http://dx.doi.org/10.2106/JBJS.RVW.17.00038

Appendix

Assessments of Patient-Reported Outcome Measures (PROMs) for Total Hip Arthroplasty Patients

Arthritis Impact Measurement Scales (AIMS)

The AIMS are a disease-specific PROM designed for patients with rheumatic diseases. Negative evidence was found for hypothesis testing and responsiveness when the scales were correlated with related constructs such as mobility, pain, and social activities of the Functional Status Index (FSI), Index of Well-Being (IWB), and Sickness Impact Profile (SIP)[38]. No additional empirical evidence was identified.

Short Arthritis Impact Measurement Scales (sAIMS)

This shortened version of the AIMS was developed to minimize the completion time and maximize the response rate. All constructs from the original version remain in this shortened version[39].

This instrument was evaluated for hypothesis testing and responsiveness but no positive evidence was found. For hypothesis testing, this was due to the lack of data reported for correlations between generic instruments with unrelated constructs, and for responsiveness, this was due to the lack of correlation between instruments measuring the same constructs[24]. No additional empirical evidence was identified.

Core Outcome Measures Index-Hip Oriented (COMI-hip)

To eliminate the need for patients to complete lengthy questionnaires, the COMI-hip was designed to cover the core domains in patients who undergo THA[23] (see Table III).

Moderate positive evidence was found for hypothesis testing. Correlations between the COMI-hip and World Health Organization Quality of Life questionnaire short version (WHOQOL-Bref), Short Form-12 (SF-12), and EuroQol (EQ)-5D were found to be higher in related constructs than in unrelated constructs such as social disability and quality of life. Indeterminate evidence was found for responsiveness because of the lack of correlation for change scores or differences between groups with closely related constructs[23]. No additional empirical evidence was identified.

Disability Rating Index (DRI)

The Disability Rating Index comprises 12 items measuring the domains of basic daily activities, daily physical activities, and work-related activities[40].

Limited positive evidence was found for structural validity. Two factors explained 66% of the score variance. In contrast, indeterminate ratings were given to internal consistency, reliability, measurement error, hypothesis testing, and responsiveness because of the poor methodological quality of the included study[40]. No additional empirical evidence was identified.

EUROHIS-Quality of Life 8-Item Index

This 8-item instrument was derived from a 100-item World Health Organization Quality of Life (WHOQOL) questionnaire that measures general patient conditions. It contains 2 items in each of the four domains (physical, social, psychological, and environmental)[41].

The only evidence was found in responsiveness. An indeterminate rating was given to it because of the lack of correlation with other instruments[41]. No other evidence was found from the included study.

Forgotten Joint Score-12 (FJS-12)

This instrument is a recently published PROM to assess awareness of hip and knee joints in daily activities[42].

Limited positive evidence was found for internal consistency. The Cronbach alpha was 0.98 for the scale, and factor analysis was done to confirm its unidimensionality. Moderate positive evidence was found for hypothesis testing. Correlations between FJS-12 and WOMAC pain and function subscales were −0.75 and −0.78, respectively, higher than for the unrelated constructs (i.e., stiffness)[43]. No additional empirical evidence was identified.

Harris Hip Score (HHS)

The HHS was originally a clinician-reported outcome instrument developed by a single surgeon; it is now used as a PROM. It includes four domains: function, pain, deformity, and range of motion. Together, these are summed as a measure of hip dysfunction.

Limited positive evidence was found for criterion validity, as the correlation of patient-reported and physician-reported scores was >0.70[44]. Limited positive evidence was also found for hypothesis testing, as correlations were found to be >0.50 for the constructs of physical function and bodily pain when compared

with the Medical Outcomes Survey Short Form-36[101]. However, indeterminate ratings were given to internal consistency, reliability, measurement error, responsiveness, and interpretability because of the poor methodological quality of the included papers[45,93]. No additional empirical evidence was identified.

Harris Hip Score (HHS)-Modified

This self-administered instrument contains two domains (pain and function) of the Harris hip score[46].

Only indeterminate evidence was found for this instrument because of the lack of correlations with other instruments[46]. No additional empirical evidence was identified.

Hip Disability and Osteoarthritis Outcome Score (HOOS)

The HOOS is an adaptation of the Knee injury and Osteoarthritis Outcome Score (KOOS), which was itself an extension of the WOMAC. It was developed for use in patients with osteoarthritis who were younger and had higher demands involving physical function. Specifically, the constructs of sport and recreational function were added to the instrument.

A strong positive rating was given to content validity, as evidenced by its items being rated relevant, important, and complete[47]. A moderate positive rating was given to reliability because of the high intraclass correlation coefficient (ICC) values (>0.83) found for all constructs[26]. Moderate positive evidence was found for hypothesis testing, since the correlation of the instrument was higher in related constructs and results confirmed all of the hypotheses established a priori[48]. Limited positive evidence was found for interpretability. Paulsen et al.[49] reported means and standard deviations of the HOOS for 5 subgroups of patients; in addition, an MID was defined. Internal consistency, measurement error, and responsiveness received indeterminate ratings because of the poor methodological quality of the studies[26,47,94]. No additional empirical evidence was identified.

Hip Disability and Osteoarthritis Outcome Score-Joint Replacement Short Form (HOOS, JR)

This 6-item instrument was derived from the WOMAC, particularly for patients undergoing THA. It measures 2 domains of pain and activities of daily living[50].

Limited positive evidence was found for hypothesis testing and criterion validity. Correlations were >0.5 between similar constructs and were higher than with unrelated constructs[50], and coefficients were high (>0.7) when this instrument was correlated

with the WOMAC. Indeterminate ratings were given to internal consistency, responsiveness, and interpretability because of the poor methodological quality of the included study[50].

Hip Disability and Osteoarthritis Outcome Score-Physical Function Short Form (HOOS-PS)

Developed from the HOOS, this physical function short form is intended to measure the functional outcome of patients with osteoarthritis who receive either conservative management or THA.

Moderate positive evidence was found for hypothesis testing, as the correlation between the HOOS-PS and WOMAC physical function subscale was >0.50, which was also higher than correlations with unrelated construct[51,104]. Also, the results were in accordance with the a priori hypotheses. Limited positive evidence was found for reliability and interpretability, with an ICC of 0.80[52], a defined MID, and data for 5 relevant patient groups[49]. Internal consistency and responsiveness were rated indeterminate because of the poor methodological quality of the included studies[51]. No additional empirical evidence was identified.

Hip Outcome Score (HOS)

The HOS has 2 subscales, activities of daily living and sport, that are meant for the evaluation of patients with acetabular tears[53].

Only negative results were found for this instrument. For internal consistency, although the Cronbach alpha was found to be 0.93 for the activities of daily living subscale and 0.88 for the sport subscale, unidimensionality was confirmed by factor analysis for the sport subscale only[97]. For structural validity, 1 study showed that 4 factors explained 56% of the variance of the activities of daily living subscale but only 2 factors explained 49% of the variance of the sport subscale[97], resulting in a negative evidence rating. Indeterminate ratings were given for reliability, hypothesis testing, responsiveness, and interpretability because of the poor methodological quality of the studies assessing them[97]. No additional empirical evidence was identified.

Hip-Specific Quality of Life Survey (HSS)

The HSS was developed to evaluate quality of life related to the hip joint[54]. It includes 4 subdomains: right hip joint evaluation, left hip joint evaluation, assessment of basic function, and assessment of advanced function.

The only evidence from the literature resulted in an indeterminate rating for internal consistency and responsiveness because of the unknown dimensionality of the instrument[98] and poor methodological quality of the included study[46]. No additional empirical evidence was identified.

Intermittent and Constant Osteoarthritis Pain (ICOAP) Questionnaire

The ICOAP questionnaire was developed for use in patients with hip or knee osteoarthritis to measure 2 domains: constant pain and intermittent pain.

Limited positive evidence was found for reliability. ICCs were >0.82 for both the constant and intermittent pain subdomains[52]. For hypothesis testing, limited negative evidence was found when correlating the scores of the ICOAP with the WOMAC pain and function subscales. Correlations were higher between the ICOAP and the WOMAC dysfunction domain ($r = 0.76$) than those with the WOMAC pain domain ($r = 0.75$)[104]. Results were unknown for responsiveness, as no complete information about the correlations measuring related or unrelated constructs was found in the studies that intended to test this property[55,104]. No additional empirical evidence was identified.

Lequesne Algofunctional Index

The Lequesne index is a disease-specific instrument designed for patients with osteoarthritis. It includes 3 subdomains: pain, maximum walking distance, and physical function disability.

The ratings for internal consistency, reliability, hypothesis testing and responsiveness were indeterminate because of the poor methodological quality[95]. No additional empirical evidence was identified.

McMaster Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR)

The MACTAR evaluates the change in condition of patients with rheumatoid arthritis by rating a set of self-nominated physical or social activities that are adversely affected by the patient's hip disease.

An indeterminate evidence rating was given for responsiveness because of the poor methodological quality of the included study[93]. No additional empirical evidence was identified.

Musculoskeletal Outcomes Data Evaluation and Management System (MODEMS) Hip/Knee Core Scale

The MODEMS hip/knee core scale focuses mainly on measuring early recovery of hip and knee function following total knee arthroplasty or THA through 2 domains: pain and function.

Results for both hypothesis testing and responsiveness were indeterminate in the only study that assessed this instrument because of the incomplete reporting and poor methodological quality[56]. No additional empirical evidence was identified.

Osteoarthritis-Computer Adaptive Testing (OA-CAT) Functional Difficulty, Functional Pain, and Disability Scales

This instrument uses a computer algorithm to tailor a selection of items to measure the level of ability reported by patients in the domains of functional difficulty, functional pain, and disability.

Only indeterminate evidence was found for responsiveness because of the lack of correlation with other instruments[46]. No additional empirical evidence was identified.

Osteoarthritis Knee and Hip Quality-of-Life (OAKHQOL) Questionnaire

The OAKHQOL is a disease-specific instrument that measures quality of life in patients with osteoarthritis of the hip or knee. It includes 5 domains: physical activities, mental health, pain, social support, and social functioning.

Studies evaluating the OAKHQOL confirmed unidimensionality of the scale by factor analyses and internal consistency with Cronbach alpha values ranging from 0.78 to 0.93, resulting in strong positive evidence for this property[57]. Assessments on patient relevance and comprehensiveness were performed for content validity through interviews of experts such as health psychologists and sociologists[58]. As a result, strong positive evidence was found for content validity of the OAKHQOL. Also, a strong positive rating was given for structural validity, as evidenced by factor analysis showing that the 4 factors explained 64% of the total variance of the OAKHQOL using principal components analysis (PCA)[58]. In contrast, strong negative evidence was found for reliability in that 3 of the 5 constructs (physical activity, pain, and mental health) had ICCs ranging from 0.85 to 0.90, but the constructs of social support and social activities in patients waiting for THA had ICCs of <0.70[57] This result was consistent with an additional study[30]. Conflicting results were found for hypothesis testing. Finally, an unknown level of evidence was assigned to responsiveness because of the

poor methodological quality of the included studies[30,57,58]. No additional empirical evidence was identified.

Oxford Hip Score (OHS)

OHS is a region-specific outcome instrument that monitors pain, function, and mobility for patients who undergo THA.

A strong positive rating was given to hypothesis testing for the OHS. Eighty percent of the findings (12 of 15 hypotheses) were in accordance with the hypotheses established a priori, and related constructs had higher correlations[59]. Strong positive evidence was also found for reliability, as evidenced by ICCs of >0.94 for all constructs[60]. Limited positive evidence was found for content validity[61]. Limited negative evidence was found for responsiveness, as correlations with related constructs were <0.5[11]. Finally, an unknown level of evidence was given to internal consistency, measurement error, criterion validity, and interpretability because of the poor methodological quality of the included studies assessing these constructs[27,29,62]. No additional empirical evidence was identified.

Patient-Administered Questionnaire (PAQ)-Hip

The PAQ-Hip is a newly developed instrument that evaluates physical symptoms, function, psychological symptoms, and satisfaction.

An unknown level of evidence was assigned to internal consistency, reliability, hypothesis testing, and responsiveness because of the poor methodological quality of the included studies[92]. No additional empirical evidence was identified.

Patient Specific Index (PASI)

PASI allows patients to rate their concerns on fixed or open questions related to their symptoms and function.

Moderate positive evidence was found for hypothesis testing, as the PASI scores were moderately correlated with the WOMAC pain and function subscales as well as related constructs of the SF-36, and >75% of the results confirmed the hypotheses in the study[93]. Limited positive evidence was found for criterion validity, as the self-reported PASI was correlated with an interviewer-administered PASI[63]. Limited positive evidence was also found for responsiveness, as evidenced by correlations between global ratings and the PASI being >0.50, which were also higher than correlations with unrelated constructs[93]. Reliability was given an unknown level of evidence because of the poor methodological quality of these assessments[63,93]. All other

properties are unknown in quality because of the poor methodology or lack of additional empirical evidence.

Patient Specific Functional Scale (PSFS)

The PSFS was first introduced in 1995 by Stratford et al.[64]. The PSFS asks patients with low back pain to name 5 important activities that they are unable to perform or have difficulty performing because of their condition[64].

Hypothesis testing and responsiveness were the only 2 properties that were assessed for the PSFS. Limited negative evidence was found for hypothesis testing due to correlations being <0.50 between the PSFS and WOMAC measuring similar constructs[10]. Indeterminate evidence was found for responsiveness, as there was no correlation between changes in scores[10]. No additional empirical evidence was identified.

Total Hip Arthroplasty Outcome Evaluation Questionnaire (THA OEQ)

Developed by the Task Force on Outcome Studies of the American Academy of Orthopaedic Surgeons, the THA OEQ comprises 3 separate forms: baseline, history, and postoperative outcomes. These forms were derived from the patients' perspectives and measure the following domains: pain, level of activity, capacity for walking (including activities of daily living), and patient satisfaction[65].

Limited negative rating was given to hypothesis testing, as <75% of the results confirmed the hypotheses and correlations were found to be <0.50 for related constructs[65]. An unknown Level of Evidence was given to reliability because of the poor methodological quality of the included studies. Overall, no evidence was found for any other property.

Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)

The WOMAC is a widely used disease-specific PROM to assess patients with hip osteoarthritis or knee problems in the domains of dysfunction, pain, and stiffness.

Moderate positive evidence was found for responsiveness, as all results confirmed the hypothesis established for predicting the correlations between WOMAC change scores and patients' global ratings of change in hip function[93]. Moderate positive evidence was also found for criterion validity, as the correlations between the electronic and paper versions were as high as 0.77[96]. Limited positive evidence was found for internal consistency and reliability; Cronbach alpha values for all constructs were >0.80,

with a COSMIN score of "good" indicating adequate methods[66]. This finding was also supported in the pain subscale by an additional study[99]. Also, ICCs were found to be >0.70 for all constructs of the WOMAC[105], again supporting the reliability of this measure. Limited positive evidence was found for hypothesis testing, as evidenced by correlations with related constructs being >0.50 in several of the studies[19]. Indeterminate evidence was found for interpretability. Quintana et al.[67] defined the MID of the WOMAC as 25 points, but the number of patients was lower than the required sample size. No additional empirical evidence was identified.

Western Ontario and McMaster Universities Osteoarthritis-Total Hip Replacement (WOMAC-THR) Function Short Form

The WOMAC-THR was developed from the original WOMAC to assess outcomes in patients undergoing THA. The instrument's 8 items were selected by THA patients as being the most important items related to their function.

We found only 1 study assessing the properties of this instrument. This study had poor methodological quality, resulting in unknown levels of evidence for internal consistency, hypothesis testing, and responsiveness[103]. No additional empirical evidence was identified.