Responsiveness and Internal Validity of Common Patient-Reported Outcome Measures Following Total Shoulder Arthroplasty

Aaron D. Sciascia, PhD, ATC, PES; Brent J. Morris, MD; Cale A. Jacobs, PhD; T. Bradley Edwards, MD

abstract

The Constant–Murley (Constant) score, Western Ontario Osteoarthritis of the Shoulder (WOOS) index, American Shoulder and Elbow Surgeons (ASES) score, and Single Assessment Numeric Evaluation (SANE) score are commonly used to assess patient-reported function following shoulder surgery. However, psychometric properties for these tools are mostly unknown for patients with primary glenohumeral arthritis who have undergone anatomic total shoulder arthroplasty (TSA). The purposes of this study were to (1) compare the responsiveness and internal validity between the 4 patient-reported outcomes (PROs) and (2) identify PRO score values associated with patient satisfaction after TSA. A total of 234 primary TSAs were performed for primary glenohumeral osteoarthritis with a 2-year or greater follow-up. The Constant score, WOOS index, ASES score, SANE score, and patient satisfaction were assessed preoperatively and 2 to 5 years postoperatively. Effect sizes, standardized response means, and relative efficiency were calculated to determine responsiveness, and internal validity was determined via the presence of floor and/or ceiling effects. Receiver operator characteristic (ROC) curves were constructed to identify the minimum outcome score that could correctly identify a satisfied patient. At final follow-up, 88% of patients were satisfied. The PROs had large effect sizes and standardized response means (≥0.83). The minimum score that most correctly identified a patient as satisfied was 78 for ASES score, 18 for WOOS index, 73 for Constant score, and 58 for SANE score. However, the ASES score, WOOS index, and SANE score had marked postoperative ceiling effects, whereas the Constant score was the most responsive and internally valid tool. These results suggest that the Constant score should serve as the primary PRO for patients with primary glenohumeral arthritis, whereas the WOOS index, ASES score, and SANE score could be supplementary assessments. [*Orthopedics*. 201x; xx(x):xx-xx.]

A recognized paradigm shift exists in modern-day health care, with a transition from surgeon-driven treatment decisions based on clinical experience to a focus on improved patient engagement with a shared decision-making process.[1,2] As such, patient perception and patient satisfaction are becoming important components of health care outcomes. In an era of public reporting and pay-for-performance, orthopedic surgeons need to better understand the association between patient-reported outcome (PRO) measures and patient satisfaction. Reimbursement for health care services is routinely being linked to patient satisfaction as part of third-party reimbursement initiative programs.[3] However, there is little guidance as to which PRO measures should be used or which values relate to a patient being satisfied following surgery.

The authors are from Eastern Kentucky University (ADS), Richmond, Lexington Clinic Orthopedics–The Shoulder Center of Kentucky (BJM), Lexington, and the Department of Orthopedic Surgery (CAJ), University of Kentucky, Lexington, Kentucky; and the Fondren Orthopedic Group, Texas Orthopedic Hospital (TBE), Houston, Texas.

Drs Sciascia and Jacobs have no relevant financial relationships to disclose. Dr Morris is a paid consultant for Tornier and Wright Medical Group. Dr Edwards is a paid consultant for Tornier, Wright Medical Group, and Smith & Nephew and receives royalties from Tornier and Wright Medical Group.

Correspondence should be addressed to: Aaron D. Sciascia, PhD, ATC, PES, Eastern Kentucky University, 521 Lancaster Ave, 228 Moberly Bldg, Richmond, KY 40475 (aaron.sciascia@eku.edu).

Received: October 27, 2016; Accepted: February 15, 2017.

doi: 10.3928/01477447-20170327-02

There has also been a lack of consensus and great variability in the application of PROs in shoulder research over the past decade.[4]

Selecting and interpreting PRO instruments to assess the patient's perception of his or her pain and ability to physically function can be challenging because many instruments exist.

Shoulder pain and function have been assessed using joint-specific, disease-specific, and other, more generalized, global assessments.[5-11] The American Shoulder and Elbow Surgeons (ASES) score[6] is a commonly used joint-specific PRO instrument. The ASES score has known psychometric properties that were established in patients primarily treated nonoperatively for various shoulder conditions.[12-15] The 11-question, Likert scale–designed ASES score has a reported administration time of 3 to 5 minutes, with an additional 2 minutes for scoring.[12] The ASES score has been used frequently as one of the postsurgical PRO measures in total shoulder arthroplasty (TSA),[16-20] although psychometric assessments have not often focused on the usefulness of the measure in TSA patients.

The Constant–Murley shoulder (Constant) score incorporates both patient subjectivity and clinical objectivity[5] (specifically strength and range of motion), has established psychometric properties,[21] and has often been used in studies focused on outcomes following TSA.[20,22] The time to complete the Constant score assessment is between 5 and 7 minutes,[23] with approximately 2.5 minutes reserved for the patient component.[24]

Disease-oriented measures have also been advocated for assessing patient outcomes following TSA, specifically the Western Ontario Osteoarthritis of the Shoulder (WOOS) index.[7] The WOOS index contains 19 questions characterized by a visual analog scale design.[7] Previous authors have noted that the time to administer the WOOS index is approximately 10 minutes, with the ease of scoring the instrument described as moderate due to the manual measurement required for each answered question and the resultant calculation.[25]

Finally, the Single Assessment Numeric Evaluation (SANE) score[8] has been advocated as a convenient measure of patient physical function because of its single question design.[26] However, a noted limitation of the SANE measure is that the psychometric properties, such as reliability and validity, have not been thoroughly examined.[27] Recent literature has reported the SANE score to be correlated with the Rowe and ASES scores[8] and to be predictive of the results from other shoulder-specific outcome measures in military personnel.[28] However, its usefulness for TSA outcomes has not been concretely established.

To date, the relationship between commonly used upper extremity PRO tools and patient satisfaction after anatomic TSA has not been evaluated. In addition, psychometric properties for these tools are mostly unknown for patients with primary glenohumeral arthritis who have undergone anatomic TSA. Therefore, the primary purpose of this study was to compare the responsiveness and internal validity of the ASES score, Constant score, WOOS index, and SANE score in patients with primary glenohumeral arthritis. It was hypothesized that the ASES score would demonstrate superior responsiveness and internal validity when compared with the Constant score, WOOS index, and SANE score. The secondary purpose was to identify PRO score values for these outcome questionnaires that are associated with patient satisfaction after anatomic TSA to provide a target value for postsurgical goal setting.

Materials and Methods

All patients who underwent primary anatomic TSA for the treatment of primary glenohumeral osteoarthritis with a minimum of 2-year follow-up were identified in a prospective shoulder arthroplasty registry. A single surgeon (T.B.E.) performed all procedures at a high-volume orthopedic hospital. All patients who underwent revision surgery or had an intra- or postoperative complication were excluded.

The Constant score, the ASES score, the WOOS index, the SANE score, and patient satisfaction were assessed preoperatively and 2 to 5 years postoperatively. All patients included in this study completed all self-reported outcome measures both pre- and postoperatively. Patient satisfaction was determined by asking the patient to rate themselves as being "very satisfied," "satisfied," "dissatisfied," or "very dissatisfied." Similar to previous studies, patients who were very satisfied or satisfied were categorized as being satisfied for the purposes of this study. Patients who were very dissatisfied or dissatisfied were considered dissatisfied.[29] Patients were provided the outcome measures by a member of the research team. Patients completed the questionnaires without assistance from the medical provider at the time of surgery scheduling but were permitted to ask questions at any time.

The Aequalis, Aequalis Ascend, and Aequalis Ascend Flex (Tornier, Bloomington, Minnesota) anatomic TSA systems were used. The TSA surgical technique is well described,[30,31] and a standardized rehabilitation protocol was followed.[30,32] All patients were prospectively enrolled in a shoulder arthroplasty outcomes registry, and the senior author (T.B.E.) performed all examinations. A handheld goniometer was used to calculate range of motion (ROM) measurements. A handheld digital dynamometer was used to measure strength of abduction (Chatillon Digital Force Gauge 200 lbf; AMETEK, Inc, Largo, Florida).

Statistical Analyses

Summary statistics for demographic items were calculated and reported as means and standard deviations for continuous variables, whereas frequencies and percentages were reported for categorical variables.

Table 1

Patient Population Demographic Summary

Demographic	Overall (n=234)
Age, y	
Mean (SD)	67 (10)
Range	38-93
Height, in	
Mean (SD)	68 (4)
Range	57-78
Weight, lb	
Mean (SD)	198 (41)
Range	118-328
Body mass index, kg/m^2	
Mean (SD)	30 (6)
Range	19-50
Sex, No. (%)	
Male	153 (65)
Female	81 (35)

Table 2

Preoperative and Postoperative Patient-Reported Outcome Measure Scores

		Mean±SD		
Patient-Reported Outcome Measure	Scale	Preoperative	Postoperative	*P*
Constant score	0-100	29±17	80±15	<.001
WOOS index[a]	100-0	66±18	13±19	<.001
ASES score	0-100	41±19	89±17	<.001
SANE score	0-100	32±25	69±36	<.001

Abbreviations: ASES, American Shoulder and Elbow Surgeons; Constant, Constant–Murley; SANE, Single Assessment Numeric Evaluation; WOOS, Western Ontario Osteoarthritis of the Shoulder.
[a]The WOOS index is scored 1900-0 (low to high function), but has been reported as a percentage to assist in interpreting it similar to other outcome measures.

Responsiveness. Effect sizes, standardized response means, and relative efficiency were calculated to establish responsiveness of each outcome tool.[33-35] Effect size was calculated by dividing the mean change score for each outcome tool by the standard deviation of the preoperative score for each tool (ie, mean change in ASES score pre- to postoperatively/preoperative ASES standard deviation). Cohen defined effect sizes as small (d=0.2), medium (d=0.5), and large (d=0.8).[36,37]

The standard response mean (SRM) is an additional measure of responsiveness that is calculated by dividing the mean change score for an outcome tool by the standard deviation of the change score of the same tool.[34] The relative efficiency was derived from the resultant t-score as provided by paired t tests that compared the preoperative score for an outcome tool to the postoperative score. Relative efficiency was calculated by dividing the t-score for one PRO by the t-score of another PRO tool and then squaring the resultant value.[34] A relative efficiency value greater than 1 indicated that an outcome tool was more responsive compared with another.[34]

Internal Validity. Floor and ceiling effects were defined as the number of patients who achieved the lowest (floor) and highest (ceiling) possible score for each outcome tool.[38] If 15% or more of the patients had either the lowest or highest score for an instrument, floor and/or ceiling effects were considered to be present.[38,39]

Patient Satisfaction. Receiver operator characteristic (ROC) curves were constructed to identify the minimum outcome score value that could best predict patient satisfaction. The area under the curve (AUC), percentage correctly classified, and positive and negative likelihood ratios for each outcome instrument are reported. Odds ratios were calculated to determine the odds of a patient being satisfied based on the results derived from the ROC curve. Statistical analyses were performed using STATA 13.1 (StatCorp, LP, College Station, Texas) and SPSS Statistics (IBM, Armonk, New York), with P<.05 being considered statistically significant.

Results

Demographic Summary

A total of 234 patients (153 men, 81 women) were identified for inclusion in this study. Demographic variables are reported in **Table 1**. Mean follow-up for this series was 3±1 years (range, 2-5 years). Of the 234 TSAs performed, 207 (88%) patients reported being satisfied, whereas 27 (12%) patients were not satisfied at final follow-up. Mean total scores significantly improved for the all PRO measures between the preoperative visit and the most recent follow-up (all P<.001) (**Table 2**).

Responsiveness

The responsiveness metrics are reported in **Table 3**. The outcome tools with the highest effect size and standardized response means were the WOOS index (effect size=3.0, SRM=2.4) and Constant score (effect size=2.9, SRM=2.4), whereas the SANE score was the least responsive (effect size=1.5, SRM=0.83). The relative efficiency values for the outcome tools are provided in **Table 4,** with the Constant score being the most responsive (relative efficiency >1) compared with all other outcome tools in this study.

Internal Validity

A low ceiling effect was noted with the Constant score (3% of patients had the highest possible score), whereas marked postoperative ceiling effects were identi-

Table 3

Responsiveness Metrics for the Constant, WOOS, ASES, and SANE Outcome Tools

Patient-Reported Outcome Measure	Effect Size	SRM	Floor	Ceiling
Constant score	2.9	2.4	1%	3%
WOOS index	3.0	2.4	1%	19%[a]
ASES score	2.5	2.2	1%	21%[a]
SANE score	1.5	0.83	4%	19%[a]

Abbreviations: ASES, American Shoulder and Elbow Surgeons; Constant, Constant–Murley; SANE, Single Assessment Numeric Evaluation; SRM, standardized response mean; WOOS, Western Ontario Osteoarthritis of the Shoulder.
[a]Outcome tool with a marked ceiling effect (>15%).

Table 4

Relative Efficiency Values for the Constant, WOOS, ASES, and SANE Outcome Tools

Patient-Reported Outcome Measure	Constant Score	WOOS Index	ASES Score	SANE Score
Constant score	-	1.1	1.3	9.2
WOOS index	0.9	-	1.2	8.0
ASES score	0.8	0.9	-	6.9
SANE score	0.1	0.1	0.1	-

Abbreviations: ASES, American Shoulder and Elbow Surgeons; Constant, Constant–Murley; SANE, Single Assessment Numeric Evaluation; WOOS, Western Ontario Osteoarthritis of the Shoulder.

Table 5

Receiver Operating Characteristics for Assessing Patient Satisfaction

				Likelihood Ratio		
Patient-Reported Outcome Measure	Score	AUC	Correctly Classified	Positive	Negative	Odds Ratio (95% Confidence Interval)
ASES score	78	0.88	89%	2.04	0.06	34 (12-95)
WOOS index	18	0.87	86%	1.74	0.06	28 (10-79)
Constant score	73	0.85	88%	1.85	0.07	28 (10-76)
SANE score	58	0.76	75%	1.34	0.13	10 (4-26)

Abbreviations: ASES, American Shoulder and Elbow Surgeons; AUC, area under the curve; Constant, Constant–Murley; SANE, Single Assessment Numeric Evaluation; WOOS, Western Ontario Osteoarthritis of the Shoulder.

fied with the ASES score (21%), WOOS index (19%), and SANE score (19%). Floor effects were not noted with any of the 4 PRO tools.

Patient Satisfaction

The ROC analysis results are detailed in **Table 5**. The AUC for each outcome instrument was 0.74 or greater, suggesting high ability to predict satisfaction. However, each outcome instrument was determined to have positive likelihood ratios that would result in small changes from pre- to post-test probability[40] when using the threshold values to predict satisfaction (positive likelihood ratios: ASES score=2.04, Constant score=1.85, WOOS index=1.74, and SANE score=1.34).

Discussion

Selecting and interpreting PRO instruments to assess the patient's perception of his or her pain and ability to physically function can be challenging. Many instruments exist and it is unclear which instruments are preferred. To help alleviate some of the clinician burden associated with interpreting outcome scores, this study aimed to compare the responsiveness and internal validity of commonly used PROs and establish minimum scores that could identify a patient as being satisfied following anatomic TSA. Overall, the PROs examined in this study were responsive, having large effect sizes and SRMs (all greater than 1).

The primary hypothesis that the ASES score would have superior responsiveness and internal validity compared with the other outcome instruments examined in this study was rejected. Conversely, the Constant score and WOOS index were the most responsive; however, the Constant score demonstrated the best relative efficiency and was the only outcome tool to have almost no floor or ceiling effects. These results suggest that the Constant score should serve as the primary PRO for TSA for patients with primary glenohumeral arthritis, whereas the WOOS index, ASES score, and SANE score could serve as supplementary assessment measures.

There are noted limitations with this study. First, not all existing PROs were examined in this study. It is possible that an outcome measure not used in this study may have a stronger correlation to patient satisfaction or higher psychometric results. Second, not all PROs distributed in this study have undergone rigorous psychometric assessment for patients with primary glenohumeral arthritis until now. Future investigations should attempt to replicate this study with similar methodology and patient population to confirm the results. Finally, previous reports have suggested that a sex and age adjusted Constant score be used to account for confounding variables that may influence the interpretation of the traditional Constant score.[41] However, the adjusted Constant score was originally derived from a group of subjects without shoulder pathology, so the current authors did not believe it would be appropriate to include the adjusted score in the current study's analysis. Several strengths of the study are also highlighted. The study consisted of a large homogeneous cohort of patients with the same preoperative diagnosis and the same surgical treatment. All procedures were performed by the same surgeon using the same surgical technique and postoperative rehabilitation protocol. Furthermore, all of the data was collected prospectively in a shoulder arthroplasty registry.

The current study hypothesized that the ASES score would demonstrate superior responsiveness and internal validity compared with the Constant score, WOOS index, and SANE score due to the ASES score being often used as one of the PROs in TSA patients.[16-20] In addition, one group examined the usefulness and validity of multiple PROs, finding both the Constant and ASES scores after TSA to be adequately responsive for patient assessment.[20] However, the current study found the Constant score to be more responsive than the ASES score. This difference from the previous work could be contributed to the variation in follow-up time between the studies (6 months vs 2 to 5 years in the current study) and sample size (81 more patients in the current study). The current study also found the WOOS index to be adequately responsive, although as with the ASES and SANE scores, it demonstrated ceiling effects (>15% of the patient sample had the highest possible scores) but the Constant score did not. This suggests that although it would not be inappropriate to use the WOOS index, ASES score, or SANE score in clinical practice, there may be instances where some patients' perception of pain and/or physical function may exceed the highest score over time, thus not allowing progress to be accurately gauged.

Previous work has suggested that patient satisfaction after TSA for primary glenohumeral osteoarthritis appears to be driven by subjective (patient-oriented) factors rather than objective (clinician-oriented) factors.[42] However, because the Constant score was found to be the most responsive instrument and to have both low floor and ceiling effects in the current study, it is possible that the design of the Constant score may help balance the information obtained from both clinician- and patient-oriented assessments, thus allowing it to have stronger internal validity.

For example, the Constant score incorporates both patient subjectivity and clinical objectivity,[5] whereas the WOOS index and ASES score are exclusively based on patient perception. These differences in instrument design (ie, the lack of an objective component with the WOOS index and ASES score) may be contributing to the high ceiling effects within those tools. Although the current study identified threshold values for each outcome tool, the positive likelihood ratios, which attempt to provide clinicians with an estimate of how much change will occur from pre-to post-test, suggest that as individual assessments of patient status, there will be small changes to the post-test probability of identifying a patient as satisfied (positive likelihood ratios=1.3-2.0).[40] In other words, the threshold values would have minimal clinical utility as individual determinants of patient satisfaction. As such, clinicians should not base clinical judgments about patient status solely on either subjective or objective findings as isolated entities but should use an integrative approach comprised of assessments for both patient perception and physical task performance.

Finally, although the SANE score has been advocated as a convenient measure of a patient's global perception of his or her shoulder because of its single question design,[26] it was found to be the least responsive compared with the other PRO measures in the current study. On further inspection of the data, 25% of patients reported a SANE score that was opposite that of their satisfaction status. For example, 11% of unsatisfied patients reported a SANE score of 86 or higher, whereas 14% of the satisfied patients reported a score of 10 or less. This implies that patients may have inadvertently inverted the scale where the lowest score indicating complete disability would be interpreted as ideal (and vice versa).

To further investigate the impact this misinterpretation may have had on the results, the authors revisited the data, converting the potentially misinterpreted value to its expected value based on each patient's satisfaction response. For example, if a patient reported being "satisfied" or "very satisfied" following surgery but reported a SANE value of 10, the SANE value was converted to 90. This was performed for all occurrences where a patient reported being satisfied and had a SANE score of 10 or less. Conversely, patients who reported being "unsatisfied" or "very unsatisfied" and had a SANE score of 86 or greater the value was converted to 14 or less as mathematically appropriate. After the conversions were performed, the average postoperative SANE score increased from a 69 to an 86. This suggests that, although convenient to implement, patients must be properly educated on how to respond to a single question outcome measure. Based on these findings, it is recommended that distribution of the SANE score occur with direct patient interaction to address any potential confusion or concern with the meaning of the question and that examiners should review the scores with the patient to ensure accuracy.

Conclusion

With the initiative of pay-for-performance entering health care systems, there is a need to assist clinicians with PRO selection and interpretation because patient perception has become an integral piece of the treatment process. All of the measures examined in this study may serve as adequate PRO assessment options for patients undergoing anatomic TSA; however, the Constant score was the most responsive outcome tool. When selecting a PRO, both completion and scoring burden should be considered, as well as understanding that complementary subjective and objective information appears to better relate to patient satisfaction compared with individual subjective or objective assessments of pain and/or physical function. The minimum values for each PRO identified in this study could be used for postsurgical goal setting purposes but clinicians should treat each patient based on individual presentation and needs.

References

1. Adams JR, Drake RE. Shared decision-making and evidence-based practice. *Community Ment Health J.* 2006; 42(1):87-105.
2. Barratt A. Evidence based medicine and shared decision making: the challenge of getting both evidence and preferences into health care. *Patient Educ Couns.* 2008; 73(3):407-412.
3. Morris BJ, Richards JE, Archer KR, et al. Improving patient satisfaction in the orthopaedic trauma population. *J Orthop Trauma.* 2014; 28(4):e80-e84.
4. Gartsman GM, Morris BJ, Unger RZ, Laughlin MS, Elkousy HA, Edwards TB. Characteristics of clinical shoulder research over the last decade: a review of shoulder articles in The Journal of Bone & Joint Surgery from 2004 to 2014. *J Bone Joint Surg Am.* 2015; 97(5):e26.
5. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res.* 1987; (214):160-164.
6. Richards RR, An KN, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg.* 1994; 3(6):347-352.
7. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: the Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage.* 2001; 9(8):771-778.
8. Williams GN, Gangel TJ, Arciero RA, Uhorchak JM, Taylor DC. Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales: outcomes measures after shoulder surgery. *Am J Sports Med.* 1999; 27(2):214-221.
9. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder and Hand). The Upper Extremity Collaborative Group (UECG) [published correction appears in *Am J Ind Med.* 1996; 30(3):372]. *Am J Ind Med.* 1996; 29(6):602-608.
10. Alberta FG, ElAttrache NS, Bissell S, et al. The development and validation of a functional assessment tool for the upper extremity in the overhead athlete. *Am J Sports Med.* 2010; 38(5):903-911.
11. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability: the Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med.* 1998; 26(6):764-772.
12. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Assessment Form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elbow Surg.* 2002; 11(6):587-594.
13. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am.* 2010; 92(2):296-303.
14. Placzek JD, Lukens SC, Badalanmenti S, et al. Shoulder outcome measures: a comparison of 6 functional tests. *Am J Sports Med.* 2004; 32(5):1270-1277.
15. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am.* 2005; 87(9):2006-2011.
16. Sershon RA, Van Thiel GS, Lin EC, et al. Clinical outcomes of reverse total shoulder arthroplasty in patients aged younger than 60 years. *J Shoulder Elbow Surg.* 2014; 23(3):395-400.
17. Latif V, Denard PJ, Young AA, Liotard JP, Walch G. Bilateral anatomic total shoulder arthroplasty versus reverse shoulder arthroplasty. *Orthopedics.* 2012; 35(4):e479-e485.
18. Razmjou H, Stratford P, Kennedy D, Holtby R. Pattern of recovery following total shoulder arthroplasty and humeral head replacement. *BMC Musculoskelet Disord.* 2014; 15:306.
19. Angst F, Pap G, Mannion AF, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Rheum.* 2004; 51(5):819-828.
20. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer HK, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Rheum.* 2008; 59(3):391-398.
21. Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. *J Shoulder Elbow Surg.* 2010; 19(1):157-164.
22. Lo IK, Litchfield RB, Griffin S, Faber K, Patterson SD, Kirkley A. Quality-of-life outcome following hemiarthroplasty or total shoulder arthroplasty in patients with osteoarthritis: a prospective, randomized trial. *J Bone Joint Surg Am.* 2005; 87(10):2178-2185.
23. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res (Hoboken).* 2011; 63(suppl 11):S174-S188.
24. Noorani AM, Roberts DJ, Malone AA, et al. Validation of the Stanmore percentage of normal shoulder assessment. *Int J Shoulder Surg.* 2012; 6(1):9-14.
25. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis.* 2004; 63(4):335-341.
26. Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg.* 2010; 18(7):436-444.
27. Smith MV, Calfee RP, Baumgarten KM, Brophy RH, Wright RW. Upper extremity-specific measures of disability and outcomes in orthopaedic surgery. *J Bone Joint Surg Am.* 2012; 94(3):277-285.
28. Provencher MT, Frank RM, Macian D, et al.

An analysis of shoulder outcomes scores in 275 consecutive patients: disease-specific correlation across multiple shoulder conditions. *Mil Med.* 2012; 177(8):975-982.

29. Jacobs CA, Morris BJ, Sciascia AD, Edwards TB. Comparison of satisfied and dissatisfied patients 2 to 5 years after anatomic total shoulder arthroplasty. *J Shoulder Elbow Surg.* 2016; 25(7):1128-1132.
30. Edwards TB, Labriola JE, Stanley RJ, O'Conner DP, Elkousy HA, Gartsman GM. Radiographic comparison of pegged and keeled glenoid components using modern cementing techniques: a prospective randomized study. *J Shoulder Elbow Surg.* 2010; 19(2):251-257.
31. Gartsman GM, Edwards TB. *Shoulder Arthroplasty.* Philadelphia, PA: Saunders Elsevier; 2008.
32. Liotard JP, Edwards TB, Padey A, Walch G, Boulahia A. Hydrotherapy rehabilitation after shoulder surgery. *Tech Shoulder Elbow Surg.* 2003; 4(2):44-49.
33. Lam ET, Lam CL, Fong DY, 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(1):200-208.
34. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990; 28(7):632-642.
35. Nilsdotter AK, Roos EM, Westerlund JP, Roos HP, Lohmander LS. Comparative responsiveness of measures of pain and function after total hip replacement. *Arthritis Rheum.* 2001; 45(3):258-262.
36. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
37. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol Methods.* 1996; 1(2):170-177.
38. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open.* 2015; 5(7):e007765.
39. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res.* 1995; 4(4):293-307.
40. Jaeschke RZ, Guyatt GH, Sackett DL. Users' guides to the medical literature: III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994; 271(9):703-707.
41. Katolik LI, Romeo AA, Cole BJ, Verma NN, Hayden JK, Bach BR. Normalization of the Constant score. *J Shoulder Elbow Surg.* 2005; 14(3):279-285.
42. Chen AL, Bain EB, Horan MP, Hawkins RJ. Determinants of patient satisfaction with outcome after shoulder arthroplasty. *J Shoulder Elbow Surg.* 2007; 16(1):25-30.