2019

Psychometric properties of patient-reported outcome measures for periacetabular osteotomy

Marcin K. Wasko
Washington University School of Medicine in St. Louis

Elizabeth L. Yanik
Washington University School of Medicine in St. Louis

Cecilia Pascual-Garrido
Washington University School of Medicine in St. Louis

John C. Clohisy
Washington University School of Medicine in St. Louis

Recommended Citation

Wasko, Marcin K.; Yanik, Elizabeth L.; Pascual-Garrido, Cecilia; and Clohisy, John C., ,"Psychometric properties of patient-reported outcome measures for periacetabular osteotomy." The Journal of Bone and Joint Surgery.101,6. e21. (2019).
https://digitalcommons.wustl.edu/open_access_pubs/7585

Psychometric Properties of Patient-Reported Outcome Measures for Periacetabular Osteotomy

Marcin K. Wasko, MD, PhD, Elizabeth L. Yanik, PhD, ScM, Cecilia Pascual-Garrido, MD, and John C. Clohisy, MD

Investigation performed at the Department of Orthopaedic Surgery, Washington University School of Medicine, Barnes-Jewish Hospital, St. Louis, Missouri

Background: Appropriate patient-reported outcome measures are paramount to determine the clinical relevance of change experienced by patients after a surgical procedure. The purpose of this study was to evaluate the psychometric properties of 3 patient-reported outcome measures (Hip disability and Osteoarthritis Outcome Score [HOOS], modified Harris hip score, and Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC]) in a large population of patients treated with periacetabular osteotomy for symptomatic developmental dysplasia of the hip.

Methods: Patients who underwent a periacetabular osteotomy for acetabular dysplasia between October 2011 and October 2016 completed multiple questionnaires preoperatively and at the 1-year follow-up. Internal consistency for subscores was evaluated with the Cronbach alpha. Validity was assessed against the Short Form-12. Interpretability included the calculation of the distribution of scores, floor and ceiling effects, and the minimal clinically important difference. Responsiveness was assessed by correlating the score changes with the Global Perceived Effect score.

Results: Of 294 patients, 246 (84%) were female. The median age was 21 years (interquartile range, 17 to 29 years). All WOMAC and HOOS subscores demonstrated adequate internal consistency, and none of the modified Harris hip score components did. All scores except the WOMAC stiffness score exhibited adequate validity and interpretability, with no floor and ceiling effects over 15%. For the HOOS subscores, the minimal clinically important difference was 10.3 for pain, 10.2 for symptoms, 12.6 for sports and recreation, 11.2 for quality of life, and 10.8 for activities of daily living. The minimal clinically important difference for the modified Harris hip score was 7.4. For the WOMAC subscores, the minimal clinically important difference was 10.8 for pain, 12.9 for stiffness, 10.8 for physical, and 10.4 for total. All scores demonstrated adequate responsiveness.

Conclusions: The HOOS, WOMAC, and modified Harris hip score have adequate psychometric properties for use in patients undergoing periacetabular osteotomy, with minor shortcomings. Among the scores analyzed, the HOOS appears to be the most appropriate measure of patient-reported outcome in patients undergoing periacetabular osteotomy.

Clinical Relevance: By establishing the psychometric properties of patient-reported outcome measures for periacetabular osteotomy, this study enables a more informed choice of measures for clinical practice and research. The estimated minimal clinically important difference values will facilitate understanding the importance of change in patient-reported outcomes after a surgical procedure in a clinical setting and sample size calculation for further studies.

Increasing numbers of studies indicate that periacetabular osteotomy is an effective intervention for improving pain and physical function in patients with developmental dysplasia of the hip[1-3]. Traditionally, the effectiveness of this surgical procedure was expressed in hard end points such as progression to osteoarthritis or conversion to total hip arthroplasty[4-6], yet improvements in functional level or quality of life may matter more to the patient than those end points. Thus, the selection of appropriate outcomes to measure subjective patient improvement remains an essential challenge.

Recognizing the importance of subjective patient improvement, currently, the most commonly used research instruments in patients undergoing periacetabular osteotomy are patient-reported outcome measures[7]. They rate the patient's

Disclosure: One author of this study (M.K.W.) received partial funding through a grant from The Kosciuszko Foundation and also received a grant from the Curing Hip Disease Fund for research personnel salary support. On the **Disclosure of Potential Conflicts of Interest** forms, *which are provided with the online version of the article*, one or more of the authors checked "yes" to indicate that the author had a relevant financial relationship in the biomedical arena outside the submitted work (http://links.lww.com/JBJS/F83).

quality of life on a numerical scale, and differences in scores can be used to evaluate the effectiveness of therapies[8,9], without observer (surgeon) bias[10,11].

Multiple studies have shown clinical outcomes of periacetabular osteotomy for developmental dysplasia of the hip utilizing patient-reported outcome measures, without estimating their psychometric properties[4,5,12]. Psychometric properties are useful for understanding the value of patient-reported outcome measures in a specific population. Only a few articles have described psychometric properties of patient-reported outcome measures in hip preservation procedures; most have focused on hip arthroscopy[9,13,14], with only 1 study on psychometric properties of patient-reported outcome measures in a periacetabular osteotomy population[15].

The principal aim of this study was to evaluate the psychometric properties of 3 patient-reported outcome measures in patients after periacetabular osteotomy. We hypothesized that the Hip disability and Osteoarthritis Outcome Score (HOOS), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and the modified Harris hip score would demonstrate appropriate internal consistency, good convergent and discriminant validity, minimal floor and ceiling effects, and appropriate responsiveness. The secondary aim of this study was to calculate minimal clinically important differences for those scores.

Materials and Methods

Study Design

This single-center, prospective, observational cohort study was approved by the institutional review board. From October 1, 2011, to October 31, 2016, a senior surgeon prospectively enrolled 423 patients with a clinical and radiographic diagnosis of acetabular dysplasia[16].

Participants

Consecutive patients who were 18 to 40 years of age and had undergone periacetabular osteotomy for the treatment of acetabular dysplasia and returned for 1-year follow-up were invited to participate. There were 302 patients who fulfilled those criteria, and all agreed to participate. Patients were excluded if they had undergone periacetabular osteotomy for diagnoses other than developmental dysplasia of the hip or if they presented with other lower-limb injuries at any time point. That left 294 patients (294 hips) who formed the study group.

Procedure

Patient demographic characteristics were documented, physical parameters of height and weight were measured by clinical staff, and body mass index (BMI) was calculated. Patients completed a comprehensive series of questionnaires before and 1 year after the surgical procedure.

Hip-Specific Patient-Reported Outcome Measures

The patient-reported outcome measures included the modified Harris hip score[17] to assess pain and function; the HOOS, with 5 subscores to evaluate pain, symptoms, function, sports participation, and quality of life[18]; and WOMAC subscores, which were calculated from the corresponding HOOS items of pain (questions P4 to P8), stiffness (S4 to S5), and function (A1 to A17) and were normalized to 0 to 100. All these patient-reported outcome measures have a highest possible score of 100 (best result) and a lowest possible score of 0 (worst result).

TABLE I Characteristics of the Study Group (N = 294 Hips), Satisfaction with Surgery, and Global Perceived Effect

Baseline Characteristic	Study Group (N = 294)
Age* *(yr)*	21 (17 to 29)
Female patients†	246 (84%)
Height* *(cm)*	167 (162 to 173)
Weight* *(kg)*	63.5 (56.7 to 74.8)
BMI* *(kg/m²)*	23.4 (20.8 to 26.3)
UCLA‡ activity score§ *(points)*	6 (4 to 10)
Previous surgery†	54 (18%)
Hip arthroscopy	33 (11%)
Childhood pelvic or femoral osteotomy	15 (5%)
Other	6 (2%)
Competitive sport participation†	
Baseball or softball	9 (3%)
Basketball	12 (4%)
Cheerleading or dancing	20 (7%)
Cycling	5 (2%)
Running	9 (3%)
Soccer	16 (5%)
Swimming	9 (3%)
Track	7 (2%)
Volleyball	11 (4%)
Other	17 (6%)
Recreational sports	185 (63%)
Satisfaction with surgery†	
Extremely satisfied	123 (42%)
Very satisfied	81 (28%)
Satisfied	34 (12%)
Somewhat satisfied	25 (9%)
Not satisfied	10 (3%)
No answer provided	21 (7%)
Global Perceived Effect†	
Better	200 (68%)
Same	78 (27%)
Worse	16 (5%)

*The values are given as the median, with the interquartile range in parentheses. †The values are given as the number of patients, with the percentage in parentheses. ‡UCLA = University of California Los Angeles. §The values are given as the mean, with the range in parentheses.

TABLE II Distribution of Scores and the Presence of Floor and Ceiling Effects in the Study Group (N = 294 Hips)

Patient-Reported Outcome Measure	Baseline *(points)*			Floor Effect	Ceiling Effect	1-Year Follow-up *(points)*		
	Mean*	Median†	Range			Mean*	Median†	Range
HOOS								
Pain	54.4 ± 20.7	52.5 (40 to 67.5)	0 to 100	<1%	1%	83.4 ± 17.5	90.0 (75 to 97.5)	7.5 to 100
Symptoms	53.8 ± 20.5	55.0 (40 to 70)	0 to 100	<1%	2%	77.6 ± 18.5	80.0 (65 to 92.5)	15 to 100
Activities of daily living	66.6 ± 21.6	67.7 (52.9 to 86)	1.5 to 100	0	5%	89.7 ± 14.4	96.3 (83.8 to 100)	22.1 to 100
Sport and recreation	40.9 ± 25.3	37.5 (25 to 59.4)	0 to 100	5%	2%	75.4 ± 23.3	81.3 (62.5 to 93.8)	0 to 100
Quality of life	32.1 ± 22.4	31.3 (12.5 to 50)	0 to 100	10%	<1%	65.7 ± 25.8	68.8 (50 to 87.5)	0 to 100
Modified Harris hip score	60.0 ± 14.8	60.5 (49.5 to 68.2)	9.9 to 100	0	1%	82.9 ± 16.5	84.7 (71.5 to 95.7)	26.4 to 100
WOMAC								
Pain	59.0 ± 21.6	60.0 (45 to 75)	0 to 100	<1%	4%	86.4 ± 16.9	92.5 (80 to 100)	5 to 100
Stiffness	56.7 ± 25.8	56.3 (37.5 to 75)	0 to 100	3%	11%	79.2 ± 20.8	87.5 (62.5 to 100)	0 to 100
Physical	66.6 ± 21.6	67.7 (52.9 to 86)	1.5 to 100	0	5%	89.7 ± 14.4	96.3 (83.8 to 100)	22.1 to 100
Total	64.2 ± 20.8	65.6 (50 to 80.2)	3.1 to 100	0	2%	88.1 ± 14.8	93.8 (82.3 to 99)	19.8 to 100

*The values are given as the mean and the standard deviation. †The values are given as the median, with the interquartile range in parentheses.

Other Patient-Reported Measures

The Short Form-12 Health Survey version 2 (SF-12) was used to measure overall health[19]. The Patient Analysis and Tracking System (Axis Clinical Software) was used to score the Physical Component Summary (PCS) and Mental Component Summary (MCS) measures. This survey has a highest possible score of 100 (best result) and a lowest possible score of 0 (worst result). At the time of follow-up, patients self-reported their overall satisfaction with the surgical procedure and their overall change in health (Global Perceived Effect). Satisfaction was measured with the question: "What is your satisfaction level with your surgery?" The 5-item Likert-style response options included "extremely satisfied," "very satisfied," "satisfied," "somewhat satisfied," and "unsatisfied." The Global Perceived Effect was scored with the question: "Compared with my general health before the surgery, my health today is:" with 3 Likert-style response options, which included "better," "same," and "worse."

Psychometric Properties and Statistical Analyses

All statistical analyses were performed using MedCalc version 18.5 software (MedCalc), and significance was determined using an alpha level of 0.05.

Internal Consistency

The items in a patient-reported outcome measure that measure the same concept should be positively correlated with one another (display internal consistency). A useful correlation coefficient for assessing internal consistency is the Cronbach alpha[20]. The Cronbach alpha was calculated for each subscore of the HOOS and WOMAC and for the function and gait components of the modified Harris hip score. Values of α above 0.70 were taken as an indication of sufficient homogeneity[20].

Validity

Validity establishes the ability of an instrument to measure an abstract concept[21]. To evaluate convergent and discriminant validity, we calculated Spearman correlation coefficients (ρ) between patient-reported outcome measures and the SF-12 PCS and MCS scores. The 0.4 threshold was used to determine whether measures were positively correlated, as previously described[13,14]. We hypothesized that the patient-reported outcome measures should correlate more strongly with the SF-12 PCS score than the SF-12 MCS score[8].

Interpretability (Minimal Clinically Important Difference)

The concept of the minimal clinically important difference was proposed to express the importance of change in patient-reported outcome measures[22]. In contrast to significance, the

TABLE III Internal Consistency (Reliability) of Patient-Reported Outcome Measures (N = 294)*

Patient-Reported Outcome Measure	Cronbach Alpha
HOOS	
Pain	0.94† (0.93 to 0.95)
Symptoms	0.79† (0.74 to 0.82)
Activities of daily living	0.97† (0.96 to 0.97)
Sports and recreation	0.90† (0.88 to 0.92)
Quality of life	0.84† (0.81 to 0.87)
Modified Harris hip score	
Gait	0.60 (0.52 to 0.68)
Activities	0.33 (0.20 to 0.45)
WOMAC	
Pain	0.89† (0.86 to 0.91)
Stiffness	0.84† (0.79 to 0.87)
Physical	0.97† (0.96 to 0.97)

*The values are given as the Cronbach alpha, with the 95% CI in parentheses. All variables achieved significant correlation with a 2-tailed test with a p <0.05. †Alpha > 0.7.

TABLE IV Validity of Patient-Reported Outcome Measures Compared with SF-12 (N = 294 Hips)*

Patient-Reported Outcome Measure	SF-12 Physical Subscale	SF-12 Mental Subscale
HOOS		
Pain	0.55† (0.46 to 0.62)	0.22 (0.11 to 0.33)
Symptoms	0.42† (0.33 to 0.51)	0.19 (0.08 to 0.29)
Activities of daily living	0.58† (0.49 to 0.65)	0.18 (0.07 to 0.29)
Sports and recreation	0.62† (0.54 to 0.68)	0.16 (0.04 to 0.26)
Quality of life	0.57† (0.48 to 0.64)	0.30 (0.19 to 0.40)
Modified Harris hip score	0.62† (0.54 to 0.68)	0.18 (0.07 to 0.29)
WOMAC		
Pain	0.53† (0.44 to 0.60)	0.21 (0.09 to 0.32)
Stiffness	0.38 (0.27 to 0.47)	0.16 (0.05 to 0.27)
Physical	0.58† (0.49 to 0.65)	0.18 (0.07 to 0.29)
Total	0.57† (0.49 to 0.65)	0.20 (0.09 to 0.31)

*The values are given as the mean correlation ρ, with the 95% CI in parentheses. †ρ > 0.4.

minimal clinically important difference represents the smallest change in patient-reported outcome measures that is important to the patients[8]. It has been commonly used for measuring outcomes after surgical interventions, and many methods for estimating the minimal clinically important difference have been proposed[23].

In this study, a distribution-based approach, which focused on the statistical characteristics of a patient sample, was used for a minimal clinically important difference calculation. We considered half a standard deviation of the baseline score to be a minimal clinically important difference based on the method established by Norman et al.[24]. The results from this method have been shown to be consistent with other minimal clinically important difference calculation approaches by other research groups[23-25]. First, the mean baseline scores and standard deviations were calculated for the study group. After calculating the minimal clinically important difference (as described above), we looked at the 1-year follow-up scores. If they were >1 minimal clinically important difference greater than baseline, we counted that as meaningful improvement. If they were >1 minimal clinically important difference less than baseline, we counted that as meaningful deterioration. After comparing the follow-up scores with values equal to the baseline plus or minus the minimal clinically important difference, the proportions of patients with an important benefit or deterioration and with an unimportant benefit or deterioration were calculated.

Responsiveness

Responsiveness represents the ability of a patient-reported outcome measure to detect change in a patient's condition over time[21]. We hypothesized that the change in patient-reported outcome measures between the preoperative administration and the 12-month administration would correlate with the Global Perceived Effect score and that the correlation would be at least 0.4 for all subscores. Furthermore, we hypothesized that the standardized response mean and effect size would be higher for patients who reported their condition to be better than for patients reporting no change or worsening on the Global Perceived Effect score. By convention, an effect size of >0.8 is considered large[15]. The standardized response mean was calculated by dividing the mean score change by the standard deviation of that score change[26]. The effect size was calculated as the score change divided by the baseline standard deviation[27].

TABLE V Responsiveness of Patient-Reported Outcome Measures for Periarticular Osteotomy (N = 294 Hips)*

			Participant Global Perceived Effect Responses*					
			"Better" (N = 200)		"Unchanged" (N = 78)		"Worse" (N = 16)	
Patient-Reported Outcome Measure	No. of Patients Available for Score or Subscore	Global Perceived Effect Score Correlation, ρ*	Effect Size	Standardized Response Mean	Effect Size	Standardized Response Mean	Effect Size	Standardized Response Mean
HOOS								
Pain	289	0.35† (0.25 to 0.45)	2.0 (1.7 to 2.2)	1.7 (1.4 to 1.9)	1.1 (0.8 to 1.3)	0.9 (0.7 to 1.1)	0.3 (−0.3 to 0.8)	0.3 (−0.3 to 0.8)
Symptoms	293	0.32† (0.21 to 0.42)	1.6 (1.4 to 1.8)	1.3 (1.1 to 1.5)	0.8 (0.5 to 1.1)	0.7 (0.5 to 1.0)	0.0 (−0.5 to 0.5)	0.0 (−0.6 to 0.5)
Activities of daily living	287	0.34† (0.24 to 0.44)	1.7 (1.5 to 1.9)	1.4 (1.2 to 1.6)	0.8 (0.6 to 1.1)	0.7 (0.5 to 0.9)	0.1 (−0.3 to 0.5)	0.2 (−0.4 to 0.7)
Sport and recreation	284	0.34† (0.23 to 0.44)	1.8 (1.6 to 2.1)	1.5 (1.3 to 1.7)	1.0 (0.7 to 1.3)	0.9 (0.6 to 1.1)	0.1 (−0.3 to 0.6)	0.1 (−0.4 to 0.7)
Quality of life	289	0.34† (0.24 to 0.44)	1.8 (1.6 to 2.0)	1.5 (1.3 to 1.7)	0.9 (0.7 to 1.2)	0.9 (0.7 to 1.1)	0.1 (−0.6 to 0.6)	0.1 (−0.5 to 0.6)
Modified Harris hip score	293	0.36† (0.26 to 0.46)	2.0 (1.7 to 2.3)	1.6 (1.4 to 1.9)	0.9 (0.7 to 1.3)	0.8 (0.6 to 1.1)	0.0 (−0.7 to 0.6)	0.0 (−0.6 to 0.5)
WOMAC								
Pain	288	0.30 (0.20 to 0.41)	1.8 (1.6 to 2.1)	1.6 (1.4 to 1.8)	1.0 (0.7 to 1.3)	0.8 (0.6 to 1.0)	0.4 (−0.2 to 1.0)	0.4 (−0.2 to 1.0)
Stiffness	288	0.27 (0.16 to 0.37)	1.2 (1.0 to 1.4)	1.0 (0.8 to 1.2)	0.7 (0.4 to 0.9)	0.6 (0.4 to 0.8)	−0.1 (−0.5 to 0.3)	−0.1 (−0.7 to 0.4)
Physical	285	0.34† (0.24 to 0.44)	1.7 (1.5 to 1.9)	1.4 (1.2 to 1.6)	0.8 (0.6 to 1.1)	0.7 (0.5 to 0.9)	0.1 (−0.3 to 0.5)	0.2 (−0.4 to 0.7)
Total	284	0.35† (0.25 to 0.45)	1.7 (1.5 to 1.9)	1.5 (1.3 to 1.7)	0.9 (0.6 to 1.2)	0.8 (0.6 to 1.0)	0.2 (−0.3 to 0.6)	0.2 (−0.4 to 0.7)

*The values are given as the mean, with the 95% CI in parentheses. †ρ ≥ 0.4.

TABLE VI Minimal Clinically Important Difference for Patient-Reported Outcome Measures for Periacetabular Osteotomy

Patient-Reported Outcome Measure	Minimal Clinically Important Difference
HOOS	
Pain	10.3
Symptoms	10.2
Activities of daily living	10.8
Sport and recreation	12.6
Quality of life	11.2
Modified Harris hip score	7.4
WOMAC	
Pain	10.8
Stiffness	12.9
Physical	10.8
Total	10.4

Floor and Ceiling Effects

The floor effect describes the inability of the patient-reported outcome measure to demonstrate deterioration after treatment, because the baseline score is already at its lowest point. Similarly, the ceiling effect describes the inability of the patient-reported outcome measure to register improvement, because the baseline score is already at its highest point[18]. The presence of both of the effects is usually established when at least 15% of the population scores either the lowest or the highest possible value[28].

Results

Baseline Participant Characteristics

In the study, 294 patients were included. Of those, 246 (84%) were female. The median age at the time of the surgical procedure was 21 years (interquartile range, 17 to 29 years). The median BMI was 23.4 kg/m^2 (interquartile range, 20.8 to 26.3 kg/m^2), and 63% of the patients had normal weight. Table I presents detailed patient characteristics, satisfaction with the surgical procedure, and the Global Perceived Effect. The score distribution is provided in Table II.

Internal Consistency

All WOMAC and HOOS subscores demonstrated adequate internal consistency. The Cronbach alpha values for the HOOS ranged from 0.79 to 0.97 (Table III). The modified Harris hip score components did not demonstrate adequate reliability for use in patients who underwent periacetabular osteotomy (Cronbach alpha subscores of 0.60 for gait and 0.33 for activities of daily living). Thus, our hypothesis of adequate internal consistency for use in patients undergoing periacetabular osteotomy was confirmed only for the HOOS and WOMAC.

External Validity

Although all of the patient-reported outcome measures correlated poorly with the SF-12 MCS score ($\rho < 0.4$), they correlated more strongly with the SF-12 PCS score: $\rho = 0.62$ for the modified Harris hip score, $\rho > 0.4$ for all HOOS components (from 0.42 for symptoms to 0.62 for sports and recreation), and $\rho > 0.5$ for all WOMAC components except stiffness ($\rho = 0.38$). These findings (summarized in Table IV) supported our a priori hypotheses.

Responsiveness

All 3 patient-reported outcome measures displayed adequate responsiveness based on effect size and standardized response mean calculations (Table V). Although none of the patient-reported outcome measures correlated with the Global Perceived Effect score with $\rho \geq 0.4$, almost all of the 95% confidence intervals (CIs) extended above 0.4.

TABLE VII The Study Group (N = 294 Hips) Stratified According to the Level of Change (Based on Minimal Clinically Important Difference) Achieved in Different Patient-Reported Outcome Measures

Patient-Reported Outcome Measure	Important Improvement	Unimportant Improvement	Remained the Same	Unimportant Deterioration	Important Deterioration	Total
HOOS						
Pain	220 (76%)	37 (13%)	6 (2%)	18 (6%)	9 (3%)	290
Symptoms	189 (67%)	40 (14%)	21 (7%)	14 (5%)	20 (7%)	284
Activities of daily living	190 (66%)	58 (20%)	9 (3%)	21 (7%)	10 (3%)	288
Sport and recreation	206 (73%)	34 (12%)	15 (5%)	13 (5%)	16 (6%)	284
Quality of life	225 (78%)	19 (7%)	14 (5%)	14 (5%)	18 (6%)	290
Modified Harris hip score	229 (78%)	24 (8%)	11 (4%)	13 (4%)	16 (5%)	293
WOMAC						
Pain	216 (75%)	33 (11%)	14 (5%)	16 (6%)	10 (3%)	289
Stiffness	154 (53%)	51 (18%)	50 (17%)	14 (5%)	20 (7%)	289
Physical	188 (66%)	58 (20%)	9 (3%)	21 (7%)	10 (3%)	286
Total	206 (72%)	46 (16%)	5 (2%)	18 (6%)	10 (4%)	285

Interpretability

The minimal clinically important differences spanned from 7.39 for the modified Harris hip score to 12.89 for WOMAC stiffness. The detailed values are presented in Table VI. The percentage of patients experiencing important improvement (achieving at least a minimal clinically important difference[8]) spanned from 53% for WOMAC stiffness to 78% for HOOS quality of life and the modified Harris hip score, and was >72% for most of the measures. The detailed values are provided in Table VII.

Floor and Ceiling Effects

No measures demonstrated floor or ceiling effects over 15% (Table II).

Discussion

The results of this study demonstrated that the HOOS, WOMAC, and modified Harris hip score all showed adequate psychometric properties in patients undergoing periacetabular osteotomy for the treatment of symptomatic developmental dysplasia of the hip. The modified Harris hip score was the only measure with poor internal consistency. The minimal clinically important difference for most of the patient-reported outcome measure subscales was <11 points.

In this study, all subscores of both the WOMAC and HOOS had acceptable internal consistency. That was not the case for the function and gait components of modified Harris hip score. That indicates that the questions within each of modified Harris hip score components may not be capturing the same construct. However, the lack of internal consistency might also be explained by the low number of questions within the subscales of modified Harris hip score, which can lead to lower Cronbach alpha values[29]. To our knowledge, no other study has assessed the internal consistency of patient-reported outcome measures for periacetabular osteotomy.

Validity ensures that the components of the questionnaire are appropriate for the target patient population. The convergent and discriminant validity of each of the 3 patient-reported outcome measures is adequate for use in patients undergoing periacetabular osteotomy, apart from WOMAC stiffness (Table IV). Interestingly, the modified Harris hip score displayed the highest correlation coefficient with the SF-12 PCS score. An explanation for that finding might be that the modified Harris hip score focuses on overall patient physical disability as a function of hip problems.

All of the scores demonstrated adequate responsiveness; the greater the improvement felt by the patient, the higher the responsiveness indices were. In the only available study on psychometric properties of patient-reported outcome measures in periacetabular osteotomy, Davidson et al. examined WOMAC and SF-36 responsiveness in 83 patients with periacetabular osteotomy. They showed that the WOMAC was more responsive than the SF-36. They also demonstrated higher changes in the SF-36 PCS scores (Bodily Pain, General Health, Vitality) than the SF-36 MCS scores (Role Emotional and Mental Health)[15].

The values of the minimal clinically important difference were <11 points for most of the scores evaluated in this study. This is consistent with contemporary studies on periacetabular osteotomy[12,30,31] and a classic study that described the long-term outcomes[32]. The limitations of the distribution-based approach (i.e., strong influence of the population characteristics and possible insensitivity to true differences in patient perception) dictate that future studies also use the anchoring approach. At the same time, the minimal clinically important differences calculated in this study were also smaller than the mean differences between symptomatic and asymptomatic hips at a minimum of 14 years of follow-up in another study[33]. Therefore, the calculated minimal clinically important differences seem to represent a true minimal difference that can be observed after periacetabular osteotomy.

Although this current study represents an important step in establishing the psychometric properties of patient-reported outcomes for periacetabular osteotomy and thus incorporating an evidence-based approach into orthopaedics, it had several limitations. First, we did not perform a test-retest reliability calculation. However, the optimal time for retesting remains arbitrary. With shorter test-retest times, correlation might be inflated; with longer test-retest times, it would decrease. One might expect those values to be more important in the chronic disease setting, where there is a subset of patients who do not change over time, and thus this psychometric property can be calculated[25]. That stands in contrast with patient-reported outcomes after a surgical procedure, which represents a dramatic change to the natural course of the disease. Thus, the Cronbach alpha might represent a more intrinsic patient-reported outcome measure characteristic than those calculated with a test-retest method. Second, the choice of the Global Perceived Effect as the measure of responsiveness limited the ability to detect the amount of change. The Global Perceived Effect is prone to present intangible random variation due to changes in general health. In the current study, responsiveness was compared between the patients who experienced a significant change in their overall health (improved and worse groups) and those who had not. It must be stressed that the overall change in health is not equal to patients' satisfaction with the surgical procedure (almost one-third of the cohort did not feel improvement in their general health, but a majority of the patients were satisfied with the surgical procedure). Importantly, the satisfaction rates were similar to those reported by other studies[12,34]. Third, since the beginning of our study, which is a part of a multicenter cohort, new patient-reported outcome measures have been developed for patients with hip problems, such as the International Hip Outcome Tool (iHOT-33) and the Copenhagen Hip and Groin Outcome Score (HAGOS)[14,35]. Therefore, further research is warranted on psychometric properties of the more recent patient-reported outcome measures in the population undergoing periacetabular osteotomy.

Interestingly, the 2-year follow-up results that were available for the entire cohort did not differ significantly from the 1-year results (see Appendix). However, we report 1-year follow-up data due to the Global Perceived Effect question recall period and because the focus of this article is on the psychometric properties of the patient-reported outcome measures and not on the evaluation of the benefits and risks of treatment. The psychometric properties of the patient-reported outcome measures do not change significantly with the follow-up period, as they represent the intrinsic properties of the outcome measures, and the evaluation of the benefits and risks of treatment may change over the course of time.

Almost 20% of our patients had undergone a previous ipsilateral hip surgical procedure, which has been shown to influence the outcomes of periacetabular osteotomy[36,37]. However, the psychometric properties were not influenced by that fact and were primarily driven by the primary periacetabular osteotomy cases in our cohort (see Appendix).

There are multiple studies on psychometric properties of patient-reported outcome measures in other young adult populations undergoing a hip surgical procedure (patients treated with hip arthroscopy for intra-articular damage[13], patients with femoroacetabular impingement[38]). However, young adult populations undergoing hip arthroscopy or periacetabular osteotomy differ with regard to 3 major characteristics: disease processes, patient characteristics, and procedures At the same time, the fact that the calculated minimal clinically important difference values for periacetabular osteotomy are similar to those of hip arthroscopy[13] reinforces confidence in the reported minimal clinically important difference estimates.

On the basis of the strength of the psychometric properties measured in this study, the HOOS appears to be the most appropriate measure of patient-reported outcome in the population undergoing periacetabular osteotomy. In addition to displaying very high internal consistency (Cronbach alpha values ranged from 0.79 to 0.97) and validity compared with the SF-12 PCS score ($\rho > 0.4$ for all subscores), this patient-reported outcome measure also showed adequate responsiveness and minimal (≤5%) floor and ceiling effects for all but the quality-of-life subscore.

The importance of this study lies in the fact that it enables a more informed choice of a patient-reported outcome measure for research and clinical practice. The minimal clinically important differences calculated for the different subscores of patient-reported outcome measures will help to calculate the sample size for further studies. Even more importantly, they can further inform clinicians of the magnitude of a clinically meaningful change (answering the question: "Did this patient experience an important change?") in conjunction with the clinical and radiographic assessment.

The findings of this study indicate that the HOOS appears to be the most appropriate current measure of patient-reported outcome in patients undergoing periacetabular osteotomy. Some psychometric properties of the modified Harris hip score and WOMAC are reduced, and these patient-reported outcome measures may be less valuable in this patient group.

Appendix

eA Tables showing the distribution of scores in the study group at 1 and 2-year follow-up, internal consistency (reliability) of patient-reported outcome measures for periacetabular osteotomy in the study group and the primary periacetabular osteotomy subgroup, validity of patient-reported outcome measures compared with the SF-12 PCS score for periacetabular osteotomy, responsiveness of patient-reported outcome measures for periacetabular osteotomy, and minimal clinically important difference for patient-reported outcome measures for periacetabular osteotomy are available with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJS/F84). ■

Marcin K. Wasko, MD, PhD[1,2]
Elizabeth L. Yanik, PhD, ScM[1]
Cecilia Pascual-Garrido, MD[1]
John C. Clohisy, MD[1]

[1]Department of Orthopaedic Surgery, Washington University School of Medicine, Barnes-Jewish Hospital, St. Louis, Missouri

[2]Department of Radiology and Imaging, The Medical Center for Postgraduate Education in Warsaw, Otwock, Poland

E-mail address for M.K. Wasko: marcin@wasko.md

ORCID iD for M.K. Wasko: 0000-0002-5475-1999
ORCID iD for E.L. Yanik: 0000-0002-5835-0201
ORCID iD for C. Pascual-Garrido: 0000-0001-7487-4753
ORCID iD for J.C. Clohisy: 0000-0001-7828-9690

References

1. Matheney T, Kim YJ, Zurakowski D, Matero C, Millis M. Intermediate to long-term results following the Bernese periacetabular osteotomy and predictors of clinical outcome: surgical technique. J Bone Joint Surg Am. 2010 Sep;92(Suppl 1 Pt 2):115-29.

2. McKinley TO. The Bernese periacetabular osteotomy: review of reported outcomes and the early experience at the University of Iowa. Iowa Orthop J. 2003;23:23-8.

3. Clohisy JC, Schutz AL, St John L, Schoenecker PL, Wright RW. Periacetabular osteotomy: a systematic literature review. Clin Orthop Relat Res. 2009 Aug;467(8):2041-52. Epub 2009 Apr 21.

4. Siebenrock KA, Leunig M, Ganz R. Periacetabular osteotomy: the Bernese experience. Instr Course Lect. 2001;50:239-45.

5. Steppacher SD, Tannast M, Ganz R, Siebenrock KA. Mean 20-year followup of Bernese periacetabular osteotomy. Clin Orthop Relat Res. 2008 Jul;466(7):1633-44. Epub 2008 May 1.

6. Troelsen A, Elmengaard B, Søballe K. Medium-term outcome of periacetabular osteotomy and predictors of conversion to total hip replacement. J Bone Joint Surg Am. 2009 Sep;91(9):2169-79.

7. Smith MV, Klein SE, Clohisy JC, Baca GR, Brophy RH, Wright RW. Lower extremity-specific measures of disability and outcomes in orthopaedic surgery. J Bone Joint Surg Am. 2012 Mar 7;94(5):468-77.

8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49. Epub 2010 Feb 19.

9. Safran MR, Hariri S. Hip arthroscopy assessment tools and outcomes. Oper Tech Orthop. 2010 Dec;20(4):264-77.

10. Lieberman JR, Dorey F, Shekelle P, Schumacher L, Thomas BJ, Kilgus DJ, Finerman GA. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. J Bone Joint Surg Am. 1996 Jun;78(6): 835-8.

11. Roos EM. Outcome after anterior cruciate ligament reconstruction—a comparison of patients' and surgeons' assessments. Scand J Med Sci Sports. 2001 Oct;11(5):287-91.

12. Clohisy JC, Ackerman J, Baca G, Baty J, Beaulé PE, Kim YJ, Millis MB, Podeszwa DA, Schoenecker PL, Sierra RJ, Sink EL, Sucato DJ, Trousdale RT, Zaltz I. Patient-reported outcomes of periacetabular osteotomy from the prospective ANCHOR cohort study. J Bone Joint Surg Am. 2017 Jan 4;99(1):33-41.

13. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. Am J Sports Med. 2013 Sep;41(9):2065-73. Epub 2013 Jul 8.

14. Thorborg K, Hölmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. Br J Sports Med. 2011 May;45(6):478-91.

15. Davidson D, Yen YM, Jette AM, Olarsch S, Kim YJ. Responsiveness of the SF-36 and WOMAC following periacetabular osteotomy for acetabular dysplasia. J Bone Joint Surg Am. 2011 Dec 7;93(23):2214-8.

16. Nunley RM, Prather H, Hunt D, Schoenecker PL, Clohisy JC. Clinical presentation of symptomatic acetabular dysplasia in skeletally mature patients. J Bone Joint Surg Am. 2011 May;93(Suppl 2):17-21.

17. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am. 1969 Jun;51(4):737-55.

18. 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 May 30;4:10. Epub 2003 May 30.

19. Ware J, Kosinski Jr M, Turner-Bowker D, Gandek B. How to score version 2 of the SF-12 Health Survey (with a supplement documenting version 1). Lincoln: QualityMetric; 2002.

20. Bland JM, Altman DG. Cronbach's alpha. BMJ. 1997 Feb 22;314(7080):572.

21. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res. 2003 Jun;12(4):349-62.

22. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989 Dec;10(4): 407-15.

23. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J. 2007 Sep-Oct;7(5):541-6. Epub 2007 Apr 2.

24. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003 May;41(5):582-92.

25. Frans FA, Nieuwkerk PT, Met R, Bipat S, Legemate DA, Reekers JA, Koelemay MJ. Statistical or clinical improvement? Determining the minimally important difference for the vascular quality of life questionnaire in patients with critical limb ischemia. Eur J Vasc Endovasc Surg. 2014 Feb;47(2):180-6. Epub 2013 Oct 22.

26. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials. 1991 Aug;12(4)(Suppl):142S-58S.

27. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989 Mar;27(3)(Suppl):S178-89.

28. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995 Aug;4(4):293-307.

29. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 Sep;16(3):297-334.

30. Ricciardi BF, Fields KG, Wentzel C, Nawabi DH, Kelly BT, Sink EL. Complications and short-term patient outcomes of periacetabular osteotomy for symptomatic mild hip dysplasia. Hip Int. 2017 Feb 21;27(1):42-8. Epub 2016 Oct 24.

31. Gray BL, Stambough JB, Baca GR, Schoenecker PL, Clohisy JC. Comparison of contemporary periacetabular osteotomy for hip dysplasia with total hip arthroplasty for hip osteoarthritis. Bone Joint J. 2015 Oct;97-B(10):1322-7.

32. Lerch TD, Steppacher SD, Liechti EF, Tannast M, Siebenrock KA. One-third of hips after periacetabular osteotomy survive 30 years with good clinical results, no progression of arthritis, or conversion to THA. Clin Orthop Relat Res. 2017 Apr; 475(4):1154-68.

33. Wells J, Millis M, Kim YJ, Bulat E, Miller P, Matheney T. Survivorship of the Bernese periacetabular osteotomy: what factors are associated with long-term failure? Clin Orthop Relat Res. 2017 Feb;475(2):396-405. Epub 2016 May 12.

34. Biedermann R, Donnan L, Gabriel A, Wachter R, Krismer M, Behensky H. Complications and patient satisfaction after periacetabular pelvic osteotomy. Int Orthop. 2008 Oct;32(5):611-7. Epub 2007 Jun 20.

35. Mohtadi NG, Griffin DR, Pedersen ME, Chan D, Safran MR, Parsons N, Sekiya JK, Kelly BT, Werle JR, Leunig M, McCarthy JC, Martin HD, Byrd JW, Philippon MJ, Martin RL, Guanche CA, Clohisy JC, Sampson TG, Kocher MS, Larson CM; Multicenter Arthroscopy of the Hip Outcomes Research Network. The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). Arthroscopy. 2012 May;28(5):595-605; quiz 606-10.e1.

36. Polkowski GG, Novais EN, Kim YJ, Millis MB, Schoenecker PL, Clohisy JC. Does previous reconstructive surgery influence functional improvement and deformity correction after periacetabular osteotomy? Clin Orthop Relat Res. 2012 Feb;470(2): 516-24.

37. Ricciardi BF, Fields KG, Wentzel C, Kelly BT, Sink EL. Early functional outcomes of periacetabular osteotomy after failed hip arthroscopic surgery for symptomatic acetabular dysplasia. Am J Sports Med. 2017 Sep;45(11):2460-7. Epub 2017 Jun 15.

38. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery-a systematic review of the literature. J Hip Preserv Surg. 2015 Feb 6;2(1):15-27.