Clinical Sports Medicine Update

High Degree of Variability in Reporting of Clinical and Patient-Reported Outcomes After Hip Arthroscopy

Austin V. Stone,*† MD, PhD, Cale A. Jacobs,‡ PhD, T. David Luo,* MD, Molly C. Meadows,§ MD, Shane J. Nho,§ MD, MS, Allston J. Stubbs,* MD, MBA, and Eric C. Makhni,‖¶ MD, MBA
Investigation performed at Department of Orthopaedic Surgery, Wake Forest Baptist Medical Center, Winston-Salem, North Carolina, USA, and at the Henry Ford Health System, Detroit, Michigan, USA

Background: Hip arthroscopy for the treatment of intra-articular pathology is a rapidly expanding field. Outcome measures should be reported to document the efficacy of arthroscopic procedures; however, the most effective outcome measures are not established.

Purpose: To evaluate the variability in outcomes reported after hip arthroscopy and to compare the responsiveness of patient-reported outcome (PRO) instruments.

Study Design: Systematic review.

Methods: We reviewed primary hip arthroscopy literature between January 2011 and September 2016 using the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines. Patient and study characteristics were recorded. Pre- and postoperative means and SDs of PROs were recorded from articles that used 2 or more PROs with a 1-year minimum follow-up. From this subset of articles, we compared the responsiveness between PRO instruments using the effect size, standard response mean, and relative efficiency.

Results: We identified 130 studies that met our inclusion/exclusion criteria, which totaled 16,970 patients (17,511 hips, mean age = 37.0 years, mean body mass index = 25.9 kg/m^2). Radiographic measures were reported in 100 studies. The alpha angle and center-edge angle were the most common measures. Range of motion was reported in 81 of 130 articles. PROs were reported in 129 of 130 articles, and 21 different PRO instruments were identified. The mean number of PROs per article was 3.2, and 78% used 2 or more PROs. The most commonly used PRO was the modified Harris Hip Score, followed by the Hip Outcome Score (HOS)–Activities of Daily Living, HOS-Sport, visual analog scale, and Nonarthritic Hip Score (NAHS). The 2 most responsive PRO tools were the International Hip Outcome Tool (iHOT)–12 and the NAHS.

Conclusion: Outcomes reporting is highly variable in the hip arthroscopy literature. More than 20 different PRO instruments have been used, which makes comparison across studies difficult. A uniform set of outcome measures would allow for clearer interpretation of the hip arthroscopy literature and offer potential conclusions from pooled data. On the basis of our comparative responsiveness results and previously reported psychometric properties of the different PRO instruments, we recommend more widespread adoption of the iHOT PROs instruments to assess hip arthroscopy outcomes.

Keywords: hip; arthroscopy; patient-reported outcomes; outcomes; PRO

The number of hip arthroscopy procedures experienced an 18-fold increase between 1999 and 2009.[5] Increases in hip arthroscopy–related research have paralleled the increased surgical volume. As the number of both cases and research studies continues to increase, however, there has been no consensus as to objective clinical data and patient-reported outcomes (PROs) reporting after hip arthroscopy procedures. Effective outcomes reporting is critical to justify early intervention for hip pathology since patients are frequently young and active. The expanding indications will also be associated with an increased cost to the health care system and efficacy, and surgical results should be reported to help justify its use. The lack of consensus guidelines on postoperative outcomes reporting has been likely driven, in part, by 2 factors: 1) The variability of reporting methodology has not been quantified in a systematic manner, and 2) there is a lack of “head-to-head” comparative responsiveness data between available PRO instruments that have been used in the literature.

The responsiveness of 2 PRO instruments can be directly compared when administered in the same patient population

The American Journal of Sports Medicine, Vol. XX, No. X
DOI: 10.1177/0363546517724743

both before and after surgery.[17] The responsiveness of a PRO instrument is defined as that instrument's ability to accurately detect change over time and is thought to be one of the most important measurement properties of a PRO instrument.[2,6,17] The purpose of this systematic review was to address these 2 potential barriers by evaluating the variability in outcomes reported after hip arthroscopy and comparing the responsiveness of commonly used PRO instruments. We hypothesized that there would be a wide range of variability in both the type of outcomes tools used and studies reporting the outcomes.

METHODS

A comprehensive literature review was performed in searchable database listings (PubMed, MEDLINE, Cochrane) using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines to identify all articles pertaining to hip arthroscopy published during the past 5 years (January 1, 2011, to September 1, 2016). Articles were identified using the keywords "hip arthroscopy," "hip arthroscopic," and "outcome(s)." Cross-referencing was performed to identify any potentially missed articles.

Inclusion criteria were English-language publication, any study with subjective or objective clinical outcomes after hip arthroscopy for any pathology, and publication after January 1, 2011. Exclusion criteria included basic science or cadaveric studies; studies reporting on exclusively open or revision surgery; review articles, meta-analyses, case reports, and editorials; studies focusing primarily on imaging without clinical correlations; and studies focusing on diagnostic or screening techniques.

Several metrics were collected for all studies meeting the inclusion criteria. Basic study descriptors included year and journal of publication, level of evidence, number of patients, number of hips studied, mean patient demographics, mean follow-up, and pathology treated. Outcomes recorded included range of motion and strength, revision rate, and conversion to total hip arthroplasty (THA). Patient satisfaction and PROs were also recorded. In addition, we recorded pre- and postoperative means and SDs of PROs from all articles that used 2 or more PROs at a minimum of 1-year follow-up. From this subset of articles, we were then able to directly compare the responsiveness between PRO instruments by calculating the effect size, standard response mean, and relative efficiency.

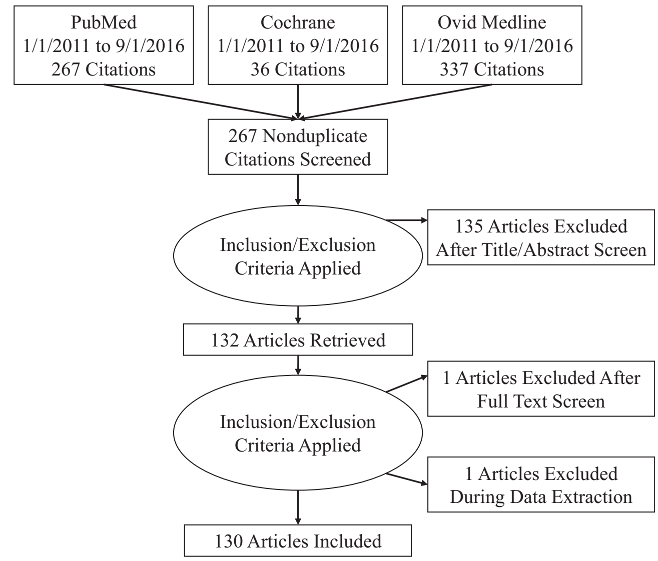

Figure 1. Flow diagram for included studies following searchable database listings per the PRISMA guidelines. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-analyses.

The responsiveness was compared between PRO instruments by calculating effect sizes, standardized response mean, and the relative efficiency of each outcome tool.[12,13,17] Effect size is a measure of the magnitude of the pre- to postoperative change accounting for the variability of a given PRO instrument, and effect sizes (d) are defined as small if they fall between 0.2 and 0.49, moderate if between 0.5 and 0.79, and large if $\geq$0.8.[4,7] We then calculated the relative efficiency to directly compare the responsiveness between 2 different PROs instruments that were used within the same patient population. When comparing 2 PRO tools, a relative efficiency value $<$1 indicates that the first PRO tool is less responsive than the other tool, whereas values $>$1 suggest that the first PRO tool is more responsive.[13,17] In the current analysis, relative efficiency values $<$.80 were defined as being indicative of poorer responsiveness, values between 0.80 and 1.20 were indicative of equal responsiveness between tools, and values $>$1.20 were defined as being indicative of greater responsiveness. On the basis of previous studies indicating large ceiling effects and questionable utility within hip preservation patient populations (Jacobs CA, Duncan ST, Muchow RD, Nunley

¶Address correspondence to Eric C. Makhni, MD, MBA, Department of Orthopedic Surgery, Division of Sports Medicine, 6777 W Maple Road, 3rd Floor East, Department of Orthopedics, West Bloomfield, MI, 48322 (email: ericmakhnimd@gmail.com).
*Department of Orthopaedic Surgery, Wake Forest Baptist Medical Center, Winston-Salem, North Carolina, USA.
†Division of Sports Medicine, Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, Illinois, USA.
‡Department of Orthopaedic Surgery & Sports Medicine, University of Kentucky, Lexington, Kentucky, USA.
§Section of Young Adult Hip Surgery, Department of Orthopedic Surgery, Division of Sports Medicine, Rush University Medical Center, Chicago, Illinois, USA.
‖Department of Orthopaedic Surgery, Henry Ford Hospital, Detroit, Michigan, USA.
A.V.S. and C.A.J. contributed equally to this article.
One or more of the authors has declared the following potential conflict of interest or source of funding: S.J.N. reports research support from Allosource, Arthrex, Athletico, DJ Orthopaedics, Linvatec, Miomed, Smith & Nephew, and Stryker; receives royalties from Ossur and Springer; and is a paid consultant for Ossur and Stryker. A.J.S. reports research support from Bauerfeind, AG, stock or stock options from Johnson & Johnson, and other financial or material support from Smith & Nephew. E.C.M. receives publishing royalties and financial or material support from Springer.

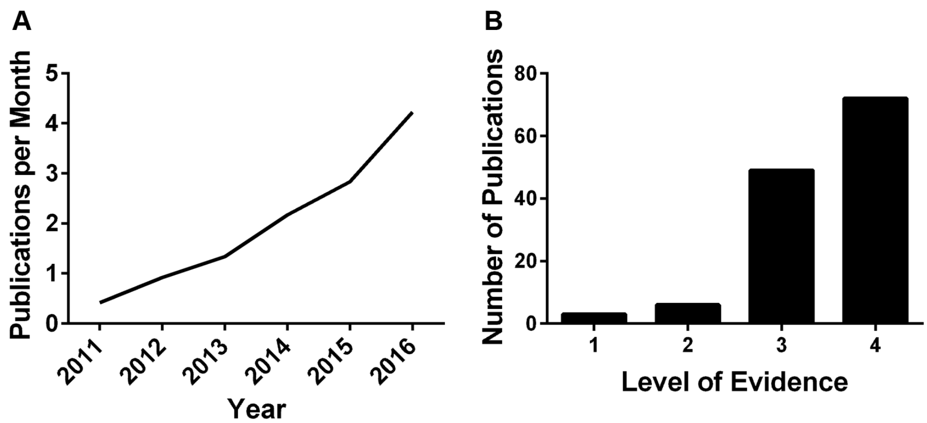

Figure 2. Rate and level of evidence of publications over the past 5 years. (A) The mean number of publications per month rapidly increased. (B) The predominate publication level of evidence was 3 and 4 since 2011.

TABLE 1
Characteristics of Included Studies[a]

Variable	Results (Range)	Studies Reporting (%)
No. of patients	130 (2-935)	130 (100)
No. of hips	135 (2-935)	130 (100)
Mean age, y	37.0 (7-87)	130 (100)
Mean BMI, kg/m^2	25.9	50 (38)
Mean follow-up, mo	26.8 (6-180)	130 (100)
Mean duration of symptoms, mo	31.9 (0.25-300)	26 (20)
Reported radiographic variables		100 (77)
Reported range of motion		82 (63)
Reported revision arthroscopy		70 (54)
Reported conversion to total hip arthroplasty		57 (44)

[a]Body mass index (BMI) was most commonly reported as a mean; consequently, a range could not be determined due to incomplete data from the index study.

RM, Clohisy JC, group TA. The HOOS, JR and other osteoarthritis-based patient-reported outcome tools demonstrated large ceiling effects after periacetabular osteotomy. Annual Meeting of the American Association of Hip and Knee Surgeons, Dallas, TX; 2016),[1,14,18] we chose not to include osteoarthritis-based PRO instruments in our responsiveness analyses, as large ceiling effects could artificially inflate the responsiveness of these tools in the hip preservation patient population. Specifically, we did not include the full or modified Harris Hip Score (HHS, mHHS), Hip Disability and Osteoarthritis and Outcome Score (HOOS), and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Also, to avoid misinterpretation due to small sample sizes, we did not include PRO tools in the responsiveness analyses if the tool appeared in only 1 article.

RESULTS

We identified 130 studies that met our inclusion/exclusion criteria, which included 16,970 patients (17,511 hips, mean age = 37.0 years, mean body mass index [BMI] = 25.9 kg/m^2) (Figure 1). The number of hip arthroscopy articles published gradually increased each year (Figure 2A). The mean number of hips per article was 135 (range, 2-935 hips), and the mean follow-up was 26.8 months (range, 6-180 months) (Table 1). The mean level of evidence was 3.5, with only 9 of 130 (6.9%) classified as level 1 or 2 evidence (Figure 2B). The journals publishing the greatest number of articles are detailed in Table 2.

Radiographic measures were reported in 100 of 130 studies (77%) (Table 1). The alpha angle (60/100) and center-edge angle (55/100) were the most common measures, while fewer studies reported the Tönnis grade (19/100) and Sharp angle (4/100). Additionally, range of motion was reported in 82 of 130 (63%) articles. The duration of preoperative symptoms was reported in 26 of 130 (20%) articles, with the overall mean duration of 31.9 months between the onset of symptoms and surgery (range, 1 week to 300 months). Reoperation was reported in 70 of 130 articles (54%), and conversion to THA was reported in 57 of 130 (44%).

PROs were reported in 129 of 130 articles, and 21 different PRO instruments were identified. The mean number of PROs per article was 3.2 (range, 0-6), and 101 of 129 (78%)

TABLE 2
Journals Most Frequently Publishing Outcome Studies in Hip Arthroscopy and the Number of Level 1 and 2 Studies

	No. of Studies	No. of Level 1 or 2 Studies
Arthroscopy	39	3
Am J Sports Med	33	2
J Hip Preserv Surg	7	2
Knee Surg Sports Traumatol Arthrosc	7	0
Hip Int	5	0
J Bone Joint Surg Am	5	1

used 2 or more PROs. The most commonly used PRO was the mHHS (n = 109 studies), followed by the Hip Outcome Score–Activities of Daily Living (HOS-ADL, n = 68), Hip Outcome Score–Sport (HOS-SSS, n = 64), visual analog scale (VAS, n = 47), and Nonarthritic Hip Score (NAHS, n = 42). Other commonly used PRO instruments included the Short Form–12 Physical Component Score (SF-12 PCS, n = 19), SF-12 Mental Component Score (SF-12 MCS, n = 16), WOMAC (n = 12), International Hip Outcome Tool–33 (iHOT-33, n = 11), HHS (n = 7), University of California at Los Angeles Activity Scale (UCLA, n = 6), and HOOS (n = 5). The iHOT-12, Copenhagen Hip and Groin Outcome Score (HAGOS), EuroQol-5D (EQ-5D), and Hip Sports Activity Scale (HSAS) were each used in 3 studies, and the VAS global hip, Oxford Hip Score, Lower Extremity Function Score, Likert, and Pain Numerical Rating Scale were each used in a single study. In addition, 41 of 130 articles (31%) reported patient satisfaction (Table 3).

From the full set of 130 articles, we identified a subset of 32 articles that met the inclusion criteria for responsiveness analyses (used 2 or more PROs, minimum follow-up of 1 year, reported both pre- and postoperative means and SDs, and appeared in more than 1 article). PRO use tended to fall into 2 groups: articles using the HOS and/or the NAHS with additional scores and articles using the iHOT-12 with additional scores (Table 4). From the set of articles that used the HOS and/or NAHS, the NAHS was the most responsive tool and was more responsive than was the HOS-ADL (relative efficiency [RE] = 1.43), HOS-SSS (RE = 1.42), and pain VAS (RE = 1.40). The pain VAS was more responsive than the UCLA (RE = 3.30), HOS-ADL (RE = 1.70), and HOS-SSS (RE = 1.72). Both the HOS-ADL and HOS-SSS were less responsive than the UCLA but more responsive than the SF-12 scales (Table 5).

There were 3 articles that each used the combination of the iHOT-12, HAGOS, EQ-5D, and HSAS (Table 6). Overall, the iHOT-12 appeared to be the most responsive, as it had greater RE values than the HSAS, EQ-5D, and the HAGOS Physical Activity, Symptom, and Daily Activity subscales. However, the responsiveness did not greatly differ between the iHOT-12 and the HAGOS Quality of Life, Sport, or Pain subscales. The 6 HAGOS subscales were all more responsive than the EQ-5D and the HSAS, and the EQ-5D was also more responsive than the HSAS (Table 7).

TABLE 3
The Most Commonly Utilized Patient-Reported Outcome Instruments

	No. of Articles (%)
Modified Harris Hip Score	109 (85)
Hip Outcome Score–Activities of Daily Living	68 (53)
Hip Outcome Score–Sports	64 (50)
Pain visual analog scale	47 (36)
Nonarthritic Hip Score	42 (33)
Short Form–12 Physical Component Scale	19 (15)
Short Form–12 Mental Component Scale	16 (12)
Western Ontario and McMaster Osteoarthritis Index	12 (9)
International Hip Outcome Tool–33	11 (9)
Harris Hip Score (full)	7 (5)
University of California at Los Angeles Activity Scale	6 (5)
Hip Disability and Osteoarthritis Outcome Score	5 (4)

DISCUSSION

In the hip arthroscopy literature, there is substantial variability in data collection and outcomes reporting. Many studies did not report routine patient characteristics including BMI, duration of symptoms, and physical examination and radiographic findings. The paucity of patient data increases the difficulty in assessing treatment efficacy and in identifying proper patient selection in the context of outcomes that were reported. Revision rates and conversions to THA were also reported only approximately 50% of the time, which precludes the assessment of the survivorship of the procedure.

More than 20 different PRO instruments have been used, which makes comparing studies difficult. This finding is not unique to hip arthroscopy, as variable outcomes reporting was previously identified in the rotator cuff, shoulder instability, and anterior cruciate ligament literature.[10,15,16] While variability in outcomes reporting appears to be common across the arthroscopy literature, it does make interpretation of the literature more difficult and limits the ability to draw conclusions from pooled data. Consistent reporting is especially important for hip arthroscopy since hip preservation is a relatively “young” subspecialty. The identification of potential subsets of patients who are more susceptible to postoperative complication and/or poor outcomes is dependent on our ability to draw meaningful conclusions from larger datasets and pooled data from the available literature; however, the ability to draw conclusions may be limited by both the variability in outcomes reporting but also the PRO instruments that are commonly used.

The majority of articles in the current review (109/130, 84%) used the mHHS. This score was originally developed in 1969 by Dr William H. Harris to assess postoperative results after arthroplasty procedures.[8] While it initially served as a valuable PRO instrument in evaluating THA outcomes, its effectiveness may be limited in accurate reflection of newer techniques and indications in both arthroplasty and arthroscopy outcomes. A systematic review by Wamper et al[19] examined the HHS in all THA

TABLE 4
Pooled Pre- and Postoperative Means and Effect Sizes for the Group of Articles That Utilized the HOS and/or NAHS With Additional PROs[a]

PRO Tool	No. of Studies	No. of Hips	Preoperative Score	Postoperative Score	ES
HOS-ADL	26	4810	63.6 ± 19.6	83.6 ± 18.7	1.02
HOS-SSS	26	4810	43.3 ± 24.7	71.7 ± 28.0	1.15
NAHS	14	3205	56.2 ± 18.6	78.5 ± 18.9	1.20
Pain VAS	12	2911	6.1 ± 2.1	3.2 ± 2.5	1.42
SF-12 PCS	5	1083	42.8 ± 14.8	52.6 ± 8.6	0.67
SF-12 MCS	4	548	44.2 ± 18.9	54.0 ± 8.5	0.52
UCLA	3	61	3.4 ± 2.2	7.3 ± 2.4	1.74

[a]ES, effect size; HOS-ADL, Hip Outcome Score–Activities of Daily Living; HOS-SSS, Hip Outcome Score–Sport; NAHS, Nonarthritic Hip Score; PRO, patient-reported outcome; SF-12 MCS, Short Form–12 Mental Component Score; SF-12 PCS, Short Form–12 Physical Component Score; UCLA, University of California at Los Angeles Activity Scale; VAS, visual analog scale.

TABLE 5
Relative Efficiency (Comparative Responsiveness) Results Between the NAHS, Pain VAS, HOS-ADL, HOS-SSS, UCLA, SF-12 PCS, and SF-12 MCS[a]

	NAHS	UCLA	Pain VAS	HOS-ADL	HOS-SSS	PCS	MCS
NAHS		–	1.40	1.43	1.42	–	–
UCLA	–		0.30	3.35	2.02	–	–
Pain VAS	0.71	3.30		1.70	1.72	–	–
HOS-ADL	0.70	0.27	0.59		0.93	2.23	3.69
HOS-SSS	0.70	0.10	0.58	1.07		2.37	3.37
SF-12 PCS	–	–	–	0.45	0.42		2.22
SF-12 MCS	–	–	–	0.27	0.30	0.45	

[a]Values <1 suggest that the PRO tool in the left column is less responsive than the corresponding PRO tool in the top row. Conversely, values >1 suggest that that the PRO tool in the left column is more responsive than the corresponding PRO tool in the top row. A dash indicates that no studies were available to compare the 2 PRO tools. See Table 4 for abbreviation definitions.

studies published in 2007. The authors found that across the studies reviewed, the ceiling effect was, on average, 20%. This finding indicates that large improvements, especially in younger or more active populations, would be missed. These effects are magnified when attempting to accurately assess hip preservation procedures outcomes. For example, when assessing a patient's walking distance capacity, the best possible option is "unlimited"; however, the second-best option is "6 blocks." This response suggests that a patient who can walk 7 blocks could be considered to have the maximum walking ability, and patients who walked 7 blocks would be equivalent with those running a 5000-m race. With our study's mean age of 37 years, a large proportion of hip arthroscopy patients can walk well over 7 blocks at a minimum follow-up of 1 year after surgery; it would be expected that a 7-block maximum is not a satisfactory result. As such, achieving the maximum possible score may be more common with this instrument when given to younger and likely more active hip arthroscopy patients, which then increases the likelihood that the mHHS may be overestimating postoperative outcomes. This concept has been previously confirmed as the mHHS has been reported to have marked ceiling effect after hip arthroscopy procedures (unpublished data, Jacob et al 2016).[11,18] The mHHS was also demonstrated to only moderately correlate with patient satisfaction,[1] which has been suggested to be related to the minimally clinically important difference achieved to reach the patient acceptable symptomatic state.[3] Similarly, other PRO instruments initially developed for evaluating osteoarthritis and/or THA outcomes, including the WOMAC, HOS, HAGOS, and HOOS, have also demonstrated notable ceiling effects when used in hip preservation populations (unpublished data, Jacob et al 2016).[11,18]

On the contrary, more general health–related quality of life scores such as the SF-12 and EQ-5D were less responsive than hip-specific PRO instruments in the current analyses. This was not driven by a potential ceiling effect like what has been reported with osteoarthritis-based instruments but rather due to the inability to accurately detect change over time. Health-related quality of life scores have been previously reported to be less responsive when used with hip osteoarthritis than hip-specific scores,[9] and the current results suggest that a similar trend is found in the hip arthroscopy literature as well.

While the utility of osteoarthritis-based and general health PRO instruments to assess outcomes after hip arthroscopy may be questioned, the results of this review provide additional guidance as to which PRO instruments are the most responsive in this patient population. The iHOT-33

TABLE 6
Pooled Pre- and Postoperative Means and Effect Sizes for the 3 Articles (524 Hips) That Used the iHOT-12 With Additional PROs[a]

	Preoperative Score	Postoperative Score	ES
iHOT-12	42.6 ± 17.3	67.0 ± 26.6	1.41
HAGOS			
Quality of Life	32.9 ± 18.2	59.0 ± 28.6	1.43
Sport	39.8 ± 20.3	66.0 ± 28.5	1.29
Physical Activity	29.3 ± 26.3	59.0 ± 33.5	1.13
Pain	56.3 ± 18.2	76.5 ± 20.8	1.11
Symptom	50.2 ± 19.2	69.3 ± 21.7	1.00
Daily Activity	60.4 ± 22.3	78.8 ± 21.7	0.83
EQ-5D	0.59 ± 0.27	0.76 ± 0.25	0.64
HSAS	3.1 ± 2.3	3.9 ± 2.2	0.34

[a]EQ-5D, EuroQol–5D; ES, effect size; HAGOS, Copenhagen Hip and Groin Outcome Score; HSAS, Hip Sports Activity Scale; iHOT-12, International Hip Outcome Tool–12; PRO, patient-reported outcome.

TABLE 7
Relative Efficiency (Comparative Responsiveness) Results Between the iHOT-12, HSAS, EQ-5D, and the 6 HAGOS Subscales[a]

	iHOT	QOL	Sport	PhysAct	Pain	Symptom	DailyAct	EQ-5D	HSAS
iHOT		0.98	1.05	1.21	1.10	1.35	1.66	2.60	9.80
HAGOS-QOL	1.02		1.07	1.23	1.12	1.37	1.68	2.64	9.96
HAGOS-Sport	0.95	0.94		1.15	1.05	1.29	1.58	2.48	9.35
HAGOS-PhysAct	0.83	0.81	0.87		1.10	1.11	1.37	2.15	8.10
HAGOS-Pain	0.91	0.90	0.96	0.91		1.23	1.51	2.37	8.93
HAGOS-Symptom	0.74	0.73	0.78	0.90	0.81		1.23	1.93	7.28
HAGOS-DailyAct	0.60	0.59	0.63	0.73	0.66	0.81		1.57	5.91
EQ-5D	0.39	0.38	0.40	0.47	0.42	0.52	0.64		3.77
HSAS	0.10	0.10	0.11	0.12	0.11	0.14	0.17	0.26	

[a]DailyAct, Daily Activity; EQ-5D, EuroQol–5D; HAGOS, Copenhagen Hip and Groin Outcome Score; HSAS, Hip Sports Activity Scale; PhysAct, Physical Activity; QOL, Quality of Life. Values <1 suggest that the patient-reported outcome (PRO) tool in the left column is less responsive than the corresponding PRO tool in the top row. Conversely, values >1 suggest that that the PRO tool in the left column is more responsive than the corresponding PRO tool in the top row.

has been previously suggested by Ramisetty et al[18] to be the most appropriate PRO tool to assess outcomes for hip arthroscopy patients, as it demonstrated the best overall psychometric properties when compared with the HOS, NAHS, HAGOS, HOOS, and HHS. The iHOT-12 is a shortened version of the iHOT-33, and scores of the iHOT-12 have been previously reported to agree well with the full version of the iHOT. In the current analyses, the iHOT-12 was more responsive than the HSAS and many of the HAGOS subscales, which further supports increased utilization in the hip arthroscopy literature. Our results are also in agreement with the findings of Kemp et al[11] in a small hip arthroscopy patient population, which found that the HOOS and iHOT-33 have better psychometric properties than the mHHS, HOS, and some subscales of the HAGOS.

While not directly compared with the iHOT-12 in the current analyses, the NAHS was more responsive than the HOS. One of the criticisms of the NAHS made by Ramisetty et al[18] was that the responsiveness of the instrument had not been established. The large effect sizes in the current study suggest that the NAHS is, indeed, responsive. We were unable to directly compare the NAHS to the iHOT-12 since no article utilized both instruments. Future comparative studies examining these 2 PRO tools are needed to determine if the NAHS may be a viable option to assess outcome after hip arthroscopy.

Our study carried common limitations of systematic reviews. By limiting our search to the past 5 years, there is the potential that additional articles that could have been utilized to assess the comparative responsiveness of PRO instruments were not identified. We selected a 5-year time frame to identify more current trends in outcomes reporting for hip arthroscopy. All intra-articular and related injuries treated with hip arthroscopy were reviewed for the purposes of this study. Our goal was not to judge treatment efficacy but to identify what outcomes measures were being reported and assess the responsiveness of those measures. Many articles were also excluded from the responsiveness analyses as only 32 of 130 articles reported pre- and postoperative means and SDs required for those calculations. While we were unable to perform a responsiveness analysis for all articles, we believe the strict criteria better preserved the integrity of our study. Future articles should report means and SDs to better enable comparisons among studies and tools.

Outcomes reporting in the hip arthroscopy literature is highly variable. More than 20 different PRO instruments have been used, which makes comparison across studies difficult. Agreement upon a uniform set of measures to assess outcomes would allow for clearer interpretation of the hip arthroscopy literature and offer the potential to draw conclusions from pooled data. On the basis of our comparative responsiveness results and previously reported psychometric properties of the different PRO instruments by Ramisetty et al,[18] we recommend more widespread adoption of the iHOT PRO instruments and that future studies are necessary to determine if the NAHS may also be a viable option for assessing hip arthroscopy outcomes. To ultimately judge the efficacy of hip arthroscopy techniques, consistent and reliable outcomes reporting is necessary across studies.

REFERENCES

1. Aprato A, Jayasekera N, Villar RN. Does the Modified Harris Hip Score reflect patient satisfaction after hip arthroscopy? *Am J Sports Med*. 2012;40(11):2557-2560.
2. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol*. 1997;24(4):799-802.
3. Chahal J, Van Thiel GS, Mather RC III, et al. The Patient Acceptable Symptomatic State for the Modified Harris Hip Score and Hip Outcome Score among patients undergoing surgical treatment for femoroacetabular impingement. *Am J Sports Med*. 2015;43(8):1844-1849.
4. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New Jersey: Lawrence Erlbaum; 1988.
5. Colvin AC, Harrast J, Harner C. Trends in hip arthroscopy. *J Bone Joint Surg Am*. 2012;94(4):e23.
6. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short Form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res*. 1999;12(3):172-179.
7. Dunlop W, Cortina J, Vaslow J, Burke M. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods*. 1996;1(2):170-177.
8. 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.
9. Hoeksma HL, Van Den Ende CH, Ronday HK, Heering A, Breedveld FC. Comparison of the responsiveness of the Harris Hip Score with generic measures for hip function in osteoarthritis of the hip. *Ann Rheum Dis*. 2003;62(10):935-938.
10. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med*. 2015;43:2572-2582.
11. 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(9):2065-2073.
12. Lam ETP, Lam CLK, Fong DYT, Huang WW. Is the SF-12 version 2 health survey a valid and equivalent substitute for the SF-36 version 2 health survey for the Chinese? *J Eval Clin Pract*. 2013;19:200-208.
13. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum*. 1985;28(5):542-547.
14. 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.
15. Lukenchuk J, Sims LA, Shin JJ. Variability in outcome reporting for operatively managed anterior glenohumeral instability: a systematic review. *Arthroscopy*. 2017;33(2):477-483.
16. Makhni EC, Padaki AS, Petridis PD, et al. High variability in outcome reporting patterns in high-impact ACL literature. *J Bone Joint Surg Am*. 2015;97(18):1529-1542.
17. Nilsdotter A-K, Roos EM, Westerlund JP, Roos HP, Lohmander LS. Comparative responsiveness of measures of pain and function after total hip replacement. *Arthritis Care Res*. 2001;45:258-262.
18. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery: a systematic review. *J Hip Preserv Surg*. 2015;2(1):15-27.
19. 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(6):703-707.