DR. JOEL J GAGNIER (Orcid ID : 0000-0002-3162-3935)

Article type : Special Article

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 Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire

Jack Weick[1], Jace Bullard[1], Jordan H. Green[1], Joel J. Gagnier[1-2]

1. Department of Orthopaedic Surgery, Michigan Medicine, University of Michigan, Ann Arbor, MI, USA.

2. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ACR.24231

Address correspondence to: Dr. Joel Gagnier ND, MSc, PhD, MedSport, Domino's Farms, 24 Frank Lloyd Wright Drive, Ann Arbor, Michigan, 48105. E-mail: jgagnier@med.umich.edu

INTRODUCTION

Accurate measurements of hip outcomes and function are vital in the study of hip pathology. Additionally, outcome measures can also be used in the clinical realm to track progress of treatment. Several outcome measures have been developed in an effort to accurately assess hip symptoms and function. A 2011 review of the most commonly used hip outcome scores was previously published in this journal (1). Since then, a number of additional studies focusing on these outcomes scores have been published to further refine our knowledge and potential utility of hip outcome scores. The purpose of review is to provide updated information with regards to the most commonly used hip outcome scores which include the Harris Hip Score (HHS), the Hip Disability and Osteoarthritis Outcome Score (HOOS), Patient-Reported Outcomes Measurement Information System (PROMIS), the Oxford Hip Score (OHS), the Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), the American Academy of Orthopedic Surgeons (AAOS) Hip and Knee questionnaire, and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).

HARRIS HIP SCORE (HHS)

Description

Purpose. The HHS was initially developed for the assessment of the results of hip surgery, specifically mold arthroplasties for post-traumatic osteoarthritis of the hip, in 1969 (2). It is now intended to evaluate various hip disabilities and methods of treatment in the adult population.

Content. The original HHS covers pain, function, absence of deformity, and range of motion. The pain domain measures pain severity and its effect on activities and need for pain medication. The function domain assesses daily activities (stair use, using public transportation, sitting, and managing shoes and socks) and gait. Deformity evaluates hip flexion, adduction, internal rotation, and extremity length discrepancy. Range of motion measures hip flexion, abduction, external and internal rotation, and adduction.

A modified version is also now available (HHS-modified) which only assess the pain and function components of the HHS. (3,4)

Number of items. In the original HHS, there are 10 items. In the HHS-modified, there are 8 items.

Response options/scale. The original HHS has a maximum of 100 points (best possible outcome) covering pain (1 item, 0-44 points), function (7 items, 0-47 points), absence of deformity (1 item, 4 points), and range of motion (2 items, 5 points). The only includes pain and function, so has a maximum of 91 points, however, the score is multiplied by 1.1 to get a maximum of 100.

Recall period for items. Not described

Cost to use. Free

How to obtain. Available in original article (2). URL: http://www.orthopaedicscore.com/ , http://www.ncbi.nlm.nih.gov/pubmed/5783851. Modified HHS can be obtained by only completing the pain and function components, then multiplying the score by 1.1.

Practical application

Method of administration. The HHS is a clinician-based outcome measure administered by a qualified health care professional such as a physician or a physical therapist. The modified HHS is a self-administered instrument.

Scoring. Each item has a unique numerical scale, which corresponds to descriptive response options. The number of response options as well as the number of points assigned varies by each item. In the original HHS, the range of motion item consists of six motions graded. Each range of motion factor is assigned an index factor and a maximum possible value. These points are added and multiplied by 0.05 to receive the total points for range of motion. The total score is calculated by summing the scores for

the 4 domains. The modified HHS is calculated by summing the pain and function components of the scale and multiplying by 1.1.

Score interpretation. Both the HHS and modified HHS have a maximum of 100 points. The higher the HHS, the less dysfunction. Scores below 70 are typically considered a poor result. 70-80 is considered fair, 80-90 is good, and 90-100 is considered excellent (2). Pain and function are the major score contributors with 44 points possible for pain and 47 possible for function (14 for activities of daily living and 33 for gait). Range of motion has a maximum of 5 points and deformity has a maximum of 4 points.

Respondent time to complete. Takes approximately 5 minutes to complete.

Administrative burden. No formal training is necessary to administer.

Translations/adaptations. The HHS has been used internationally (USA, Canada, Sweden, Europe, etc.), but there are no validated versions in other languages.

Psychometric information

Floor and ceiling effects. Unacceptable ceiling effects for the HHS were reported in 31 of 59 studies (5). Pooled data across studies (n of 6,667 patients) suggested ceiling effects of 20% (95% CI 18-22). The modified HHS had reported ceiling effects of 27.5% (6). In a population of 294 patients who underwent periacetabular osteotomy, the modified HHS had reported ceiling effects of 1% (7).

Reliability. For the original HHS, the test-retest reliability for total score was excellent for physicians (r = 0.94) and physiotherapists (r = 0.95). Both physicians and physiotherapists had excellent test-retest reliability for pain and function (r = 0.93-0.98) (8). The interrator correlations were good to excellent in two previous studies (0.74 – 1.0) (8,9).

The modified HHS was shown to have excellent reliability with a Cronbach's alpha score of 0.95 and high intraclass correlation (0.91 – 0.95) (10).

Validity. The content validity of the HHS has been tested by comparisons to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Short Form 36 (SF-36). The HHS has demonstrated no major differences when compared to these scores (8). In assessment of construct validity, the pain and function domains in HHS have been shown to correlate with similar domains in the

WOMAC, the Nottingham Health Profile, and the SF-36 (8,11). The correlation between the HHS and SF-36 is notably strong in the physical domains, and weak in the mental domains.

The modified HHS has been compared to the SF-12 physical and mental subscale for validity and was shown to have poor correlation with the mental subscale, but strong correlation with the physical subscale (7).

Responsiveness. In a study of 335 total hip arthriplasties (THA), the HHS was found to be responsive to pain and function at six-month post-operative follow-up, but weak at two-year follow-up (12). In a study of 293 periacetabular osteotomies, the modified HHS was found to be adequately responsive (7).

Minimally important differences. In a study assessing the HHS in femoral neck fractures at four and 12 months post-operatively, the standardized response mean was 0.75 which was the best in ability to detect change when compared to the Barthel Index and the EuroQol 5-domain (EQ-5D) (13). The modified HHS was shown to have the lowest minimal clinically important difference when compared to the WOMAC and HOOS (7).

Generalizability. Both the HHS and the modified HHS are generalizable to adult patients with hip disabilities or undergoing hip surgeries.

Use in clinical trials. Both the HHS and the modified HHS have been used extensively in clinical trials involving hip pathology and treatment of different hip conditions. The HHS has been used in trials that range from assessment of treatment choice for femoral neck and intertrochanteric femur fractures fractures to assessment of effectiveness of different injections for osteoarthritis of the hip (14–17). The HHS is also a popular tool in studies relating to THAs given the focus of the measure on pain and impaired physical function, which are the main indications for a THA (18–20). Similarly, the modified HHS is widely used in clinical trials relating to hip pathology including injections, management of hip fractures, hip arthroscopy, and THAs (21–25). The modified HHS has the benefit in clinical trials over HHS of being self-administered and not requiring a clinician for assessment.

Critical appraisal of overall value to the rheumatology community

Strengths. As discussed above, the HHS and modified HHS are both widely used in evaluating outcomes of THA given the focus on evaluation of pain and physical function (26). Additionally, the HHS and modified HHS have been shown to be appropriate for use in measurement of outcomes of intervention

for treatment of femoral neck fractures, hip arthroscopy, and conservative management of hip osteoarthritis (7,13,27).

Caveats and cautions. The main criticism of the HHS are the ceiling effects which limit its validity (5).

Clinical usability. The psychometric evaluation does not support interpretation of scores to make decisions for individuals.

Research usability. As long as the researcher is aware of the ceiling effects, both the HHS and modified HHS can be used for clinical outcome studies for a variety of hip pathologies.

HIP DISABILITY AND OSTEOARTHRITIS OUTCOME SCORE (HOOS)

Description

Purpose. The HOOS was developed as an adaptation of the Knee injury and Osteoarthritis Outcome Score (KOOS), which itself is an adaptation of the WOMAC. It was created as an instrument of assessment of an adult patients' opinion about their hip and hip disability. The HOOS was originally developed in 2003 and has been validated in two different versions (28).

Several variations of the HOOS have also been developed, the HOOS – Joint Replacement (HOOS-JR) short form and the HOOS – Physical Function (HOOS-PS) short form (29,30) and the HOOS-12. The HOOS-JR is a six-item instrument developed specifically for patients undergoing THA measuring the domains of pain and activities of daily living (ADLs). The HOOS-PS is a five-item measure of physical function designed to elicit patients' opinions about difficulties experienced due to their hip problems. The HOOS-12 is a 12-item HOOS retains the 3 subscores (pain, function/ADL, and quality of life).

Content. HOOS consists of five subscales: pain, other symptoms, function in ADLs, function in sport and recreation, and hip related quality of life (QOL). The HOOS-JR measures only the domains of pain and ADLs. The HOOS-PS measures physical function.

Number of items. The original HOOS contains 40 items total: 10 for pain, 5 for other symptoms, 17 for function in ADLs, 4 for function in sports and recreation, and 4 for hip-related QOL. The HOOS-JR contains 6 items total: 2 for pain and 4 for function. The HOOS-PS contains 5 items total all relating to physical function. The HOOS-12 has 12 items, 4 pain, 4 function, 4 QOL.

Response options/scale. In the HOOS, HOOS-JR, and HOOS-PS, standardized answer options are given in 5 Likert boxes. Each question is scored 0-4. Scores are summarized and transformed into a 0-100 scale. A score of 0 indicates extreme problems and 100 indicating no problems.

Recall period for items. The last week is taken into consideration for the questions.

Cost to use. Free

How to obtain. The HOOS, HOOS-PS, and HOOS-12 can be obtained at www.koos.nu. The HOOS-JR can be obtained from www.aaos.org/uploadedFiles/HOOS-JR-2016.pdf.

Practical application

Method of administration. The questionnaire for all forms of HOOS is patient reported.

Scoring. The manual scoring sheet is available from the above websites. Included are instructions for handling missing values. Computer scoring can increase clinical usefulness.

Score interpretation. Scores can range from 0-100 and each subscale can have scores from 0-100 where 0 indicates extreme problems and 100 indicates no problems. The HOOS scores within each subscale (pain, symptoms, ADLs, sports/rec, QOL) can be plotted for comparison of pre-intervention and post-intervention comparison visualization.

Respondent time to complete. The original HOOS takes 10-15 minutes for a subject to complete. The HOOS-JR, HOOS-PS, and HOOS-12 all take under 5 minutes to complete.

Administrative burden. Minimal administrative burden. Hand scoring of the HOOS can take 10-15 minutes without additional training. Computerized scoring can automate this process for instantaneous results without further administrative need.

Translations/adaptations. The HOOS is available in 25 total languages including English, French, German, Dutch, Italian, and Spanish among others (available on www.koos.nu). A number of adaptations of the HOOS exist including the knee injury and osteoarthritis outcome score (KOOS), knee injury and osteoarthritis score for children (KOOS-Child), foot and ankle outcome score (FAOS), rheumatoid and arthritis outcome score (RAOS), hip and groin outcome score (HAGOS), and neck outcome score (NOOS).

189

190 ***Psychometric information***

191 **Floor and ceiling effects.** Floor and ceiling effects have been reported in different subscales within the
192 HOOS. Floor effects have been reported to range 4.1-17.8% in the sport/rec subscale in subjects eligible
193 for THA and patients with hip osteoarthritis (31–33). Ceiling effects have been reported in THA in all
194 subscales of the HOOS (31). No floor effects have been reported in the hip arthroscopy population,
195 however, ceiling effects were reported 12 and 24 months post-operatively in the ADL and sport/rec
196 subscales (34).

197 The HOOS-JR has been shown to have low floor effects (0.6-1.9%), but ceiling effects up to 37-
198 46% after THA (29). Floor and ceiling effects of the HOOS-PS have only been studied in the French
199 translation of the HOOS-PS, but no floor or ceiling effects were observed (33).

200 **Reliability.** The reliability of the HOOS has been examined in patients treated conservatively for hip
201 osteoarthritis, patients treated with THA, and for patient undergoing hip arthroscopy (28,31–34). In THA
202 patients, the internal consistency ranged from 0.82-0.98 (Cronbach's alpha coefficient) with the ADL
203 subscale having the highest consistency of 0.94-0.98 (28,32,33). The test-retest reproducibility has been
204 shown to be high in the THA population with an intraclass correlation coefficient of 0.75-0.97 (28,32,33).
205 In the hip arthroscopy population, the test-retest reliability was excellent with intraclass correlation
206 coefficients ranging from 0.91-0.97 (34).

207 The HOOS-JR was shown to have acceptable internal consistency as measured by Person
208 Separation Index (PSI) of 0.86 (29). Test-retest reliability was not reassessed in the HOOS-JR. The HOOS-
209 PS was shown to have an intraclass correlation coefficient of 0.86 (33).

210 **Validity.** The HOOS content validity was performed by asking patients to rate item importance (28,31).
211 Construct validity has been confirmed in hip osteoarthritis, THA, and hip arthroscopy studies by
212 comparison to the SF-36, Oxford Hip Score (OHS), and Lequesne Index, and visual analog scale for pain
213 (31–34). The HOOS-JR and HOOS-12 construct validity was confirmed with high correlations with the
214 pain and ADL domains of HOOS and the pain and function domains of the WOMAC with moderate
215 correlations with the other domains of the HOOS and WOMAC scores (29).

216 **Responsiveness.** All domains of the HOOS were found to be responsive in hip arthroscopy at 9 and 12-
217 month post-operative follow-up (34). In a study of patients who underwent a periacetabular osteotomy,

the HOOS was found to be responsive at 12-month follow-up (7). The HOOS-JR and HOOS-12 were found to have very high responsiveness up to 2-years post-operatively (29). The HOOS-PS was also found to have good responsiveness, but was only assessed 1-month post-operatively (33).

Minimally important differences. The smallest detectable difference of the HOOS ranged from 9.6 for the ADL domain to 16.2 in the QOL domain (33). In hip arthroscopy, the minimal detectable change ranged from 9 for ADL to 19 in the QOL domain (34).

Generalizability. The HOOS is generalizable to adult populations with hip ailments or undergoing hip procedures. The HOOS-JR was developed for patients undergoing THA.

Use in clinical trials. The HOOS has been used extensively in clinical trials relating to hip osteoarthritis and hip pathology. The HOOS is one of the more common patient reported measures of hip function and post-operative outcomes. It has been used extensively in THA clinical studies (35–37), hip dysplasia surgery (38,39), hip arthroscopy (40), and even in development of pain management protocols after hip surgery (41). The HOOS-JR has only existed since 2016 and is intended mainly for THA, but it has been used in recent trials related to THA outcomes (42,43). The HOOS-PS has also been used in clinical trials relating to hip surgery, but is used far less frequently than the original HOOS (18,44,45).

Critical appraisal of overall value to the rheumatology community

Strengths. The HOOS can be used for younger and more active patients given the subscales and individual domains. The HOOS has shown favorable qualities in reviews of psychometric properties in assessment of patients after THA, hip arthroscopy, periacetabular osteotomy, and patients with hip osteoarthritis not undergoing surgery (7,34,46,47).

Caveats and cautions. The HOOS is overall well received, however, it does have ceiling effects in all domains which are also present in the HOOS-JR (29,34).

Clinical usability. The HOOS can be used to follow patients with hip conditions over time in clinic. The HOOS-JR can be used in clinics which perform THA for monitoring patients.

Research usability. The HOOS, HOOS-JR, HOOS-12, and HOOS-PS are all usable in the research setting for hip conditions.

PATIENT-REPORTED OUTCOMES MEASUREMENT INFORMATION SYSTEM (PROMIS)

Description

Purpose. PROMIS was developed in 2004 by the National Institutes of Health (NIH) for use in clinical care and medical research (48). This system was intended to be used across multiple specialties, be publically available, precise, and flexible.

Content. The PROMIS adult profile covers three domains: physical health, mental health, and social health. Within each of these domains are more specific profile domains. Profile domains under physical health include fatigue, pain intensity, pain interference, physical function, and sleep disturbance. Profile domains under mental health include anxiety and depression. Within orthopaedics, the PROMIS physical function, pain interference, and depression domains have been most commonly used (49–53).

Number of items. There are multiple options of measures within PROMIS which can vary in the number of items. The short forms (SFs) are a fixed set of 4-10 items for one individual domain. Computer adaptive tests (CATs) provide items selected dynamically from an item bank based upon the subject's previous answers. The CATs are usually 4-12 items in length. Profiles are fixed collections of SFs from 7 different domains. Several versions of PROMIS Profiles exist ranging from 4-8 questions per domain.

Response options/scale. PROMIS uses a T-score metric. 50 is designed to be the mean of the population and 10 is one standard deviation (SD) of the population. Therefore, a score of 60 would indicate that subject is one SD above the mean for the population within the measured domain. Response options of the questions depend on the category of question. Pain is typically measured on a 0-10 point scale. Frequency, duration, intensity, and capability related questions typically have a scale of five variable options.

Recall period for items. Most PROMIS items use a 7-day recall period. However, typically within the physical function domain, the question is assessed in the present tense.

Cost to use. Free

How to obtain. The different PROMIS measures can be obtained from the HealthMeasures website at www.healthmeasures.net/explore-measurement-systems/promis.

Practical application

Method of administration. PROMIS scales are administered via paper or computer-based forms.

Scoring. Scoring manuals are available on the HealthMeasures website to allow conversion of the score to a T-score. PROMIS measures are scored on a T-score metric. 50 is designed to be the mean of the population and 10 is one SD of the population. Therefore, a score of 60 would indicate that subject is one SD above the mean for the population within the measured domain.

Score interpretation. A score of 50 indicates the subject measures at the mean of the population in that particular metric. One SD is 10 points. Therefore, a score of 60 would indicate that subject is one SD above the mean for the population within the measured domain.

Respondent time to complete. PROMIS domains relating to hip pathology typically take less than 5-10 minutes for the subject to complete.

Administrative burden. Minimal overall administrative burden. However, the distributor must have access to the scoring manuals for interpretation of the subjects' responses.

Translations/adaptations. The PROMIS adult domains are available in over 45 different languages including English, Spanish, French, German, Dutch, and Italian.

Psychometric information

Floor and ceiling effects. There has been little overall study on psychometric properties of the PROMIS score in relation to hip pathology. In a study of patients who underwent periacetabular osteotomy for symptomatic hip dysplasia, PROMIS showed no floor or ceiling effects both pre and post-operatively (39). When studied in patients who underwent THA for symptomatic hip osteoarthritis, the PROMIS depression domain showed floor effects up to 20% pre-operatively and 30-45% post-operatively (54). In the same study, the PROMIS pain interference domain showed no floor effects pre-operatively, but floor effects of 21-26% at 1 year post-operatively. No ceiling effects were observed in the depression, pain interference or physical function scores pre or post-operatively.

Reliability. Few studies have assessed the reliability of PROMIS instruments related to treatment of hip pathology. One study examining the psychometric properties of one CAT of the PROMIS (the lower extremity CAT) compared to the modified HHS and the Hip Outcome Score showed excellent reliability with a Cronbach's alpha score of 1.00 (6)

Validity. To our knowledge, content validity of PROMIS instruments have not been evaluated specifically for hip pathology. Group validity has been assessed previously in patients with a general diagnosis of "osteoarthritis", but not specifically for hip osteoarthritis (55).

Responsiveness. Responsiveness of PROMIS instruments for hip pathology has been evaluated in study of patients who underwent THA. The pain interference and physical function domains were found to have high responsiveness post-operatively in patients with hip osteoarthritis after THA (54).

Minimally important differences. Minimally clinically important difference (MCID) has typically been defined as 5, half of the normalized SD of reported PROMIS scores (SD=10). In a study of patients who underwent THA, 66-78% of patients had a MCID post-operatively in the pain interference domain, 61-75% of patients had a MCID in the physical function domain, and 44-46% of patients had MCID in the depression domain (54).

Generalizability. PROMIS scores are generalizable to adult populations with hip ailments or undergoing hip procedures. There are separate PROMIS instruments for pediatric patients.

Use in clinical trials. Compared to other measures of hip function and symptoms, PROMIS instruments have been used relatively less frequently in clinical trials. However, more recent studies have begun to use PROMIS instruments in assessment of outcomes related to surgical intervention for different hip pathologies. Clinical studies on outcomes after THA have used PROMIS scores for assessment of both functional outcomes and pain improvement post-operatively (56,57). PROMIS instruments have been used in clinical trials assessing baseline disability and functional performance in patients with femoracetabular impingement (FAI) (58,59). Lastly, PROMIS scores were used in one clinical trial studying post-operative outcomes of correction of mildly dysplastic hips with periacetabular ostoetomy (60).

Critical appraisal of overall value to the rheumatology community

Strengths. PROMIS instruments have the benefit of adaptability to patient responses to enable the questionnaire be adaptable to individual patients. Additionally, PROMIS instruments are becoming more widely used in all fields of healthcare. Therefore, if used in hip pathology, it has the potential added benefit of easier interpretation by healthcare practitioners or researchers who may not be as familiar with other hip-specific outcome measures.

Caveats and cautions. There is a relative paucity of literature involving psychometric properties of PROMIS instruments in patients with hip pathology. Further investigation should be performed in this area. Additionally, high floor effects have been reported in a previous THA study as mentioned previously (54).

Clinical usability. Although PROMIS instruments have not been extensively used in assessment of outcomes in patients with hip pathology, it can be used to follow patients both pre and post-operatively.

Research usability. Further evaluation of the psychometric properties of PROMIS instruments specifically for hip pathology would be beneficial, however, PROMIS instruments have been used in several clinical trials involving THA, FAI, and hip dysplasia (56–60). Investigators should take note of the previously reported high floor effects in THA populations (54).

OXFORD HIP SCORE (OHS)

Description

Purpose. To assess outcome after total hip arthroplasty (THA) by measuring a patient's perceptions after surgery. Originally described in 1996 and was updated in 2007 (61,62).

Content. OHS assess his pain (6 items) and function (6 items) of the hip in relation to daily activities such as walking, dressing, slipping, etc.

Number of Items. 12 items with 5 categories of response; no subscale's.

Response options/scale. The original scoring system as described by Dawson et al in 1996 ranged from 1-5 (best to worst) with a total score of 12-60 (least difficulties to most difficulties). This scoring system was modified in 2007 by Murray et al to new item ranges of 0-4 (worse to best) with total scores ranging from 0-48 (most difficulties to least difficulties) (62,63).

Recall period for Items. During the past 4 weeks.

Examples of Use. This patient reported outcome measure has been used and several countries in both clinical and research settings. It has been validated and used in both primary and revision hip replacements (64–75).

Practical application

How to obtain. The Oxford Hip Score questionnaire is free to use and is available online at http://www.orthopaedicscore.com/scorepages/oxford_hip_score.html. Further information the Oxford Hip Score and all the other Oxford Orthopaedic Scores can be found at https://phi.uhce.ox.ac.uk/ox_scores.php.

Method of Administration. The orthopedic hip score questionnaire can be self-administered or completed over the phone (61,62,76).

Scoring. As previously stated, each item (12) contains five possible responses. According to the updated scoring system, these five responses are scored from 0-4 (worse to best) resulting in a possible overall score range of 0-48 (most difficulties to least difficulties). The maximum of 2 missing values can be accepted and replaced by mean values. Overall scores should not be calculated if more than 2 items are left unanswered. If a patient marks multiple responses for 1 item, the worst response should be used for calculation of scores (62).

Score Interpretation. Categories for the OHS based on data from the Harris Hip Score and translated to the 0-48 scoring has suggested cut off scores: >41 as excellent, 34-41 as good, 27-33 as fair, and <27 as poor (77). More recent research has shown that a postoperative OHS of greater than 37.5 is associated with a successful outcome (78–80). Additionally, clinicians have attempted to use the OHS as a method of screening out patients who do will not require total hip replacement. Neufeld et al found that an OHS of 34 or higher was a good predictor of successful nonoperative management of hip arthritis (81). With use of the classification system described by Kalairajah et al, the OHS at 6 months is a useful predictor of early revision THA. A poor score was associated with a revision risk of 7.6% compared to a revision risk of 0.7% in patients with good/excellent scores (67,77). Lastly, normative values were established in 2015 and published by Hamilton et al for patients undergoing THA. They were able to establish normative values for male and female patients in 4 age categories (<60, 60-70, 70-80, and >80) preoperatively and 12 months after THA (79).

Respondent Burden. The OHS takes between 2-15 minutes to complete. Based on patient interviews, issues have been raised regarding item clarity and double-barreled questions (82,83).

Administrative Burden. The burden of administering the OHS is minimal. As the OHS is a patient-reported questionnaire, the time to score is short and involves only elementary arithmetic. No specific training is required to score the OHS.

Translations/Adaptations. The OHS has been widely used in many countries. It has been translated and validated in Japanese, Dutch, German, Turkish, Spanish, Mandarin, Italian, Danish, and with the use of on-site translator (66,80–89). It has also been translated to French, Iranian, and Korean forms, without the supporting validation studies (84–86). The Orthopaedic Oxford Scores also include similar questionnaires for assessing outcome after knee replacement, shoulder replacement, elbow replacement, ankle replacement, and shoulder instability.

Pyschometric Information

Method of Development. Questions were made based on patient interviews where hip arthritis patients were asked to report their experiences and frustrations. Patients were involved in content validity of the questionnaire (61). The OHS underwent item-response theory testing in 2004 by Fitzpartick et al, and there was an overall good item fit of the data to the Rasch model (87).

Acceptability. In a 2000 study, Fitzpatick et al showed that 90% of patients fill out the questionnaire to completion. In general, older patients and patients with more severe medical problems were less likely to complete the questionnaire completely compared to younger and healthier patients. In their study, the most problems referred to the item regarding distance walked before severe pain (88,89). In one study, up to 10% of English-speaking Americans misinterpreted this item. The authors hypothesize that the use of “Not at all” as an answer choice implies that the patient never has pain, so this confusion may lead to an overall underestimation of their hip function. Like other Hip patient reported outcome measures, the OHS is subject to statistical ceiling effects (approximately 13.5%), however very low levels of statistical floor effects are observed with the OHS (90,91).

Reliability. Several studies have investigated the internal consistency of the OHS preoperatively and postoperatively. The studies have shown the internal consistency of the OHS to be high (0.84–0.93) preoperatively and at 3, 6, 12, and 24 months postoperatively (61,92,93). Reproducibility, as measured by the coefficient of repeatability or inter-class correlations using the Bland and Altman method, has also been studied extensively and shown to be consistently strong (61,92,94).

Validity. During the development of the OHS, patients were asked to comment on and to include hip related problems not addressed by the draft questionnaire for content validity (61). High correlations (0.67 – 0.85) have been described when comparing the OHS to other patient reported outcomes of pain and function related to hip pathology (61,70,77,90,95,96).

Ability to Detect Change. The OHS has favorable responsiveness when compared with generic measures, such as will Short Form 36 and EuroQol 5–domain, and disease-specific measures, including the Western Ontario, McMaster Universities Osteoarthritis Index, and Arthritis Impact Measurement Scales. Effect size varied from 2.1-3.1 at 6-24 months after THA and was 1.84 after revision THA (61,70,90–93,96,97). In 2007, Murray et al estimated the minimal clinically important difference to be between 3-5 points after joint replacement (62). Further study in 2015 by Beard et al found the minimal detectable change to be 5 points after THA (80).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The OHS assess his pain and functional outcomes in patients undergoing total hip arthroplasty. It has been shown to provide good psychometric properties and has been reported to be a useful predictor and early revision after total hip arthroplasty.

Caveats and Cautions. As previously discussed, the OHS has some questions that may be misinterpreted, therefore leading to an underestimation of hip pain and function. Additionally, concerns have been raised regarding the lack of items concerning activities requiring a large angle of hip flexion. Lastly, it has been shown that 10% patients who was first language is not English may misinterpret one of items.

Clinical Usability. The questionnaire is quick, easy to use, free, and self-administered. Therefore, clinical usability is high. With that said, a single administration will not provide useful information on individual, however with repeated administrations useful information can be gleaned.

Research Usability. Due to his ease of use and high response rate, the OHS is one of the preferred patient reported outcome measures for large studies on long-term hip replacement outcomes (65).

LEQUESNE INDEX OF SEVERITY FOR OSTEOARTHRITIS OF THE HIP (LISOH)

Description

Purpose. The LISOH was developed in France in the early 1980's to evaluate the severity of hip osteoarthritis in drug trials in an adult French population, the long-term treatment effects for hip

444 osteoarthritis, and as a help in decision making regarding the need for hip replacement (98). It was
445 modified in 1997 and became known to some as the Lequesne Algofunctional Index.

446 **Content.** The LISOH is an index that covers osteoarthritis-specific symptoms and physical function
447 disability. It is composite measure of aggregating symptoms and function, which are not graded
448 separately, where pain is analyzed by 5 items, maximum distance walked by 2 items, and activities of
449 daily living (ADL) by 4 items (98). This instrument is available in several versions: Interview based, self-
450 administered, and in modified versions due to changed scoring and wording (98–101).

451 **Number of Items.** There are 11 items.

452 **Response options/scale.** The score ranges from 0 (no pain or disability) to 24 (maximum pain or
453 disability) and is scored as a sum of all the items (98).

454 **Recall period for Items.** Not specified.

455 **Examples of Use.** The LISOH has been used in both clinical and research settings since its development
456 in the 1980s. Clinically it has been used to assess the severity of hip osteoarthritis and help with
457 indications for total hip arthroplasty (101–103). In the research realm, the LISOH has been used to
458 determine the effectiveness of pharmacologic interventions on hip osteoarthritis and to assess the long-
459 term impact of post-THA rehabilitation (98,104).

460

461 ***Practical application***

462 **How to obtain.** The LISOH is free to use and can be accessed at
463 https://oarsi.org/sites/default/files/docs/2013/lequesne_eng_ndex.pdf.

464 **Method of Administration.** The LISOH can be self-administered, interviewer-based, or completed by a
465 clinician during a clinical assessment.

466 **Scoring.** The original scoring consists of score ranges from 0-8 for each part of the LISOH questionnaire
467 (Pain /discomfort, maximum distance walked, and ADL) resulting in a total score range of 0-24. A
468 modification in 1991 added a question regarding sexual activity to be included when appropriate,
469 resulting in a total score range of 0-28 (98,99).

470 **Score Interpretation.** With the original scoring consisting of a total score range of 0-20 four-point, 0 = no
471 handicap, 1-4 = mild handicap, 5-7 = moderate handicap, 8-10 = severe handicap, 11-13 = very severe

handicap, >/= 14 = extremely severe handicap. A score > 11-12 points has been suggested to indicate need for total hip arthroplasty (103). The questions are suggested to score disabilities connected with a single hip. There are no indications of how to score in the case of bilateral hip osteoarthritis, which complicates interpretation of the LISOH in those patients (102).

Respondent Burden. The LISOH questionnaire takes less than 5 minutes to complete (102,105,106).

Administrative Burden. While scoring of the LISOH questionnaire takes only a few minutes, some training may be required for use of the questionnaire in an interview based environment to achieve interobserver reproducibility (98,105).

Translations/Adaptations. The LISOH has been translated and validated for hip osteoarthritis in English, French, German, Turkish, Korean, Spanish, Greek, Persian, and Portuguese (101,102,107–115).

Pyschometric Information

Method of Development. Developed in France by hip specialists in the early 1980's through patient interviews in an adult French population.

Acceptability. Several studies have assessed the LISOH questionnaire using Rasch analysis, and unfavorable results have caused some to question the psychometric properties of the questionnaire. Furthermore, in direct comparison to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), the LISOH exhibits worse internal consistency reliability and construct validity in multiple patient populations (107,110,111,116). Other issues raised regarding the LISOH questionnaire include the clarity of the items, with one study determining that 2 out of 10 patients in a French population require additional explanation to the fill out the questionnaire, and poor item response rate, with one study noting an item response rate of approximately 71% (102,105).

Reliability. In general, internal consistency has been found to be satisfactory for the LISOH composite score (Cronbach's alpha 0.83-0.84) (101,102,108). With that said, the internal consistency has been shown to be lower for the pain/discomfort part of the LISOH in comparison to the function part (Cronbach's alpha 0.63 vs 0.84, respectively) (101). Satisfactory test–retest reliability has been shown for the composite score, with interclass correlation coefficient's ranging from 0.51-0.96 (101,107,111). With regards to interrater reliability, the interview–based questionnaire had a mean deviation of 0.55 points when rated by 2 observers (98).

Validity. The construct validity and convergent validity of the LISOH questionnaire have been shown to be inferior to other patient reported outcome measures (47,101,102,107,111).

Ability to Detect Change. The MCID of the LISOH remains to be elucidated. When using the LISOH to assess the long-term impact of active drug treatment for hip osteoarthritis, an effect size of 1.3-1.8 was observed (100).

Critical Appraisal of Overall Value to the Rheumatology Community

Strengths. The LISOH is quick, free, and easy to access.

Caveats and Cautions. Problems have been raised regarding the validity of the LISOH and its utility as a single measure of outcomes after THA.

Clinical Usability. Given its poor validity and reliability relative to other patient reported outcomes, it is not recommended to use the LISOH as the sole outcome measure on an individual patient level.

Research Usability. Given its ease of use, availability, and relatively high interrater/interobserver reliability, the LISOH questionnaire may have some utility in the study of THA outcomes in large patient populations.

American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire

Description

Purpose: Throughout the 1990s and early 2000s the AAOS created a series of questionnaires designed to measure and analyze musculoskeletal outcomes. These assessments covered all body regions in adults and children. The hip and knee questionnaire was a specific version of the more general lower limb questionnaire and was published in 2004. In combination with the SF-36, the AAOS hip and knee questionnaire was effective at assessing hip and knee conditions and the effects of treatment (117). The survey was previously available through the AAOS website but has now been replaced by the HOOS and KOOS surveys.

Content: The questionnaire is identical to the AAOS lower limb questionnaire with the attribution of pain to the hip or knee. It asks respondents to answer questions regarding hip or knee stiffness,

swelling, and function. Specifically, the functional questions assess the respondent's pain while walking on flat surfaces, going up or down stairs, and while lying in bed at night. The final two questions assess the ability to get around and the level of difficulty with taking on and off socks.

Number of Items: 7 items. The worse hip is given preference in the case of bilateral symptoms.

Response Options/Scale: For questions regarding stiffness and swelling, respondents choose from 5 possible options on a Likert scale from "not at all" to "extremely". For questions regarding pain during functional activity, there are 7 response options on a Likert scale from "not painful" to "could not do", as well as a "could not do for other reasons" option. There are 7 response options for ability to get around and 6 response options for difficulty with taking on and off socks.

Recall Period for Items: 1 week

Cost to use: Free

How to obtain: Previously available on aaos.org

Practical Application

Method of administration: Self-administered questionnaire

Scoring: Standardized and normative scores may be calculated. Unanswered items are not to be included in calculation of mean scores. In addition, if more than half of the items are missing, a score cannot be calculated. This includes items marked "could not do for other reasons", which is considered equal to a missing response. Instructions on how to calculate standardized and normative scores are included in a worksheet with the survey.

Score interpretation: Standardized scores are calculated from 0-100, 0 being most disability and 100 being least disability. Normative scores are calculated to a mean population score of 50, with higher scores indicating higher function (118).

Respondent time to complete: 2-3 minutes

Administrative burden: The questionnaire can be scored very quickly with a scoring sheet, and within 15 minutes if scored by hand.

Translations/adaptations: The hip and knee questionnaire is an adaptation of the lower limb questionnaire. There are also sports/knee and foot and ankle adaptations (117).

Psychometric Information:

Floor and Ceiling Effects: Not measured (117).

Reliability: The hip and knee questionnaire has a good internal consistency with Cronbach alpha coefficient of 0.8, calculated from 43 patients with a hip or knee complaint. For test-retest reliability a Pearson's correlation coefficient was calculated on 40 patients and found to be 0.91 (117).

Validity: All AAOS instruments were developed by clinicians who continually confirmed the face and content validity for each questionnaire. Construct validity was obtained by correlating scores from the hip and knee questionnaire with the AAOS lower limb core scale (r=0.95), the unweighted mean of three SF-36 physical subscales (r=0.7), and a scale created from a physician assessment of function and pain (r=0.73 and r=0.69 respectively). In addition, a global score for the WOMAC was calculated and correlated with the AAOS hip and knee questionnaire (r=0.89) (117).

Responsiveness: There is no direct measure of the hip and knee questionnaire's responsiveness, but the lower limb core scale – from which the hip and knee core scale was adapted – has been assessed. An absolute change score, calculated as the difference between the baseline and follow-up scores, was found to be moderately correlated with a transition score (r=0.53). This transition score was calculated from the combined responses of the patient and a physician on their perceptions of improvement over 1 year. Based on the strong correlation of lower limb questionnaire scores to hip and knee questionnaire scores, it is likely that the hip and knee questionnaire possesses similar ability to detect change.

Minimally important differences: To our knowledge, no MID for the AAOS hip and knee questionnaire has been calculated.

Generalizability: The AAOS hip and knee questionnaire is applicable to adult populations over age 18 with pain attributable the hip or knee.

Use in clinical trials: To our knowledge, the AAOS hip and knee questionnaire has not been used in any major clinical trials. It has been used to assess outcomes in patients with slipped capital femoral epiphysis (119), and the AAOS lower limb questionnaire appears in numerous studies covering a broad range of musculoskeletal topics including outcomes limb lengthening and lower extremity amputation (120,121).

Critical appraisal of overall value to the rheumatology community

Strengths: The AAOS Hip and Knee questionnaire proved to be an effective instrument in assessing hip and knee conditions and the effect of appropriate treatments. It is a short survey that is easily administered and scored. A thorough psychometric evaluation was performed in 2004 to demonstrate its reliability and validity, and in conjunction with the SF-36 it can be an overall effective outcomes measure.

Caveats and cautions: As patient reported outcomes have become an integral part of the overall effort to deliver quality care in the United States, other surveys specific to hip and knee outcomes have been developed and endorsed by the Centers for Medicaid Services (122). The AAOS hip and knee questionnaire is not included in that group of endorsed outcomes measures. This is likely the reason this survey is seen less commonly in the literature, and why there have not been any follow-up studies to the original evaluations of the AAOS questionnaires published in 2004.

Clinical/Research usability: A useful tool given its size and ease of administration, but not practical given the rising popularity of other surveys as above.

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

Description

Purpose: The WOMAC was developed to measure the symptoms and physical disability of patients with hip or knee arthritis (123). It was created as part of a randomized, controlled trial of two anti-inflammatory medications in the treatment of hip and knee arthritis (124). As such, the WOMAC was intended to be able to detect clinically important and relevant changes to treatment (125,126).

Content: The survey consists of 3 subscales that evaluate pain during certain positions or movements, stiffness at different times of day, and difficulty with performing certain activities.

Number of Items: 24 total items. There are 5 items in the pain section, 2 items in the stiffness section, and 17 items in the functional section.

Response Options/Scale: The WOMAC is available in 5-point Likert, 100mm visual analog scale (VAS), and 11-box numerical rating scales (127). The Likert scale offers five response options ranging from none to extreme, with corresponding score values from 0-4. The VAS and numerical rating scales offer responses selected on 100mm or 11-box horizontal scales, ranging from "none" on the left to "extreme" on the right (128).

Recall Period for Items: 48 hours

Cost to use: There is a fee to access the questionnaire and user guide. Fee information is available after submitting a request to use.

How to obtain: Requests to use the WOMAC for clinical or research purposes must be submitted via the contact section on the website www.womac.com.

Practical Application

Method of administration: Primarily designed as a self or interview-administered questionnaire. The WOMAC has been validated for use in person, over the phone, or via computer or mobile phone (129–131). Patients have been shown to respond similarly to paper or online versions (132).

Scoring: For the Likert scale version, the pain, stiffness, and function sections possess potential summative score totals of 20, 8, and 68 respectively. A global score is calculated by combining the 3 sub-scores (125). For the VAS version, respondents mark a point along the horizontal 100mm line. This is measured in millimeters and totaled out of 2400 (133). Normative values were published from an Australian population in 2010 (134). In the case that 2 or more items are missing from the pain subscale, both items missing from the stiffness subscale, or 4 or more items missing from the functional subscale, responses are declared invalid and not included for analysis.

Score interpretation: On both the Likert and VAS versions, lower scores equate to less pain and disability.

Respondent time to complete: 5-10 minutes

Administrative burden: Questionnaire responses require approximately 5 minutes to score with minimal training required.

Translations/adaptations: There have been multiple versions of the WOMAC questionnaire, with the most recent being version 3.1, updated in 2016. Short-form versions of the WOMAC have been developed, specific to total hip or total knee replacement, and were found to be equally responsive compared to the original WOMAC (135). The WOMAC has validated language translations in Arabic, Chinese, Dutch, Finnish, German, Hebrew, Italian, Japanese, Korean, Moroccan, Singapore, Spanish, Swedish, Thai, and Turkish (107,108,136–148).

Psychometric Information:

Floor and Ceiling Effects: Floor effects for WOMAC subscales or total scores are generally minimal or 0 (149). However, in one 2005 study, floor effects at 6 months and 2 years following total hip arthroplasty were significant in the pain (25% and 39%) and stiffness (30% and 46%) subscales (150). In a study of patients after periacetabular osteotomy, floor effects in the pain and stiffness subscales were <1% and 3% respectively. Floor effects were absent in the function subscale and aggregate score (7).

Ceiling effects are more commonly reported in all WOMAC subscales, particularly in pain and stiffness (7,95,149). A ceiling effect has been demonstrated in the pain and stiffness subscales when evaluating patients at 10 weeks and 12 months after hip fracture (149). These effects appear to be significantly different between young (18%-36%) and old (38%-53%) age groups (149). In patients after periacetabular osteotomy, the ceiling effect is more substantial than the floor effect, but still small (4%, 11%, 5%, 2% for pain, stiffness, function, total scores) (7). Findings in the literature are mixed, as one study demonstrated low ceiling effects of <5% for all WOMAC subscales at 6months and 2 years after total hip arthroplasty, while another demonstrated significant ceiling effects (14-38%) at 1 year post-operatively (95,150).

Reliability: The WOMAC has demonstrated excellent internal and test-retest reliability. The Cronbach alpha coefficient for the global score is reported as high as 0.97, with the lowest subscale being the stiffness subscale (alpha= 0.86) (7,128). Test-retest reliability is equally as strong, with Pearson's correlation coefficients reported as high as 0.96 (128). In addition to patients with osteoarthritis,

WOMAC scores have demonstrated good internal reliability amongst patients after hip fracture, with Cronbach coefficients ranging from 0.83-0.98 and 0.79-0.97 for young and old age groups respectively (149).

Validity: Since its creation in the late 1980s, the WOMAC has been extensively validated for patients with hip and knee arthritis (123). Additional validation has occurred for each of the aforementioned translations. More recently, the WOMAC has demonstrated good construct validity when compared to other PROMs. There is strong correlation of total WOMAC and SF-12 scores in patients after hip fracture, ranging from 0.78-0.84 for young and old age groups respectively (149). Similar results have been published correlating WOMAC to EQ-5D subscales (149). In patients after periacetabular osteotomy, Spearman correlation coefficients between WOMAC subscales and the SF-12 physical subscale are moderate to strong ($p>0.5$) except for the stiffness subscale ($p=0.38$) (7).

Responsiveness: For patients with hip osteoarthritis, the standardized response means (SRM) calculated for the WOMAC are high, exceeding 1.0 (151,152). In 162 patients who underwent total hip arthroplasty for osteoarthritis there was a mean WOMAC score change of 29, effect size of 1.84, and SRM of 1.6 (153). SRM values at 6 months and 2 years post op have been shown to continue increasing to 1.86 and 1.98 respectively (150). In addition, a statistically significant improvement in scores has been shown up to 2 years following total hip replacement (150). In patients who sustain hip fractures, responsiveness is moderate (SRM=0.66) and small (SRM=0.24) amongst patients less than and greater than 80 years old respectively (149). These differences are likely due to the fact that WOMAC scores are lowest just before a hip replacement, and not likely to be obtained prior to a patient sustaining a hip fracture (149).

Minimally important differences: The minimal clinically important difference (MCID) for patients who underwent total hip arthroplasty for hip arthritis is 10.2 points (153). The MCID for patients who undergo periacetabular osteotomy for hip pain secondary to hip dysplasia is approximately 11 points for the total WOMAC score (7). A prospective cohort study of over 1300 patients identified a minimal clinically important improvement of 7.9 on the WOMAC-function subscale in patients with hip arthritis who initiated non-operative treatment (154). In addition, mean changes of 9-12mm (100mm normalized VAS) on WOMAC scales were perceptible by patients with hip and knee osteoarthritis (155).

Generalizability: The WOMAC questionnaire has been applied across wide groups of adult populations. To our knowledge, no study has validated the WOMAC in pediatric populations. One active study

validating WOMAC scores in a pediatric population with Perthes disease can be found in the recruitment phase on clinicaltrials.gov

Use in clinical trials: As one of the oldest PROMs validated for assessing the hip, the WOMAC has been used extensively in studies affecting many types of hip pathology and responses to treatment, including osteoarthritis, hip dysplasia, and hip trauma.

Critical appraisal of overall value to the rheumatology community

Strengths: The WOMAC is a historically significant and widely used PROM that is found commonly in the literature. It has validated short-form versions and has been translated into multiple languages. It also serves as the foundation for other PROMS like the HOOS. The WOMAC subscales may also be valuable in stratifying and more thoroughly evaluating data.

Caveats and cautions: The WOMAC is proprietary, and as such, is less accessible and less attractive to potential clinicians or researchers. In addition, with 24 items, it is longer than newer, non-proprietary PROMS like the HOOS-JR. The full WOMAC is also not included in the list of PROMS recognized by the Centers for Medicaid and Medicare Services. However, the HOOS, a derivative of the WOMAC, is included on that list. As is common to all functional assessments, some activities included on the functional subscale may be impossible some patients with severe disease to complete. This may result in missing data.

Clinical/Research usability: If purchased, the WOMAC is relevant to both clinicians and researchers alike. It is reliable, valid, and responsive to treatment. It's ubiquitous presence in hip and knee outcomes literature makes further use appropriate and guarantees the ability to compare new findings to old data.

CONCLUSION

We reviewed seven of the most commonly used instruments in assessment of hip outcomes and function. There has been an extensive body of work in terms of evaluation of psychometric properties and use in clinical trials since the last review of these instruments (1). This review should be used a

reference when comparing hip outcome measures and deciding which measures should be used for research or clinical purposes for researchers and clinicians.

REFERENCES

1. 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 Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S200-207.

2. 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;51:737–755.

3. McDonough CM, Stoiber E, Tomek IM, Ni P, Kim Y-J, Tian F, et al. Sensitivity to Change of a Computer Adaptive Testing Instrument for Outcome Measurement After Hip and Knee Arthroplasty and Periacetabular Osteotomy. *J Orthop Sports Phys Ther* 2016;46:756–767.

4. Byrd JW, Jones KS. Prospective analysis of hip arthroscopy with 2-year follow-up. *Arthroscopy* 2000;16:578–587.

5. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score: Do ceiling effects limit its usefulness in orthopedics? *Acta Orthop* 2010;81:703–707.

6. Hung M, Hon SD, Cheng C, Franklin JD, Aoki SK, Anderson MB, et al. Psychometric Evaluation of the Lower Extremity Computerized Adaptive Test, the Modified Harris Hip Score, and the Hip Outcome Score. *Orthop J Sports Med* 2014;2:2325967114562191.

7. Wasko MK, Yanik EL, Pascual-Garrido C, Clohisy JC. Psychometric Properties of Patient-Reported Outcome Measures for Periacetabular Osteotomy. *J Bone Joint Surg Am* 2019;101:e21.

8. Söderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? *Clin Orthop Relat Res* 2001:189–197.

9. Kirmit L, Karatosun V, Unver B, Bakirhan S, Sen A, Gocen Z. The reliability of hip scoring systems for total hip arthroplasty candidates: assessment by physical therapists. *Clin Rehabil* 2005;19:659–661.

741 10. Kumar P, Sen R, Aggarwal S, Agarwal S, Rajnish RK. Reliability of Modified Harris Hip Score as a tool
742 for outcome evaluation of Total Hip Replacements in Indian population. *J Clin Orthop Trauma*
743 2019;10:128–130.

744 11. Garellick G, Malchau H, Herberts P. Specific or general health outcome measures in the evaluation of
745 total hip replacement. A comparison between the Harris hip score and the Nottingham Health Profile. *J*
746 *Bone Joint Surg Br* 1998;80:600–606.

747 12. Shi H-Y, Mau L-W, Chang J-K, Wang J-W, Chiu H-C. Responsiveness of the Harris Hip Score and the SF-
748 36: five years after total hip arthroplasty. *Qual Life Res* 2009;18:1053–1060.

749 13. Frihagen F, Grotle M, Madsen JE, Wyller TB, Mowinckel P, Nordsletten L. Outcome after femoral
750 neck fractures: a comparison of Harris Hip Score, Eq-5d and Barthel Index. *Injury* 2008;39:1147–1156.

751 14. Ruecker AH, Rupprecht M, Gruber M, Gebauer M, Barvencik F, Briem D, et al. The treatment of
752 intertrochanteric fractures: results using an intramedullary nail with integrated cephalocervical screws
753 and linear compression. *J Orthop Trauma* 2009;23:22–30.

754 15. Dolatowski FC, Frihagen F, Bartels S, Opland V, Šaltytė Benth J, Talsnes O, et al. Screw Fixation Versus
755 Hemiarthroplasty for Nondisplaced Femoral Neck Fractures in Elderly Patients: A Multicenter
756 Randomized Controlled Trial. *J Bone Joint Surg Am* 2019;101:136–144.

757 16. Barenius B, Inngul C, Alagic Z, Enocson A. A randomized controlled trial of cemented versus
758 cementless arthroplasty in patients with a displaced femoral neck fracture: a four-year follow-up. *Bone*
759 *Joint J* 2018;100-B:1087–1093.

760 17. Doria C, Mosele GR, Caggiari G, Puddu L, Ciurlia E. Treatment of Early Hip Osteoarthritis: Ultrasound-
761 Guided Platelet Rich Plasma versus Hyaluronic Acid Injections in a Randomized Clinical Trial. *Joints*
762 2017;5:152–155.

763 18. Winther SB, Foss OA, Husby OS, Wik TS, Klaksvik J, Husby VS. Muscular strength and function after
764 total hip arthroplasty performed with three different surgical approaches: one-year follow-up study. *Hip*
765 *Int* 2019;29:405–411.

766 19. Zomar BO, Bryant D, Hunter S, Howard JL, Vasarhelyi EM, Lanting BA. A randomised trial comparing
767 spatio-temporal gait parameters after total hip arthroplasty between the direct anterior and direct
768 lateral surgical approaches. *Hip Int* 2018;28:478–484.

769 20. Atrey A, Wolfstadt JI, Hussain N, Khoshbin A, Ward S, Shahid M, et al. The Ideal Total Hip
770 Replacement Bearing Surface in the Young Patient: A Prospective Randomized Trial Comparing Alumina
771 Ceramic-On-Ceramic With Ceramic-On-Conventional Polyethylene: 15-Year Follow-Up. *J Arthroplasty*
772 2018;33:1752–1756.

773 21. Fitzpatrick J, Bulsara MK, O'Donnell J, McCrory PR, Zheng MH. The Effectiveness of Platelet-Rich
774 Plasma Injections in Gluteal Tendinopathy: A Randomized, Double-Blind Controlled Trial Comparing a
775 Single Platelet-Rich Plasma Injection With a Single Corticosteroid Injection. *Am J Sports Med*
776 2018;46:933–939.

777 22. Tol MCJM, Bekerom MPJ van den, Sierevelt IN, Hilverdink EF, Raaymakers ELFB, Goslings JC.
778 Hemiarthroplasty or total hip arthroplasty for the treatment of a displaced intracapsular fracture in
779 active elderly patients: 12-year follow-up of randomised trial. *Bone Joint J* 2017;99-B:250–254.

780 23. Makki D, Mohamed AM, Gadiyar R, Patterson M. Addition of an anti-rotation screw to the dynamic
781 hip screw for femoral neck fractures. *Orthopedics* 2013;36:e865-868.

782 24. Madan SS, Cooper AP, Davies AG, Fernandes JA. The treatment of severe slipped capital femoral
783 epiphysis via the Ganz surgical dislocation and anatomical reduction: a prospective study. *Bone Joint J*
784 2013;95-B:424–429.

785 25. Larson CM, Giveans MR, Stone RM. Arthroscopic debridement versus refixation of the acetabular
786 labrum associated with femoroacetabular impingement: mean 3.5-year follow-up. *Am J Sports Med*
787 2012;40:1015–1021.

788 26. Söderman P. On the validity of the results from the Swedish National Total Hip Arthroplasty register.
789 *Acta Orthop Scand Suppl* 2000;71:1–33.

790 27. Hoeksma HL, Van Den Ende CHM, Ronday HK, Heering A, Breedveld FC. Comparison of the
791 responsiveness of the Harris Hip Score with generic measures for hip function in osteoarthritis of the
792 hip. *Ann Rheum Dis* 2003;62:935–938.

28. Klässbo 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.

29. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: A Short-form Hip Replacement Survey. *Clin Orthop Relat Res* 2016;474:1472–1482.

30. 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. *Osteoarthr Cartil* 2008;16:551–559.

31. 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.

32. Groot IB de, Reijman M, Terwee CB, Bierma-Zeinstra SMA, Favejee M, Roos EM, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthr Cartil* 2007;15:104–109.

33. Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson J-N, Roos EM, et al. Cross-cultural adaptation and validation of the French version of the Hip disability and Osteoarthritis Outcome Score (HOOS) in hip osteoarthritis patients. *Osteoarthr Cartil* 2010;18:522–529.

34. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 2013;41:2065–2073.

35. Bini SA, Shah RF, Bendich I, Patterson JT, Hwang KM, Zaid MB. Machine Learning Algorithms Can Use Wearable Sensor Data to Accurately Predict Six-Week Patient-Reported Outcome Scores Following Joint Replacement in a Prospective Trial. *J Arthroplasty* 2019;34:2242–2247.

36. Busch A, Jäger M, VITAS group, Wegner A, Haversath M. Vitamin E-blended versus conventional polyethylene liners in prostheses : Prospective, randomized trial with 3-year follow-up. *Orthopade* 2019.

37. Otero JE, Martin JR, Rowe TM, Odum SM, Mason JB. Radiographic and Clinical Outcomes of Modular Tapered Fluted Stems for Femoral Revision for Paprosky III and IV Femoral Defects or Vancouver B2 and B3 Femoral Fractures. *J Arthroplasty* 2019.

819 38. Lerch TD, Steppacher SD, Liechti EF, Tannast M, Siebenrock KA. One-third of Hips After
820 Periacetabular Osteotomy Survive 30 Years With Good Clinical Results, No Progression of Arthritis, or
821 Conversion to THA. *Clin Orthop Relat Res* 2017;475:1154–1168.

822 39. Li DJ, Clohisy JC, Schwabe MT, Yanik EL, Pascual-Garrido C. PROMIS Versus Legacy Patient-Reported
823 Outcome Measures in Patients Undergoing Surgical Treatment for Symptomatic Acetabular Dysplasia.
824 *Am J Sports Med* 2020:363546519894323.

825 40. Flores SE, Sheridan JR, Borak KR, Zhang AL. When Do Patients Improve After Hip Arthroscopy for
826 Femoroacetabular Impingement? A Prospective Cohort Analysis. *Am J Sports Med* 2018;46:3111–3118.

827 41. Padilla JA, Gabor JA, Schwarzkopf R, Davidovitch RI. A Novel Opioid-Sparing Pain Management
828 Protocol Following Total Hip Arthroplasty: Effects on Opioid Consumption, Pain Severity, and Patient-
829 Reported Outcomes. *J Arthroplasty* 2019;34:2669–2675.

830 42. Dietz MJ, Klein AE, Lindsey BA, Duncan ST, Eicher JM, Gillig JD, et al. Posterior Hip Precautions Do
831 Not Impact Early Recovery in Total Hip Arthroplasty: A Multicenter, Randomized, Controlled Study. *J*
832 *Arthroplasty* 2019;34:S221-S227.e1.

833 43. Banerjee S, Plummer O, Abboud JA, Deirmengian GK, Levicoff EA, Courtney PM. Accuracy and
834 Validity of Computer Adaptive Testing for Outcome Assessment in Patients Undergoing Total Hip
835 Arthroplasty. *J Arthroplasty* 2019.

836 44. Westermann RW, Lynch TS, Jones MH, Spindler KP, Messner W, Strnad G, et al. Predictors of Hip
837 Pain and Function in Femoroacetabular Impingement: A Prospective Cohort Analysis. *Orthop J Sports*
838 *Med* 2017;5:2325967117726521.

839 45. Peters RM, Beers LWAH van, Steenbergen LN van, Wolkenfelt J, Ettema HB, Ten Have BLEF, et al.
840 Similar Superior Patient-Reported Outcome Measures for Anterior and Posterolateral Approaches After
841 Total Hip Arthroplasty: Postoperative Patient-Reported Outcome Measure Improvement After 3 months
842 in 12,774 Primary Total Hip Arthroplasties Using the Anterior, Anterolateral, Straight Lateral, or
843 Posterolateral Approach. *J Arthroplasty* 2018;33:1786–1793.

844 46. Gagnier JJ, Huang H, Mullins M, Marinac-Dabić D, Ghambaryan A, Eloff B, et al. Measurement
845 Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Hip Arthroplasty:
846 A Systematic Review. *JBJS Rev* 2018;6:e2.

847 47. Veenhof C, Bijlsma JWJ, Ende CHM van den, Dijk GM van, Pisters MF, Dekker J. Psychometric
848 evaluation of osteoarthritis questionnaires: a systematic review of the literature. *Arthritis Rheum*
849 2006;55:480–492.

850 48. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes
851 Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported
852 health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010;63:1179–1194.

853 49. Ho B, Houck JR, Flemister AS, Ketz J, Oh I, DiGiovanni BF, et al. Preoperative PROMIS Scores Predict
854 Postoperative Success in Foot and Ankle Patients. *Foot Ankle Int* 2016;37:911–918.

855 50. Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN. Computerized Adaptive Testing Using the
856 PROMIS Physical Function Item Bank Reduces Test Burden With Less Ceiling Effects Compared With the
857 Short Musculoskeletal Function Assessment in Orthopaedic Trauma Patients. *J Orthop Trauma*
858 2014;28:439–443.

859 51. Hung M, Saltzman CL, Greene T, Voss MW, Bounsanga J, Gu Y, et al. Evaluating instrument
860 responsiveness in joint function: The HOOS JR, the KOOS JR, and the PROMIS PF CAT. *J Orthop Res*
861 2018;36:1178–1184.

862 52. Stiegel KR, Lash JG, Peace AJ, Coleman MM, Harrington MA, Cahill CW. Early Experience and Results
863 Using Patient-Reported Outcomes Measurement Information System Scores in Primary Total Hip and
864 Knee Arthroplasty. *J Arthroplasty* 2019;34:2313–2318.

865 53. Beleckas CM, Guattery J, Chamberlain AM, Khan T, Kelly MP, Calfee RP. Using Patient-reported
866 Outcomes Measurement Information System Measures to Understand the Relationship Between
867 Improvement in Physical Function and Depressive Symptoms. *J Am Acad Orthop Surg* 2018;26:e511–
868 e518.

869 54. Lawrie CM, Abu-Amer W, Barrack RL, Clohisy JC. Is the Patient-Reported Outcome Measurement
870 Information System Feasible in Bundled Payment for Care Improvement in Total Hip Arthroplasty
871 Patients? *J Arthroplasty* 2019.

872 55. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, Stone AA. Validity and reliability of Patient-
873 Reported Outcomes Measurement Information System (PROMIS) Instruments in Osteoarthritis. *Arthritis*
874 *Care Res (Hoboken)* 2013;65:1625–1633.

875 56. Martusiewicz A, Delagrammaticas D, Harold RE, Bhatt S, Beal MD, Manning DW. Anterior versus
876 posterior approach total hip arthroplasty: patient-reported and functional outcomes in the early
877 postoperative period. *Hip Int* 2019:1120700019881413.

878 57. Austin DC, Torchia MT, Moschetti WE, Jevsevar DS, Keeney BJ. Patient outcomes after total hip
879 arthroplasty in extreme elderly patients older than 80 years. *Hip Int* 2019:1120700019837943.

880 58. Sheean AJ, Schmitz MR, Ward CL, Barrow AE, Tennent DJ, Roach CJ, et al. Assessment of Disability
881 Related to Femoroacetabular Impingement Syndrome by Use of the Patient-Reported Outcome
882 Measure Information System (PROMIS) and Objective Measures of Physical Performance. *Am J Sports*
883 *Med* 2017;45:2476–2482.

884 59. Nwachukwu BU, Beck EC, Chapman R, Chahla J, Okoroha K, Nho SJ. Preoperative Performance of the
885 PROMIS in Patients Undergoing Hip Arthroscopic Surgery for Femoroacetabular Impingement Syndrome.
886 *Orthop J Sports Med* 2019;7:2325967119860079.

887 60. Livermore AT, Anderson LA, Anderson MB, Erickson JA, Peters CL. Correction of mildly dysplastic hips
888 with periacetabular osteotomy demonstrates promising outcomes, achievement of correction goals, and
889 excellent five-year survivorship. *Bone Joint J* 2019;101-B:16–22.

890 61. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total
891 hip replacement. *J Bone Joint Surg Br* 1996;78:185–190.

892 62. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford hip and
893 knee scores. *J Bone Joint Surg Br* 2007;89:1010–1014.

894 63. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total
895 knee replacement. *J Bone Joint Surg Br* 1998;80:63–69.

64. Baker PN, Meulen JH van der, Lewsey J, Gregg PJ, National Joint Registry for England and Wales. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. *J Bone Joint Surg Br* 2007;89:893–900.

65. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. *J Bone Joint Surg Br* 2001;83:339–344.

66. Pynsent PB, Adams DJ, Disney SP. The Oxford hip and knee outcome questionnaires for arthroplasty. *J Bone Joint Surg Br* 2005;87:241–248.

67. Rothwell AG, Hooper GJ, Hobbs A, Frampton CM. An analysis of the Oxford hip and knee scores and their relationship to early joint revision in the New Zealand Joint Registry. *J Bone Joint Surg Br* 2010;92:413–418.

68. Uesugi Y, Makimoto K, Fujita K, Nishii T, Sakai T, Sugano N. Validity and responsiveness of the Oxford hip score in a prospective study with Japanese total hip arthroplasty patients. *J Orthop Sci* 2009;14:35–39.

69. Wylde V, Blom AW, Whitehouse SL, Taylor AH, Pattison GT, Bannister GC. Patient-reported outcomes after total hip and knee arthroplasty: comparison of midterm results. *J Arthroplasty* 2009;24:210–216.

70. Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br* 2001;83:1125–1129.

71. Field RE, Cronin MD, Singh PJ. The Oxford hip scores for primary and revision hip replacement. *J Bone Joint Surg Br* 2005;87:618–622.

72. Yeoman TFM, Clement ND, Macdonald D, Moran M. Recall of preoperative Oxford Hip and Knee Scores one year after arthroplasty is an alternative and reliable technique when used for a cohort of patients. *Bone Joint Res* 2018;7:351–356.

921 73. Eibich P, Dakin HA, Price AJ, Beard D, Arden NK, Gray AM. Associations between preoperative Oxford
922 hip and knee scores and costs and quality of life of patients undergoing primary total joint replacement
923 in the NHS England: an observational study. *BMJ Open* 2018;8:e019477.

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

926 75. Scott CEH, Turnbull GS, Powell-Bowns MFR, MacDonald DJ, Breusch SJ. Activity levels and return to
927 work after revision total hip and knee arthroplasty in patients under 65 years of age. *Bone Joint J*
928 2018;100-B:1043–1053.

929 76. Abdel Messih M, Naylor JM, Descallar J, Manickam A, Mittal R, Harris IA. Mail versus telephone
930 administration of the Oxford Knee and Hip Scores. *J Arthroplasty* 2014;29:491–494.

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

934 78. Hamilton DF, Loth FL, MacDonald DJ, Giesinger K, Patton JT, Simpson AH, et al. Treatment Success
935 Following Joint Arthroplasty: Defining Thresholds for the Oxford Hip and Knee Scores. *J Arthroplasty*
936 2018;33:2392–2397.

937 79. Hamilton DF, Giesinger JM, Patton JT, MacDonald DJ, Simpson AHRW, Howie CR, et al. Making the
938 Oxford Hip and Knee Scores meaningful at the patient level through normative scoring and registry data.
939 *Bone Joint Res* 2015;4:137–144.

940 80. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, et al. Meaningful changes for the Oxford
941 hip and knee scores after joint replacement surgery. *J Clin Epidemiol* 2015;68:73–79.

942 81. Neufeld ME, Masri BA. Can the Oxford Knee and Hip Score identify patients who do not require total
943 knee or hip arthroplasty? *Bone Joint J* 2019;101-B:23–30.

944 82. McMurray R, Heaton J, Sloper P, Nettleton S. Measurement of patient perceptions of pain and
945 disability in relation to total hip replacement: the place of the Oxford hip score in mixed methods. *Qual*
946 *Health Care* 1999;8:228–233.

947 83. Wylde V, Learmonth ID, Cavendish VJ. The Oxford hip score: the patient's perspective. *Health Qual*
948 *Life Outcomes* 2005;3:66.

949 84. Delaunay C, Epinette J-A, Dawson J, Murray D, Jolles B-M. Cross-cultural adaptations of the Oxford-
950 12 HIP score to the French speaking population. *Orthop Traumatol Surg Res* 2009;95:89–99.

951 85. Nourbakhsh M, Zarezadeh A, Shemshaki H, Etemadifar MR, Moezi M, Mazoochian F. Translation and
952 cultural adaptation of the oxford hip score for Iranian population. *Int J Prev Med* 2013;4:141–145.

953 86. Lee J, Ozdoganlar OB, Rabin Y. An experimental investigation on thermal exposure during bone
954 drilling. *Med Eng Phys* 2012;34:1510–1520.

955 87. Fitzpatrick R, Norquist JM, Jenkinson C, Reeves BC, Morris RW, Murray DW, et al. A comparison of
956 Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery.
957 *Qual Life Res* 2004;13:331–338.

958 88. Fitzgerald AM, Gaston P, Wilson Y, Quaba A, McQueen MM. Long-term sequelae of fasciotomy
959 wounds. *Br J Plast Surg* 2000;53:690–693.

960 89. Péchon PHM, Butler K, Murphy G, Singer GC. The perils of PROMs: question 5 of the Oxford Hip
961 Score is ambiguous to 10% of English-speaking patients: a survey of 135 patients. *Hip Int* 2019;29:299–
962 302.

963 90. Ostendorf M, Stel HF van, Buskens E, Schrijvers AJP, Marting LN, Verbout AJ, et al. Patient-reported
964 outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br*
965 2004;86:801–808.

966 91. Impellizzeri FM, Mannion AF, Naal FD, Leunig M. Validity, reproducibility, and responsiveness of the
967 oxford hip score in patients undergoing surgery for femoroacetabular impingement. *Arthroscopy*
968 2015;31:42–50.

969 92. Gosens T, Hoefnagels NHM, Vet RCW de, Dhert WJA, Langelaan EJ van, Bulstra SK, et al. The "Oxford
970 Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total
971 hip arthroplasty. *Acta Orthop* 2005;76:204–211.

93. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, et al. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Qual Health Care* 2000;9:146–150.

94. Kwong E, Neuburger J, Black N. Agreement between retrospectively and contemporaneously collected patient-reported outcome measures (PROMs) in hip and knee replacement patients. *Qual Life Res* 2018;27:1845–1854.

95. Garbuz DS, Xu M, Sayre EC. Patients' outcome after total hip arthroplasty: a comparison between the Western Ontario and McMaster Universities index and the Oxford 12-item hip score. *J Arthroplasty* 2006;21:998–1004.

96. Weel H, Lindeboom R, Kuipers SE, Vervest TMJS. Comparison between the Harris- and Oxford Hip Score to evaluate outcomes one-year after total hip arthroplasty. *Acta Orthop Belg* 2017;83:98–109.

97. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care* 1996;5:81–88.

98. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation--value in comparison with other assessment tests. *Scand J Rheumatol Suppl* 1987;65:85–89.

99. Lequesne M. Indices of severity and disease activity for osteoarthritis. *Semin Arthritis Rheum* 1991;20:48–54.

100. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. *J Rheumatol* 1997;24:779–781.

101. Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W, et al. Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-Algofunctional index in patients with knee and hip osteoarthritis. *Osteoarthr Cartil* 1998;6:79–86.

102. Dawson J, Linsell L, Doll H, Zondervan K, Rose P, Carr A, et al. Assessment of the Lequesne index of severity for osteoarthritis of the hip in an elderly population. *Osteoarthr Cartil* 2005;13:854–860.

998 103. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. Requirement for
999 total hip arthroplasty: an outcome measure of hip osteoarthritis? *J Rheumatol* 1999;26:855–861.

1000 104. Grimmig H, Melzer C, Ludwig F-J, Daalmann HH. [Using the Lequesne questionnaires for patients
1001 with hip and knee disabilities in daily routine to reflect the quality of outcome in rehabilitation]. *Z*
1002 *Orthop Ihre Grenzgeb* 2002;140:452–459.

1003 105. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional
1004 assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the
1005 WOMAC and Lequesne indexes. *Osteoarthr Cartil* 2002;10:602–610.

1006 106. Lequesne MG, Samson M. Indices of severity in osteoarthritis for weight bearing joints. *J Rheumatol*
1007 *Suppl* 1991;27:16–18.

1008 107. Basaran S, Guzel R, Seydaoglu G, Guler-Uysal F. Validity, reliability, and comparison of the WOMAC
1009 osteoarthritis index and Lequesne algofunctional index in Turkish patients with hip or knee
1010 osteoarthritis. *Clin Rheumatol* 2010;29:749–756.

1011 108. Bae SC, Lee HS, Yun HR, Kim TH, Yoo DH, Kim SY. Cross-cultural adaptation and validation of Korean
1012 Western Ontario and McMaster Universities (WOMAC) and Lequesne osteoarthritis indices for clinical
1013 research. *Osteoarthr Cartil* 2001;9:746–750.

1014 109. Espinosa-Cuervo G, Guillermin F, Rat A-C, Duarte-Salazar C, Alemán-Hernández S-I, Vergara-Álvarez
1015 Y, et al. Transculturalization and validation of a Spanish translation of the specific lower limb
1016 osteoarthritis and quality of life questionnaire AMICAL: Arthrose des Membres Inférieurs et Qualité de
1017 vie AMIQUAL. *Reumatol Clin* 2014;10:241–247.

1018 110. Konstantinidis GA, Aletras VH, Kanakari K-A, Natsis K, Bellamy N, Niakas D. Comparative validation
1019 of the WOMAC osteoarthritis and Lequesne algofunctional indices in Greek patients with hip or knee
1020 osteoarthritis. *Qual Life Res* 2014;23:539–548.

1021 111. Nadrian H, Moghimi N, Nadrian E, Moradzadeh R, Bahmanpour K, Iranpour A, et al. Validity and
1022 reliability of the Persian versions of WOMAC Osteoarthritis Index and Lequesne Algofunctional Index.
1023 *Clin Rheumatol* 2012;31:1097–1102.

1024 112. Basaran S, Guzel R, Seydaoglu G, Guler-Uysal F. Validity, reliability, and comparison of the WOMAC
1025 osteoarthritis index and Lequesne algofunctional index in Turkish patients with hip or knee
1026 osteoarthritis. *Clin Rheumatol* 2010;29:749–756.

1027 113. Ludwig FJ, Melzer C, Grimmig H, Daalmann HH. [Cross cultural adaptation of the lequesne
1028 algofunctional indices for german speaking patients with osteoarthritis of the hip and the knee].
1029 *Rehabilitation (Stuttg)* 2002;41:249–257.

1030 114. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Elke R, et al. [Evaluation of a German
1031 questionnaire version of the Lequesne cox- and gonarthrosis indices]. *Z Rheumatol* 1996;55:50–57.

1032 115. Santos JPM, Silva RA da, Fernandes MTP, Poli-Frederico RC, Santos DC, Andraus RAC, et al. Use of
1033 the Lower Extremity Functional Scale (LEFS-Brazil) questionnaire compared to Lequesne Algofunctional
1034 Index for definition of knee and hip osteoarthritis severity. *Rev Bras Reumatol Engl Ed* 2017;57:274–277.

1035 116. Theiler R, Sangha O, Schaeren S, Michel BA, Tyndall A, Dick W, et al. Superior responsiveness of the
1036 pain and function sections of the Western Ontario and McMaster Universities Osteoarthritis Index
1037 (WOMAC) as compared to the Lequesne-Algofunctional Index in patients with osteoarthritis of the lower
1038 extremities. *Osteoarthr Cartil* 1999;7:515–519.

1039 117. Johanson NA, Liang MH, Daltroy L, Rudicel S, Richmond J. American Academy of Orthopaedic
1040 Surgeons lower limb outcomes assessment instruments. Reliability, validity, and sensitivity to change. *J*
1041 *Bone Joint Surg Am* 2004;86:902–909.

1042 118. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American academy of orthopaedic
1043 surgeons outcomes instruments: normative values from the general population. *J Bone Joint Surg Am*
1044 2002;84:208–215.

1045 119. DeLullo JA, Thomas E, Cooney TE, McConnell SJ, Sanders JO. Femoral remodeling may influence
1046 patient outcomes in slipped capital femoral epiphysis. *Clin Orthop Relat Res* 2007;457:163–170.

1047 120. Kim S-J, Balce GC, Agashe MV, Song S-H, Song H-R. Is bilateral lower limb lengthening appropriate
1048 for achondroplasia?: midterm analysis of the complications and quality of life. *Clin Orthop Relat Res*
1049 2012;470:616–621.

121. Polfer EM, Hoyt BW, Bevevino AJ, Forsberg JA, Potter BK. Knee Disarticulations Versus Transfemoral Amputations: Functional Outcomes. *J Orthop Trauma* 2019;33:308–311.

122. Anon. Measure Methodology | CMS. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology. Accessed March 5, 2020.

123. 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;15:1833–1840.

124. Bellamy N, Buchanan WW, Grace E. Double-blind randomized controlled trial of isoxicam vs piroxicam in elderly patients with osteoarthritis of the hip and knee. *Br J Clin Pharmacol* 1986;22 Suppl 2:149S-155S.

125. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum* 2001;45:453–461.

126. Bellamy N. Outcome measurement in osteoarthritis clinical trials. *J Rheumatol Suppl* 1995;43:49–51.

127. Ornetti P, Dougados M, Paternotte S, Logeart I, Gossec L. Validation of a numerical rating scale to assess functional impairment in hip and knee osteoarthritis: comparison with the WOMAC function scale. *Ann Rheum Dis* 2011;70:740–746.

128. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S208-228.

129. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. *J Rheumatol* 1997;24:2413–2415.

130. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S, et al. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. *J Clin Epidemiol* 2011;64:182–190.

131. Theiler R, Spielberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version. *Osteoarthr Cartil* 2002;10:479–481.

132. Marsh JD, Bryant DM, Macdonald SJ, Naudie DDR. Patients respond similarly to paper and electronic versions of the WOMAC and SF-12 following total joint arthroplasty. *J Arthroplasty* 2014;29:670–673.

133. Bruce B, Fries J. Longitudinal comparison of the Health Assessment Questionnaire (HAQ) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). *Arthritis Rheum* 2004;51:730–737.

134. Bellamy N, Wilson C, Hendrikz J. Population-based normative values for the Western Ontario and McMaster (WOMAC) osteoarthritis index and the Australian/Canadian (AUSCAN) hand osteoarthritis index functional subscales. *Inflammopharmacology* 2010;18:1–8.

135. Liebs TR, Herzberg W, Gluth J, Rüther W, Haasters J, Russlies M, et al. Using the patient's perspective to develop function short forms specific to total hip and knee replacement based on WOMAC function items. *Bone Joint J* 2013;95-B:239–243.

136. Guermazi M, Poiraudeau S, Yahia M, Mezganni M, Fermanian J, Habib Elleuch M, et al. Translation, adaptation and validation of the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) for an Arab population: the Sfax modified WOMAC. *Osteoarthr Cartil* 2004;12:459–468.

137. Xie F, Li S-C, Goeree R, Tarride J-E, O'Reilly D, Lo N-N, et al. Validation of Chinese Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in patients scheduled for total knee replacement. *Qual Life Res* 2008;17:595–601.

138. Soininen JV, Paavolainen PO, Gronblad MA, Kaapa EH. Validation study of a Finnish version of the Western Ontario and McMasters University osteoarthritis index. *Hip Int* 2008;18:108–111.

139. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, et al. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. *Z Rheumatol* 1996;55:40–49.

140. Wigler I, Neumann L, Yaron M. Validation study of a Hebrew version of WOMAC in patients with osteoarthritis of the knee. *Clin Rheumatol* 1999;18:402–405.

141. Salaffi F, Leardini G, Canesi B, Mannoni A, Fioravanti A, Caporali R, et al. Reliability and validity of the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index in Italian patients with osteoarthritis of the knee. *Osteoarthr Cartil* 2003;11:551–560.

142. Faik A, Benbouazza K, Amine B, Maaroufi H, Bahiri R, Lazrak N, et al. Translation and validation of Moroccan Western Ontario and McMaster Universities (WOMAC) osteoarthritis index in knee osteoarthritis. *Rheumatol Int* 2008;28:677–683.

143. Hashimoto H, Hanyu T, Sledge CB, Lingard EA. Validation of a Japanese patient-derived outcome scale for assessing total knee arthroplasty: comparison with Western Ontario and McMaster Universities osteoarthritis index (WOMAC). *J Orthop Sci* 2003;8:288–293.

144. Thumboo J, Chew LH, Soh CH. Validation of the Western Ontario and Mcmaster University osteoarthritis index in Asians with osteoarthritis in Singapore. *Osteoarthr Cartil* 2001;9:440–446.

145. Escobar A, Quintana JM, Bilbao A, Azkárate J, Güenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clin Rheumatol* 2002;21:466–471.

146. Roos EM, Klässbo M, Lohmander LS. WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and MacMaster Universities. *Scand J Rheumatol* 1999;28:210–215.

147. Söderman P, Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index: a self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). *Acta Orthop Scand* 2000;71:39–46.

148. Kuptniratsaikul V, Rattanachaiyanont M. Validation of a modified Thai version of the Western Ontario and McMaster (WOMAC) osteoarthritis index for knee osteoarthritis. *Clin Rheumatol* 2007;26:1641–1645.

149. Burgers PTPW, Poolman RW, Van Bakel TMJ, Tuinebreijer WE, Zielinski SM, Bhandari M, et al. Reliability, validity, and responsiveness of the Western Ontario and McMaster Universities Osteoarthritis Index for elderly patients with a femoral neck fracture. *J Bone Joint Surg Am* 2015;97:751–757.

150. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthr Cartil* 2005;13:1076–1083.

151. Naal FD, Impellizzeri FM, Eisenhart-Rothe R von, Mannion AF, Leunig M. Reproducibility, validity, and responsiveness of the hip outcome score in patients with end-stage hip osteoarthritis. *Arthritis Care Res (Hoboken)* 2012;64:1770–1775.

152. Bilbao A, Quintana JM, Escobar A, Las Hayas C, Orive M. Validation of a proposed WOMAC short form for patients with hip osteoarthritis. *Health Qual Life Outcomes* 2011;9:75.

153. SooHoo NF, Li Z, Chenok KE, Bozic KJ. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. *J Arthroplasty* 2015;30:176–191.

154. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29–33.

155. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol* 2000;27:2635–2641.

1152 **Table 1: Practical applications**

1153

Measure	Number of items	Content/ Domains	Method of administration	Recall period	Response format	Range of scores	Score interpretation	Availability of normative data	Cross-cultural validation
Harris Hip Score (HHS)	10, 8 in the modified-HHS	Pain, function, absence of deformity and ROM	Clinician-based. Modified-HHS is self-administered	Not described	Numerical scale within each domain	0-100	<70 indicated poor result, 70-80 fair, 80-90 good, 90-100 excellent	No	Used internationally, but no validated versions in other languages
Hip Disability and Osteoarthritis Score (HOOS)	40. HOOS-JR has 6 items, HOOS-PS has 5 items	Pain, other symptoms, function in ADLs, function in sport/rec, QOL	Self	Last week	5 Likert boxes	0-100	0 indicates extreme problems, 100 indicates no problems	No	25 total languages including English, French, German, Dutch, Italian, Spanish
Patient-Reported Outcomes Measurement Information System (PROMIS)	4-12. Vary depending on exact measure chosen.	Physical health, mental health, social health	Self	7-day. Some questions present-tense	0-10 for pain. Five options for other domains	T-score metric with 50 as mean of population	50 is mean of population, 10 is one SD of population. 60 indicates one SD above mean	Yes	45 total languages including English, French, German, Dutch, Italian, Spanish

Oxford Hip Score (OHS)	12	Pain, function	Self or over the phone	Last 4 weeks	0-4 (worst to best)	0-48	0 indicates severe problems, 48 indicates no problems	No	English, Japanese, Dutch, German, Turkish, Spanish, Mandarin, Italian. Used, but not validated in French, Iranian, and Korean
Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH)	11	Osteoarthritis-related symptoms, function	Self-administered, interviewer-based, or clinician-administered	Not specified	0-8 (best to worst)	0-28	0 indicates no handicap, >11-13 indicates severe handicap	No	English, French, German, Turkish, Korean, Spanish, Greek, Persian, Portuguese
American Academy of Orthopedic Surgeons (AAOS) Hip and Knee questionnaire	7	Pain, stiffness, swelling, function	Self	Last week	Likert scale for all domains with slight differences in options	0-100	0 indicates most disability, 100 indicates no disability	Yes	English, Spanish
Western Ontario McMaster Universities Osteoarthritis Index (WOMAC)	24	Pain, stiffness, difficulty with certain activities	Self or interview-administered	Last 48 hours	Likert scale, Visual analog scale (VAS), and 11-box numerical rating scales	Pain: 0-20, Stiffness: 0-8, Function: 0-68. VAS totaled to 2400	Lower scores indicate less pain and disability	Yes	English, Arabic, Chinese, Dutch, Finnish, German, Hebrew, Italian, Japanese, Korean, Moroccan, Singapore, Spanish, Swedish, Thai, Turkish

1154

1155

1156

1157 **Table 2: Psychometrics**

1158

Measure	Floor, ceiling effects	Reliability	Validity	Responsiveness	Minimally important differences	Generalizability	Used in RCTs
Harris Hip Score (HHS)	Ceiling effects ~20%. Floor effects not reported	Internal consistency: Cronbach's alpha score 0.94. Test-test reliability: r=0.93-0.98	Content and construct validity correlates with similar domains of other scores	Responsive to pain, function at six-month follow-up, weakly responsive at 2-year follow-up. mHHS adequately responsive after PAO	Standardized response mean SRM) 0.75	Widely generalizable to adult patients with hip disabilities or undergoing hip surgeries	Yes
Hip Disability and Osteoarthritis Score (HOOS)	Floor effects 4.1-17.8%. Ceiling effects up to 19%	Internal consistency: Cronbach's alpha 0.82-0.98. Test-test 0.75-0.97	Content validity performed by asking patients rating item importance. Construct validity confirmed by correlation with other scores	Good responsiveness in short term and long-term follow-up up to 2-years post-operatively	9.0-9.6 in ADL domain, 16.2-19.0 in QOL domain	Widely generalizable to adult patients with hip disabilities or undergoing hip procedures. HOOS-JR developed for THA patients	Yes
Patient-Reported Outcomes Measurement Information System	Not observed in PAO patients. Floor effects in THA patients 20-45% in depression	Internal consistency: Cronbach's alpha 1.00	Not evaluated	High responsiveness post-operatively in THA patients	5 (half of normalized SD of reported scores)	Generalizable to adult patients with hip ailments or undergoing hip procedures. Separate	Overall less frequently used in RCTs, but some recent use

(PROMIS)	domain, 21-26% in pain domain. No ceiling effects in THA patients					PROMIS instruments available for pediatric patients	
Oxford Hip Score (OHS)	Ceiling effects 13.5%, low floor effects	Internal consistency: Cronbach's alpha: 0.84-0.93	High correlation with content validity when compared to other PROMs	High responsiveness compared to generic measures	Estimated to be 3-5 points (0-48 point scale)	Widely generalizable to adult patients with hip disabilities or undergoing hip surgeries	Yes
Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH)	Not reported	Internal consistency: Cronbach's alpha 0.83-0.94 for composite score, pain/discomfort lower at 0.63. Test-test 0.51-0.96	Construct and convergent validity inferior to other PROMs	Not reported	Unknown	Not recommended as sole outcome measure clinically given poor validity and reliability. May have some utility in larger patient populations	Yes
American Academy of Orthopedic Surgeons (AAOS) Hip and Knee questionnaire	Not measured	Internal consistency: Cronbach's alpha: 0.8. Test-test 0.91	Construct validity with moderate to high correlation to other PROMs	Not directly measured	Unknown	Widely generalizable to adult patients with hip disabilities or undergoing hip surgeries	Little use in major clinical trials
Western Ontario McMaster Universities Osteoarthritis Index (WOMAC)	Floor effects generally thought of as minimal, but one report of significant floor effects in pain and stiffness after THA (25-46%). Ceiling	Internal consistency: Cronbach's alpha 0.86-0.97. Test-test 0.96	Good construct validity compared to other PROMs	High SRM (>1.0)	MCID after THA 10.2, after PAO 11, 7.9 in non-op management of hip OA (Scale of 0-68)	Widely generalizable to adult patients with hip disabilities or undergoing hip surgeries. Currently being assessed for validity in pediatric	Yes

effects more common up to 53%

population

1159