Published in final edited form as:
J Am Acad Orthop Surg. 2013 ; 21(0 1): S39–S46. doi:10.5435/JAAOS-21-07-S39.

Clinical Outcomes Assessment in Clinical Trials to Assess Treatment of Femoroacetabular Impingement: Use of Patient-reported Outcome Measures

Marcie Harris-Hayes, PT, DPT, MSCI, OCS, **Christine M. McDonough, PT, PhD, MS**, **Michael Leunig, MD**, **Cara Beth Lee, MD**, **John J. Callaghan, MD**, and **Ewa M. Roos, PT, PhD**
Program in Physical Therapy, Washington University School of Medicine, St. Louis, MO (Dr. Harris-Hayes), The Health & Disability Research Institute, Boston University School of Public Health, Boston, MA (Dr. McDonough), the Department of Orthopaedics, Schulthess Clinic, Zurich, Switzerland (Dr. Leunig), the Department of Orthopedic Surgery, Virginia Mason Medical Center, Seattle, WA (Dr. Lee), the University of Iowa and the Iowa City VA Medical Center, Iowa City, IA (Dr. Callaghan), and the Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark (Dr. Roos)

Abstract

Patient-reported outcome measures are an important component of outcomes assessment in clinical trials to assess the treatment of femoroacetabular impingement (FAI). This review of disease-specific measures and instruments used to assess the generic quality of life and physical activity levels of patients with FAI found no conclusive evidence to support a single disease-specific questionnaire. Using a systematic review of study methodology, the Copenhagen Hip and Groin Outcome Score and the 33-item International Hip Outcome Tool scored the best. Nevertheless, both of these instruments were developed recently and have not been established in the literature. Although currently used generic and activity-level measures have limitations, as well, they should be considered, depending on the specific goals of the study. Additional research is needed to assess the properties of these measures fully when used to evaluate patients with FAI.

Patient-reported outcome measures (PROs) often are the preferred primary outcome metrics to assess symptom modification in clinical trials. They are an important component of outcomes assessment because they represent the patient's health status as assessed by the patient, without interpretation of the healthcare provider.[1] To be useful, PROs must be reliable, valid, responsive, and representative of the patient population of interest.

This article provides recommendations for the PROs to be used in clinical trials investigating the efficacy of treatments for femoroacetabular impingement (FAI). It describes and provides quality ratings for disease-specific PROs developed for young–to–middle-aged adults with hip pain and dysfunction and presents common instruments to assess generic quality of life (QOL) and physical activity levels. Perspectives on future relevant directions and methodologies, such as computer-adaptive testing (CAT), are discussed, as well.

Dr. Leunig or an immediate family member serves as a paid consultant to Smith & Nephew and has stock or stock options held in Pivot Medical. Dr. Callaghan or an immediate family member has received royalties from DePuy. Dr. Roos or an immediate family member serves as a board member, owner, officer, or committee member of the Osteoarthritis Research Society International. None of the following authors or any immediate family member has received anything of value from or has stock or stock options held in a commercial company or institution related directly or indirectly to the subject of this article: Dr. Harris-Hayes, Dr. McDonough, and Dr. Lee.

Disease-specific Measures for Femoroacetabular Impingement

Several measures have been reported in the FAI literature[2–5] (Table 1). In this review, however, we used stringent criteria for instrument selection and therefore only those disease-specific PROs are included in which content validity was ensured through input from patients of similar age, sex, and activity level who had experienced symptoms and limitations due to FAI. Accordingly, we excluded instruments such as the Hip Outcome Score that although specifically developed and validated for impingement patients undergoing hip arthroscopy did not involve the patient opinion in the developmental process. Therefore, based on previous reviews of the literature[10–13] and the authors' collective knowledge, focus was placed on three disease-specific PROs that explicitly included young to middle-aged adults in the development of the measures.

Each PRO described below is patient administered, with a user-friendly format that requires ≤10 minutes to complete. All of them are self-explanatory and can be administered in the waiting room or mailed so that the patient can complete it at home. The quality of each PRO was assessed using the COnsensus-based Standards for the selection of health Measurement INstrument (COSMIN) checklist.[14] See Table 2 for COSMIN summary ratings.

Copenhagen Hip and Groin Outcome Score

The Copenhagen Hip and Groin Outcome Score (HAGOS) was developed in 2011[6] using the COSMIN recommendations to achieve the best possible quality of the instrument and the clinical study.[15,16] The HAGOS is a quantitative measure of hip and groin disability based on the different levels of the International Classification of Functioning, Disability, and Health. The HAGOS content validity was ensured through a systematic literature review, interviews with 25 Danish patients with hip and/or groin pain, and an expert panel, as well as by testing 101 physically active Danish patients (50 women) with a mean age of 36 years (range, 18 to 63 years) who sought medical care because of hip and/or groin pain.

The HAGOS consists of six separately scored subscales: pain, other symptoms, physical function in daily living, function in sport and recreation, participation in physical activities, and hip-related QOL. Test-retest reliability was substantial, with intraclass correlation coefficients (ICCs) ranging from 0.82 to 0.91 for the six subscales. The smallest detectable change ranged from 2.7 to 5.2 points at the group level for the different subscales indicating that changes greater than 5.2 are detectable for all subscales. Construct validity and responsiveness were confirmed with statistically significant correlation coefficients of 0.37 to 0.73 ($P < 0.01$) for convergent construct validity and from 0.56 to 0.69 ($P < 0.01$) for responsiveness.

The past week is taken into consideration when answering the questions. Standardized answer options are given in five Likert boxes, and each question is scored from zero to 4. A normalized score is calculated for each subscale, with 100 indicating no symptoms and zero indicating extreme symptoms. The HAGOS is meant to be used over short and long time intervals, to assess changes from week to week induced by treatment such as medication, surgery, or physical therapy and to assess changes over a period of years due to primary or posttraumatic injuries. The result can be plotted as an outcome profile. The HAGOS currently is available in two language versions: Danish and English. These and other upcoming language versions are available at the website www.koos.nu.

33-item International Hip Outcome Tool

The 33-item International Hip Outcome Tool (iHOT-33) was developed in 2012 by members of the Multicenter Arthroscopy of the Hip Outcomes Research Network

(MAHORN).[7] More than 400 active adult patients of both sexes ranging in age from 16 to 60 years with hip joint pathology were recruited from MAHORN members' practices in Canada, England, Switzerland, and the United States to participate in various phases of iHOT-33 development and testing. Face validity and content validity were established by involving patients, surgeons, and physical therapists during item development and item reduction. Test-retest reliability was moderate to good, with an ICC of 0.78. Convergent construct validity was confirmed with a statistically significant correlation coefficient of 0.81, compared with the Nonarthritic Hip Score (NAHS). The minimal clinically important difference after hip arthroscopy was calculated to be 6 points. The ICC of 0.78 indicates that the iHOT-33 cannot reliably detect this suggested minimally clinical important difference.

The past month is taken into consideration when answering the questions. Each question is scored using a 100-point visual analog scale, and a total score is calculated, with 100 indicating the best possible score. The iHOT-33 subscales include symptoms and functional limitations; sports and recreational physical activities; job-related concerns; and social, emotional, and lifestyle concerns. These subscales are not intended for individual use and have not been validated for use as subscales, however. A shorter version, the iHOT-12, recently was introduced for clinical use.[17]

Nonarthritic Hip Score

The NAHS was developed in 2003[8] by modifying the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC),[4] a tool that was originally developed to assess symptoms and function in patients with arthritis. Unlike the WOMAC, the NAHS includes questions related to mechanical symptoms and physical activities relevant to the relatively younger, more active patient. To ensure content validity, patients, surgeons, physical therapists, and epidemiologists were involved in creating the questionnaire.

A total of 48 patients ranging in age from 16 to 45 years, 62% of whom were women, participated in various phases of testing.[8] Test-retest reliability, assessed using a Pearson correlation coefficient, was 0.96 for the total score and ranged from 0.87 to 0.95 for each of the four subscales (ie, pain, mechanical symptoms, physical function, activity level). Convergent construct validity was confirmed with a statistically significant correlation coefficient of 0.82, compared with the Harris Hip Score. Responsiveness was not reported.

The questions are meant to assess patient factors in the previous 48 hours. Standardized answer options are given in five Likert boxes, and each question is scored from 0 to 4. A normalized score is calculated for the total score for each subscale, with a score of 100 indicating normal hip function.

Recommendation

Based on the authors' review, no conclusive evidence exists to support a single questionnaire for use in all patients with FAI. Although the aforementioned PROs are promising, further investigation is needed into the properties of these PROs. Future studies and head-to-head comparisons are needed to determine whether one particular PRO is superior. Investigators should consider using subscale scores in addition to the overall total score. Keeping constructs such as pain, function, and QOL in separate subscales may reduce the number of patients needed in clinical trials and aid in the clinical interpretation of the results. Although all PROs reviewed have subscales, only the HAGOS and the NAHS have been validated for use as separate subscales.

Generic Outcome Measures

Generic outcome measures are health-related QOL instruments that are suitable for use in the general population, regardless of age, disease, or treatment. They allow comparison of the condition of interest with other diseases; however, for some conditions their content may be redundant, and they may be inadequate at detecting change.

Medical Outcomes Study Short Forms

The Medical Outcomes Study 36-item Short Form (SF-36) comprises 36 items scored as eight domain profiles, including physical functioning, role limitations–physical (bodily pain, general health, vitality, social functioning), role limitations–emotional, and mental health, as well as two summary measures: physical and mental. Shorter versions, the SF-12 and the SF-8, use selected items from the SF-36. The SF-6D, developed recently as a preference-based, health utility measure, has 11 items.

EuroQol-5D

The EuroQol-5D (EQ-5D) comprises five items, including mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each of which is rated on a three- or five-point scale. It is both a generic outcome measure and a utility measure. The EQ-5D yields unique health states and a summary total score. The accompanying EQ–visual analog scale measures health-related QOL on a scale of zero to100.

Recommendation

Well-developed and validated generic measures can be regarded as being “fit for purpose,” even in clinical settings in which content validity has not been documented previously.[18] Hence, any of the established generic measures may be considered suitable for use in patients with FAI. Concerns have been expressed regarding the bimodal distribution of the EQ-5D and its ability to measure change due to to its limited response options; however, good responsiveness has been shown in FAI patients undergoing surgery.[19] If economic evaluations are of interest and the results of clinical trials are to be compared with other conditions or hip registry data, the EQ-5D may be recommended. If more detailed information on the various health-related profiles is required, either the SF-36 or the SF-6D may be preferred.

Activity Level Measures

Activity level instruments may provide valuable information that is not represented in the disease-specific and generic outcome measures. Although the instruments described below have been used in studies of FAI, their test properties have not been assessed adequately in young to middle-aged patients with hip pain.

The University of California at Los Angeles (UCLA) activity score initially was described in a study comparing total joint arthroplasty with hip resurfacing.[20] The UCLA score provides descriptive activity levels ranging from 1 to 10, with 1 meaning “wholly inactive; dependent on others,” and 10 meaning “regularly participates in impact sports” (eg, jogging). It is the most frequently used activity scoring system in studies of FAI; however, it has been shown to have a low correlation to daily step-count in young and middle-aged adults with hip pain ($\rho = 0.30$).[21]

The Tegner Activity Scale was developed to assess patients with anterior cruciate ligament injury.[22] The Tegner scale provides descriptive activity levels ranging from zero to 10, with zero meaning “sick leave or disability” and 10 meaning participation in “competitive

sports." Unlike the UCLA score, the Tegner scale provides work components stratified by physical demand in addition to sport-specific activities.

The Marx Activity Level Scale was validated in persons with knee injuries.[23] It queries the frequency of participation in pivoting, cutting, and deceleration activities to assess patient participation in high-demand sports.

The Baecke Questionnaire is a multiscale instrument that measures habitual physical activity, with subscales for work, sports, and leisure.[24] A positive feature of this scale, compared with others, is that it assesses the frequency, duration, and intensity of activity.[25]

In addition to the patient-reported activity level, investigators have included objective testing such as the 6-minute walk test[26,27] and accelerometer recordings in their studies to assess activity level.[21,28] Recent studies have documented discrepancies between the patient-reported activity level and objective testing, indicating that patient-reported activity and objectively assessed physical activity are correlated but distinctly different constructs.[29]

The measurement of activity level in patients with FAI may provide important information for clinical trials; however, no specific measure can be recommended. Investigators should consider their patient population when determining which instruments are appropriate for their studies. In addition, investigators should consider using objective testing as a component of clinical outcomes assessment.

Future Directions: Computer-adaptive Testing Systems

The ideal instrument precisely measures across the entire continuum of the construct of interest. Classical methods require that everyone answer every question, increasing respondent burden with content coverage. CAT is built on item response theory and methods that allow the selection of questions from a large calibrated bank covering the continuum of the construct of interest.[30] With good precision, CAT scores are calculated at the item level, using up to 10 questions.[31,32] Item response theory/CAT methods allow the addition of new items into calibrated banks with replenishment calibrations studies.[33] CAT systems currently in use include the Patient Reported Outcomes Measurement Information System[34] and the osteoarthritis (OA) CAT systems ie OA-DISABILITY-CAT, OA-FUNCTION-CAT).[35,36] Both instruments address pain and function, but neither tool focused specifically on younger persons with FAI during initial development. Future work may address the performance of these measures in this population and whether they are appropriate to serve as the basis for the development of an instrument to measure across the continuum of hip dysfunction.

Summary

PROs are necessary to determine the effects of FAI treatment. The HAGOS and the iHOT-33 may be promising disease-specific PROs to use as a primary measure, yet both instruments were developed recently and their longitudinal measurement properties in FAI populations are yet to be determined. Therefore, further testing and direct comparisons are needed to determine whether one instrument is superior to the other. Generic measures such as the SF-36 and the EQ-5D should be considered as secondary measures. Activity level is best evaluated by a combination of self-reported and objective measures.

Acknowledgments

The authors would like to recognize the attendees of the symposium for their contributions.

This work was supported by the following grant: Dr. Harris-Hayes was supported by grant K23 HD067343 from the National Center for Medical Rehabilitation Research, National Institute of Child Health and Human Development.

Clinical Outcomes Assessment in Clinical Trials to Assess Treatment of Femoroacetabular Impingement: Use of Patient-reported Outcome Measures Marcie Harris-Hayes, PT, DPT, MSCI, OCS, et al

References

1. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. Qual Life Res. 2010; 19(4):539–549. [PubMed: 20169472]
2. 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. [PubMed: 12777182]
3. Byrd JW, Jones KS. Prospective analysis of hip arthroscopy with 2-year follow-up. Arthroscopy. 2000; 16(6):578–587. [PubMed: 10976117]
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: A health status instrument for measuring clinically important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol. 1988; 1:95–108.
5. Martin RL, Kelly BT, Philippon MJ. Evidence of validity for the hip outcome score. Arthroscopy. 2006; 22(12):1304–1311. [PubMed: 17157729]
6. Thorborg K, Hölmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): Development and validation according to the COSMIN checklist. Br J Sports Med. 2011; 45(6):478–491. [PubMed: 21478502]
7. Mohtadi NG, Griffin DR, Pedersen ME, et al. Multicenter Arthroscopy of the Hip Outcomes Research Network: The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: The International Hip Outcome Tool (iHOT-33). Arthroscopy. 2012; 28(5):595–605. quiz 606, e1. [PubMed: 22542433]
8. Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: Reliable and validated. Clin Orthop Relat Res. 2003; 406:75–83. [PubMed: 12579003]
9. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br. 1996; 78(2):185–190. [PubMed: 8666621]
10. Lodhia P, Slobogean GP, Noonan VK, Gilbart MK. Patient-reported outcome instruments for femoroacetabular impingement and hip labral pathology: A systematic review of the clinimetric evidence. Arthroscopy. 2011; 27(2):279–286. [PubMed: 21035994]
11. Tijssen M, van Cingel R, van Melick N, de Visser E. Patient-Reported Outcome questionnaires for hip arthroscopy: A systematic review of the psychometric evidence. BMC Musculoskelet Disord. 2011; 12:117. [PubMed: 21619610]
12. Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: A systematic review. Br J Sports Med. 2010; 44(16):1186–1196. [PubMed: 19666629]
13. Ahmad MA, Xypnitos FN, Giannoudis PV. Measuring hip outcomes: Common scales and checklists. Injury. 2011; 42(3):259–264. [PubMed: 21163481]
14. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Qual Life Res. 2012; 21(4):651–657. [PubMed: 21732199]
15. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Med Res Methodol. 2010; 10:82. [PubMed: 20860789]
16. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. BMC Med Res Methodol. 2010; 10:22. [PubMed: 20298572]

17. Griffin DR, Parsons N, Mohtadi NG, Safran MR. Multicenter Arthroscopy of the Hip Outcomes Research Network: A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. Arthroscopy. 2012; 28(5):611–616. quiz 616–618. [PubMed: 22542434]
18. Magasi S, Ryan G, Revicki D, et al. Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. Qual Life Res. 2012; 21(5):739–746. [PubMed: 21866374]
19. Impellizzeri FM, Mannion AF, Naal FD, Hersche O, Leunig M. The early outcome of surgical treatment for femoroacetabular impingement: Success depends on how you measure it. Osteoarthritis Cartilage. 2012; 20(7):638–645. [PubMed: 22469846]
20. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip: A comparison of total joint and surface replacement arthroplasty. J Bone Joint Surg Am. 1984; 66(2):228–241. [PubMed: 6693450]
21. Harris-Hayes M, Steger-May K, Pashos G, Clohisy JC, Prather H. Stride activity level in young and middle-aged adults with hip disorders. Physiother Theory Pract. 2012; 28(5):333–343. [PubMed: 22191475]
22. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. Clin Orthop Relat Res. 1985; 198:43–49. [PubMed: 4028566]
23. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. Am J Sports Med. 2001; 29(2):213–218. [PubMed: 11292048]
24. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr. 1982; 36(5):936–942. [PubMed: 7137077]
25. Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. Osteoarthritis Cartilage. 2011; 19(6):620–633. [PubMed: 21251989]
26. Keener JD, Callaghan JJ, Goetz DD, Pederson D, Sullivan P, Johnston RC. Long-term function after Charnley total hip arthroplasty. Clin Orthop Relat Res. 2003; 417:148–156. [PubMed: 14646712]
27. Enright PL, McBurnie MA, Bittner V, et al. Cardiovascular Health Study: The 6-min walk test: A quick measure of functional status in elderly adults. Chest. 2003; 123(2):387–398. [PubMed: 12576356]
28. Silva M, Shepherd EF, Jackson WO, Dorey FJ, Schmalzried TP. Average patient walking activity approaches 2 million cycles per year: Pedometers under-record walking activity. J Arthroplasty. 2002; 17(6):693–697. [PubMed: 12216021]
29. Busse ME, van Deursen RW, Wiles CM. Real-life step and activity measurement: Reliability and validity. J Med Eng Technol. 2009; 33(1):33–41. [PubMed: 19116852]
30. Lord, FM. Applications of Item Response Theory to Practical Testing Problems. Erlbaum Associates; Hillsdale, NJ: 1990.
31. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther. 2011; 13(5):R147. [PubMed: 21914216]
32. Jette AM, Haley SM, Tao W, et al. Prospective evaluation of the AM-PAC-CAT in outpatient rehabilitation settings. Phys Ther. 2007; 87(4):385–398. [PubMed: 17311888]
33. Haley SM, Ni P, Jette AM, et al. Replenishing a computerized adaptive test of patient-reported daily activity functioning. Qual Life Res. 2009; 18(4):461–471. [PubMed: 19288222]
34. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol. 2008; 61(1):17–33. [PubMed: 18083459]
35. Jette AM, McDonough CM, Haley SM, et al. A computer-adaptive disability instrument for lower extremity osteoarthritis research demonstrated promising breadth, precision, and reliability. J Clin Epidemiol. 2009; 62(8):807–815. [PubMed: 19216052]
36. Jette AM, McDonough CM, Ni P, et al. A functional difficulty and functional pain instrument for hip and knee osteoarthritis. Arthritis Res Ther. 2009; 11(4):R107. [PubMed: 19589168]

Table 1

Disease-specific Patient-Reported Outcome Measures Used to Assess Patients With Femoroacetabular Impingement

PRO (Year Developed)	Target Population[a]	Content	No. of Items and Recall Period	Response Option/Scale	Score Interpretation	Strengths	Cautions
HAGOS[16] (2011)	Young to middle-aged, physically active patients with longstanding hip and/or groin pain	Pain, symptoms, ADL, sports/recre ation function, PA, QOL	37 items. Past week considered for all items.	All items rated on a 5-point Likert scale (0–4). Subscale scores are summed, then transformed to 0–100 worst to best scale.	0 = extreme symptoms, 100 = no symptoms	Patients representative of those with FAI involved in the development. Freely available. Subscale scores are calculated. COSMIN ratings scored fair to excellent.[b] PA subscale includes relevant questions about sports and other PA. Few patients report floor or ceiling effect.	Has not yet been used extensively in the literature
HOS[5] (2006)	Patients with acetabular labral tears who may be functioning throughout a wide range of ability	ADL, sport	26 scored items. Additional nonscored items to rate current level of function. Past week considered for all items.	All items rated on a 5-point Likert scale (0–4). Subscale scores summed, then transformed to 0–100 worst to best scale.	Scores range from 0–100 Higher score represents a higher level of physical function.	Can detect change at higher functional levels (sports). Calculation of subscale scores. Validated using both Classical and Item Test theory.	Patients not involved in the development[c]
HOOS[2] (2003)	Patients with and without hip OA	Pain, symptoms (stiffness, ROM), ADL, sport and recreation function, hip-related QOL	40 items. Past week considered for all items.	All items rated on a 5-point Likert scale (0–4). Subscale scores are summed, then transformed to 0–100 worst to best scale.	0 = extreme symptoms, 100 = no symptoms.	Able to detect change at higher functional levels in people with OA. WOMAC score may be calculated from data. Freely available. Available in many languages.	Patients aged <42 years representative of those with FAI were not involved in the development. Psychometric properties in young patients not known.[c]
iHOT-33[7] (2012)	Young, active patients with hip disorders	Symptoms and functional limitations; sports and recreational physical activities; job-	33 items. Past month considered for all items.	All items are rated using 100-point VAS. Total score is calculated.	100 is best possible score	Patients representative of those with FAI involved in the development. Freely available. COSMIN	Has not been used extensively in the literature. Subscale scores have not been validated.

PRO (Year Developed)	Target Population[a]	Content	No. of Items and Recall Period	Response Option/Scale	Score Interpretation	Strengths	Cautions
		related concerns; social, emotional, and lifestyle concerns				ratings scored fair to excellent.[b]	
MHHS[3] (2000)	Patients undergoing hip arthroscopy	Pain, function	8 items. Recall period not specified.	Arbitrary weights assigned to each item. Total score is calculated.	Scores of 0–100, with higher scores indicating better function.	Based on the Harris Hip score, one of the oldest and most commonly used surgeon-derived scores in hip and THA research.	Patients representative of those with FAI not involved in the development. No sports-specific items.[c]
NAHS[8] (2003)	Young, active patients with activity-limiting hip pain	Pain, mechanical symptoms, physical function, activity level	20 items. Past 48 hours considered for all items.	All items rated on a 5-point Likert scale (0–4). Total score and subscale scores may be calculated.	100 = normal hip function	Patients representative of those with FAI involved in the development. Subscale scores may be calculated. Includes questions related to mechanical symptoms.	Poor to fair COSMIN ratings[b]
OHS[9] (1996)	Patients having total hip arthroplasty	Items related to symptoms and function. No subscales.	12 items. Past 4 wk considered.	All items rated on a 5-point Likert scale (1–5). Total score is calculated.	12 = least difficulties, 60 = most difficulties, (original paper). Modification s have been reported.	Intended to provide a short questionnaire specific to THA population.	Not intended for use across the range of hip disorders. No sports-specific items. Psychometric properties in young patients not known.[c]
WOMAC[4] (1998)	Patients with hip or knee OA	Pain, stiffness, physical activity	24 items. Past 48 hours considered.	All items rated on a 5-point Likert scale (0–4). Total score and subscale scores are calculated.	Sum score. Lower scores indicate less symptomology.	Among the most widely used outcome measures in OA. Available in many languages.	No sports-specific items. May have ceiling effect in younger patients. Various versions published. License and fees apply. Psychometric properties were not reviewed.[c]

ADL = activities of daily living, COSMIN = COnsensus-based Standards for the selection of health Measurement INstrument, FAI = femoroacetabular impingement, HAGOS = Copenhagen Hip and Groin Outcome Score, HOOS = hip disability and osteoarthritis outcome score, HOS = Hip Outcome Score, iHOT-33 = 33-item International Hip Outcome Tool, MHHS = Modified Harris Hip Score, NAHS = Nonarthritic Hip Score, OA = osteoarthritis, OHS = Oxford Hip Score, PA = participation in physical activity, PRO = patient-reported outcomes measure, QOL = quality of life, ROM = range of motion, THA = total hip arthroplasty, VAS = visual analog scale, WOMAC = Western Ontario and McMaster Universities Osteoarthritis Index

[a] Described by the authors who developed the questionnaires.

[b] See Table 2 for COSMIN ratings

[c] Psychometric properties of the HOS, HOOS, MHHS, OHS and WOMAC were not reviewed in this summary. The authors reviewed the psychometric properties of only those disease-specific PROs in which content validity was ensured through input from patients of similar age, sex, and activity level and who had experienced symptoms and limitations due to FAI. Investigators are encouraged to search for literature describing psychometric properties of potential PROs prior to use in clinical trials. The COSMIN guidelines may be used to determine adequate quality of the reports describing PROs..

J Am Acad Orthop Surg. Author manuscript; available in PMC 2014 March 31.

Table 2

COSMIN Ratings for Disease-specific Patient-reported Outcome Measures[a,b]

Measure	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypothesis Testing	Cross-cultural Validity	Criterion Validity	Responsiveness
HAGOS[6]	E	F	F	E	E	E	NT	NT	G
iHOT-33[7]	P	F	NT	E	P	F	NT	NT	F
NAHS[8]	P	P	NT	P	NT	F	NT	NT	NT

COSMIN = COnsensus-based Standards for the selection of health Measurement INstrument, HAGOS = Copenhagen Hip and Groin Outcome Score, iHOT-33 = 3 3-item International Hip Outcome Tool, NAHS = Nonarthritic Hip Score

[a] Each article was assessed by two independent reviewers (M.H.H., C.M.M.). Disagreements were discussed and consensus determined. Where consensus was not reached, a third reviewer (E.M.R.) was consulted.

[b] Scoring: E = excellent, G = good, F = fair, P = poor, NT = measurement property was not assessed.