaire on health survey response rates: a randomized study. *J Clin Epidemiol.* 1999;52:1117–1122.

18. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28:632–642.

19. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3:85–106.

20. Linacre JM. A User's Guide to Winsteps-Ministep: Rasch-model Computer Programs. Program Manual 3.68.0. Available at: http://www.winsteps.com/manuals.htm. Accessed December 1, 2015.

21. Linacre JM, Wright BD. A User's Guide to Bigsteps. Winsteps.com. Available at: http://www.winsteps.com/a/bigsteps.pdf. Accessed December 1, 2015.

22. Masters GN. A Rasch model for partial credit scoring. *Psychometrika.* 1982;47:149–174.

23. Matharu GS, McBryde CW, Robb CA, Pynsent PB. An analysis of Oxford hip and knee scores following primary hip and knee replacement performed at a specialist centre. *Bone Joint J.* 2014;96:928–935.

24. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ.* 2015;350:g7818.

25. Nilsdotter A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopaedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res.* 2011;63(suppl 11):S200–207.

26. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4:10.

27. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, Dhert WJ. Patient-reported outcome in total hip replacement: a comparison of five instruments of health status. *J Bone Joint Surg Br.* 2004;86:801–808.

28. Pallant JF, Miller RL, Tennant A. Evaluation of the Edinburgh Post Natal Depression Scale using Rasch analysis. *BMC Psychiatry.* 2006;6:28.

29. Rahman WA, Greidanus NV, Siegmeth A, Masri BA, Duncan CP, Garbuz DS. Patients report improvement in quality of life and satisfaction after hip resurfacing arthroplasty. *Clin Orthop Relat Res.* 2013;471:444–453.

30. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int.* 2001;22:788–794.

31. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health.* 2009;12:1075–1083.

32. Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodol.* 2008;8:33.

33. Smith RM, Miao CY. Assessing unidimensionality for Rasch measurement. In: Wilson M, ed. *Objective Measurement: Theory Into Practice.* Vol 2. Norwood, NJ: Ablex Publishing Corp; 1994:316–328.

34. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* 3rd ed. Oxford, UK: Oxford University Press; 2003.

35. Wainer H, Kiely GL Item clusters and computerized adaptive testing: a case for testlets. *J Educ Meas.* 1987;24:185–201.

36. Walker JA. Total hip replacement: improving patients' quality of life. *Nurs Stand.* 2010;24:51–57; quiz 58.

37. Wechsler S. *Statistics at Square One.* 9th ed. London, UK: BMJ Publishing Group; 1996.

38. Wright BD, Stone MH. *Measurement Essentials.* 2nd ed. Wilmington, DE: Wide Range Inc; 1999. Available at: http://www.rasch.org/measess/me-all.pdf. Accessed January 4, 2016.