This document shows the changes between the original (retracted) version of the article and the current (updated) version.

Background:

Health-related quality-of-life (HRQoL) scores are required for cost-effectiveness and health-care value analysis. We evaluated HRQoL scores and patient-reported outcome measures (PROMs) in patients with advanced glenohumeral osteoarthritis treated with anatomic total shoulder arthroplasty to establish values of HRQoL scores that can be used for cost-effectiveness and value analysis and to assess relationships between HRQoL scores and shoulder and upper-extremity PROMs.

Methods:

We analyzed ~~145~~143 patients (~~145~~143 shoulders) with glenohumeral osteoarthritis treated with anatomic total shoulder arthroplasty; ~~93~~92 patients had 1-year follow-up. Preoperative and postoperative functional outcomes were assessed with the Disabilities of the Arm, Shoulder and Hand (DASH) score, the American Shoulder and Elbow Surgeons (ASES) score, the Simple Shoulder Test (SST), and a visual analog scale (VAS) for shoulder pain and function. Health utility was assessed with the EuroQol-5 Dimensions (EQ-5D), Short Form-6 Dimensions (SF-6D), and VAS Quality of Life (VAS QoL). HRQoL score validity was determined through correlations between the PROMs and HRQoL scores. The responsiveness of HRQoL scores was measured through the effect size and the standardized response mean.

Results:

There were significant improvements in all PROMs and HRQoL scores (p < 0.001) at 1 year after the surgical procedure. The changes in VAS QoL ~~and~~ (very weak to moderate), EQ-5D (weak), and SF-6D (weak) were significantly correlated (~~weak to moderate~~p < 0.05) with the changes in ~~all~~ PROMs ~~except the SST~~, demonstrating comparably acceptable validity. ~~The VAS QoL had a~~There were large effect ~~size~~sizes in the VAS QoL (1.~~833)~~843), EQ-5D (1.186), and SF-6D (1.084) and large standardized response mean ~~(1.603), and~~values in the VAS QoL (1.622), EQ-5D ~~also had a large effect size (1.163)~~(1.230), and ~~standardized response mean~~SF-6D (1.~~228~~083), demonstrating responsiveness. The effect sizes of all PROMs were larger than those of the HRQoL scores~~. The change in SF-6D had only a moderate effect size and standardized response mean and was not significantly correlated with the change in any of the PROMs~~.

Conclusions:

PROMs and HRQoL scores are not interchangeable, and studies of the cost-effectiveness and value of shoulder arthroplasty should incorporate both shoulder and upper-extremity PROMs and HRQoL scores. The findings of this study provide data on HRQoL scores that are specific to the treatment of advanced glenohumeral osteoarthritis with anatomic total shoulder ~~replacement~~arthroplasty and can be used for future cost-effectiveness and value analysis studies.

Level of Evidence:

Therapeutic <u>Level IV</u>. See Instructions for Authors for a complete description of levels of evidence.

Ever-increasing health-care costs have driven developed countries to focus on value-based health care to manage cost and to optimize outcome. Patient-reported outcome measures (PROMs) are used to evaluate functional outcomes following shoulder arthroplasty[1-3]. However, shoulder and upper-extremity PROMs cannot be used in an analysis of comparative value relative to other orthopaedic and non-orthopaedic conditions.

Quality-adjusted life-years (QALYs) are a summary measure of health outcome that combines the impact of a treatment on a patient's length of life and health-related quality of life (HRQoL) and can be used to compare alternative treatments of a specific condition, as well as treatments of disparate conditions[4,5]. QALYs are used with costs (direct and indirect) to determine cost utility (monetary cost/QALY), which can be used to perform cost utility analysis.

The determination of QALYs requires HRQoL scores, which weight the preference for various health states. QALYs are the product of the time spent in a particular health state and the health utility associated with that health state[6]. The EuroQol-5 Dimensions (EQ-5D)[7], Short Form-6 Dimensions (SF-6D)[8], and Visual Analog Scale Quality of Life (VAS QoL) are the commonly used health utility indices[9]. These metrics have not been widely used to assess HRQoL in patients with glenohumeral osteoarthritis[5,10-12].

Anatomic total shoulder arthroplasty is the gold-standard surgical treatment for advanced glenohumeral osteoarthritis, and patients have substantial and significant improvement in both pain and shoulder function. However, improvements in quality of life have not been thoroughly evaluated. Carter et al. demonstrated a significant improvement in the physical component of the Short Form-36 (SF-36), indicating an improvement in a patient's overall physical well-being following an anatomic total shoulder arthroplasty[13]. Cho et al. reported significant improvement in HRQoL assessed with the World Health Organization Quality of Life Scale Abbreviated Version at 1 year after an anatomic total shoulder arthroplasty[14].

Although PROMs were developed as joint, diagnosis, and region-specific outcome tools, HRQoL measures were not. Therefore, if the outcome of anatomic total shoulder arthroplasty is assessed with HRQoL scores, the psychometric properties, especially the validity and responsiveness, need to be defined and understood. The primary goal of this study was to establish the values of HRQoL that can be used for cost-effectiveness and value analysis research and to assess relationships between PROMs and HRQoL measures in patients with advanced primary glenohumeral osteoarthritis treated with anatomic total shoulder arthroplasty. We hypothesized that PROMs and HRQoL scores would improve from baseline to 1 year after anatomic total shoulder arthroplasty and that the different PROMs would have varying correlations with the HRQoL scores. We also hypothesized that the HRQoL scores would be both valid and responsive to changes in health after anatomic total shoulder arthroplasty.

Materials and Methods

Institutional review board approval was obtained to perform this study. Patients treated with anatomic total shoulder arthroplasty for primary glenohumeral osteoarthritis between November 2011 and February 2015 were included. The data were collected and were recorded prospectively, were maintained in a clinical database, and were retrospectively studied. Patients with other etiologies of glenohumeral arthropathy were excluded to limit the focus of this study to a specific treatment of a single unique diagnosis. Patients who did not complete the 1-year evaluation were excluded from the analysis.

Baseline Evaluation

Patients completed baseline questionnaires within 2 weeks of the surgical procedure with regard to demographic characteristics (age, sex, and hand dominance), functional outcome, and health status. PROMs were assessed with a VAS for pain and function, the Disabilities of the Arm, Shoulder and Hand (DASH) score, the Simple Shoulder Test (SST), and the American Shoulder and Elbow Surgeons (ASES) score, and HRQoL was assessed with the EQ-5D, SF-6D, and VAS QoL (Table I).

Follow-up Evaluation

Patients returned for routine postoperative evaluations at 6 weeks, 3 months, 6 months, and 1 year. The same questionnaires with regard to functional outcome and health status were administered beginning at the 3-month evaluation.

Subjects

One hundred and forty-fivethree patients (145143 shoulders), 77 male and 6866 female, were treated by the senior author with a press-fit, short-stem anatomic humeral component (AEQUALIS ASCEND FLEX Convertible Shoulder Arthroplasty System; Wright Medical Group) and a hybrid, all-polyethylene, pegged glenoid (AFFINITI CORTILOC Glenoid Implant; Wright Medical Group). Ninety-threetwo patients (64%) returned as scheduled for follow-up at 1 year after the surgical procedure. Despite having scheduled appointments, not all patients returned for a 1-year postoperative evaluation. With respect to the longest follow-up for the remaining patients, 87 patients had a 3-month follow-up, 14 patients had a 6-month follow-up, 25 patients had a 2-year follow-up, 2 patients had a 3-year follow-up, 1 patient had a 4-year follow-up, and 2 patients did not have follow-up. Thus, 121 (83120 (84%) of the original 145143 patients had a minimum 1-year follow-up. The subjects with a 1-year follow-up were compared with those without a 1-year follow-up. The 95% confidence intervals (CIs) for the mean age, percentage of male sex, body mass index (BMI), percentage ofwith a dominant affected arm, and all preoperative PROMs except for the DASH score, and HRQoLsSF-6D overlapped, suggesting a lack of significant difference between the 2 groups (Table II).

There were no infections, dislocations, fractures, revision arthroplasties, or deaths. One patient had a postoperative subscapularis failure that was repaired.

Statistical Analysis

Statistical analysis was performed using the Stata Programming Language (StataCorp)[15]. Because no variable had >15% of data missing, missing values were filled using mean imputation. Significance was set a priori at $p < 0.05$. A Bonferroni correction was applied to calculate the appropriate p value for each statistical analysis to correct for the risk of type-I error inflation associated with multiple hypothesis testing.

Two-sided t tests were used to compare baseline and 1-year scores for the PROMs and HRQoL scores. Pairwise correlations between the change in score from baseline to 1 year in PROMs and HRQoL scores were calculated to measure the criterion-related validity of the HRQoL scores (i.e., how well the HRQoL scores correlated with an external criterion, in this case, the PROMs[16]). Using classifications reported by Evans, correlation magnitudes were

designated as very weak (r = 0.00 to 0.19), weak (r = 0.20 to 0.39), moderate (r = 0.40 to 0.59), strong (r = 0.60 to 0.79), or very strong (r = 0.80 to 1.00)[17].

The responsiveness of the PROMs and HRQoL scores was assessed by the effect size and the standardized response mean[18]. The effect size was calculated by dividing the difference between baseline and 1-year scores by the standard deviation of the baseline score, and the standardized response mean was calculated by dividing the difference between baseline and 1-year scores by the standard deviation of the difference between the scores[18-23]. Using the Cohen criteria, an effect size of <0.50 was deemed small, ≥0.50 to 0.80 was deemed moderate, and >0.80 was deemed large[21,24].

With regard to the post hoc power analysis, to obtain a power of 0.8 given a sample size of 9392 patients and a significance level of 0.05, the necessary correlation coefficient value was a minimum of 0.29. Therefore, any correlation coefficients of <0.29 were not considered significant, regardless of the associated p value.

Results

Change in Score from Baseline to 1 Year

There were significant (p < 0.001) and clinically relevant improvements from baseline in all of the PROMs (Table III). There were also significant improvements (p < 0.001) in all of the HRQoL scores (Table III).

Correlations Between Changes in PROMs and HRQoL Scores

There were very weak to moderate associations between the changes in PROMs and changes in HRQoL scores (Table IV). The SF-6D did notwas weakly and significantly correlate (correlated with the DASH score (r = 0.322; p > 0.0033) with any of the PROMs.= 0.002). The EQ-5D was weakly and significantly correlated with the DASH score (r = 0.3287320; p = 0.00130019) and the VAS pain score (r = 0.3240304; p = 0.00150032). The VAS QoL was significantly and weakly correlated with the ASES score (r = 0.3610340; p = 0.00040009) and the DASH score (r = 0.3706363; p = 0.00030004) and was moderately correlated with the VAS function score (r = 0.5300518; p < 0.0001) and the VAS pain score (r = 0.5843574; p < 0.0001). After the Bonferroni correction (p >> 0.05/15 = p > 0.0033) was applied, the SST was the only PROM that did not significantly correlate with any of the HRQoL scores.

Responsiveness

There was a wide range of effect size and standardized response mean values for the PROMs and HRQoL scores (Table V). The PROMs had large effect sizes: 3.679705 for the ASES score, 1.928926 for the DASH score, 3.187178 for the SST, 3.473567 for VAS function, and 3.156213 for VAS pain. Moreover, the effect sizes for all of the PROMs were greater than the effect size of every HRQoL score. In contrast to the large effect size for the EQ-5D (1.163) and the VAS QoL (1.833), the SF-6D had a moderate effect size of 0.530. There were larger standardized response means for the EQ-5D (1.228) and the VAS QoL (1.603) than the SF-6D (0.372). Therefore, the EQ-5D and VAS QoL were more responsive to change after anatomic total shoulder arthroplasty than the SF-6D.

Discussion

The findings of this study supported our hypotheses. The subjects experienced substantial and significant improvements in shoulder function and quality of life after anatomic total shoulder arthroplasty. The ~~HRQoL~~HRQoLs evaluated had variable validity and responsiveness to changes in outcome after anatomic total shoulder arthroplasty. The PROMs and the HRQoL scores that we studied ask very different types of questions and, for this reason, did not have strong positive correlations. For example, the ASES score, DASH score, SST, and VAS function solely address symptom and functional aspects related to the shoulder and upper extremity, the EQ-5D and VAS QoL assess more general function, and the SF-6D assesses more general function and mental health. Although the ASES score, DASH score, and VAS scores had significant positive correlations with ~~either or both of the EQ-5D and VAS QoL, there were no significant correlations with the SF-6D. Interestingly, the SST, which is often used to evaluate outcome after shoulder arthroplasty[25,26], did not significantly correlate with any of the HRQoL scores~~the EQ-5D, SF-6D, and VAS QoL, these correlations were generally weak.

Over the past 2 decades, there has been a dramatic increase in the utilization of shoulder ~~arthroplasty[27].~~arthroplasty[25]. The current focus on cost-effectiveness and value in response to rising utilization, increasing cost, and treatment innovation requires outcome assessments that can be used to compare treatment modalities for the same condition, as well as those of different conditions. HRQoL scores are necessary to determine cost-effectiveness of different treatments of various conditions. Although numerous studies have assessed patient-reported outcomes of anatomic total shoulder arthroplasty, few have considered HRQoL scores. The ideal HRQoL score should be valid and responsive and be convenient to administer, and it should be a universally applicable outcome score that can be used to compare benefit from different treatments for shoulder osteoarthritis as well as treatments for other conditions.

This study examined the validity and responsiveness of 3 commonly used HRQoL scores (EQ-5D, SF-6D, and VAS QoL). Although previous studies have examined the validity and responsiveness of HRQoL measures in shoulder surgical procedures, to our knowledge, none have compared the EQ-5D, SF-6D, and VAS QoL. With regard to validity, we found that PROMs commonly used to assess shoulder arthroplasty have varying correlations with commonly used HRQoL scores. The ~~change~~changes in ~~SF-6D was not~~the HRQoL scores were significantly correlated with the ~~change~~changes in ~~any~~all of the PROMs, suggesting ~~limited, if any,~~ validity. ~~The changes in EQ-5D and VAS QoL had significant weak and moderate positive correlations with the changes in~~ of these HRQoL scores for the ~~ASES score, DASH score, VAS pain, and VAS function and, thus, greater validity than~~evaluation of the ~~SF-6D~~outcomes of anatomic total shoulder arthroplasty. Nevertheless, these correlations were not strong, demonstrating that they do not assess the same aspects of outcome as the PROMs.

The ~~EQ-5D and~~ VAS QoL had not only greater validity, but also more responsiveness to the change in outcome after anatomic total shoulder arthroplasty than ~~the SF-6D. The effect size of the SF-6D was moderate at 0.53, in contrast to the large effect sizes of the EQ-5D (1.16) and VAS QoL (1.83).~~did the EQ-5D and SF-6D. It is not unexpected that the HRQoL scores had different degrees of validity and responsiveness as they assess different aspects of quality of life and determine preference weightings with different methods. The SF-6D emphasizes non-physical aspects (mental health and psychological parameters) and uses the standard gamble technique to determine preferences, and the EQ-5D emphasizes physical aspects and uses the time trade-off technique to determine preferences. As anatomic total shoulder arthroplasty is

primarily performed to reduce shoulder pain and improve function, the findings of this study are not unexpected.

The VAS QoL score differs considerably from other outcome measures. VAS scores provide a simple method of capturing a patient's perceived quality of life and have been extensively used in assessing shoulder ~~conditions[28-30]~~conditions[26-28]. This study found that the VAS QoL had greater validity and responsiveness compared with the EQ-5D and SF-6D. Although there are limited data on the psychometric properties of VASs in shoulder surgical procedures, they have been shown to be reliable, valid, and sensitive measures in other areas of health ~~care[31-33]~~care[29-31]. The strength of the VAS QoL lies in its simplicity, reliance on individual patients' experience, and independence from specific parameters.

Of the HRQoL scores that we studied, the EQ-5D and SF-6D had ~~the least~~comparable validity and responsiveness in measuring health outcomes for patients with advanced glenohumeral osteoarthritis who are treated with anatomic total shoulder arthroplasty. This contrasts with some previous literature ~~on the~~that has shown variable efficacy of the SF-6D. The SF-6D appears to be an especially effective outcome measure for stable angina and cervical spine ~~pathology[34,35]~~pathology[32,33]. Relevant to shoulder conditions, Slobogean et al. reported that DASH and SF-6D scores had the best psychometric properties among patients with proximal humeral ~~fractures[36]~~fractures[34]. Burton et al. found that the SF-6D was more responsive than the EQ-5D to outcome changes after lower-limb reconstruction ~~surgery[37]~~surgery[35]. Other investigations of outcome measures in osteoarthritis, chronic low back pain, and rheumatoid arthritis have found the SF-6D and EQ-5D to be equally effective in measuring cost ~~utility[38-40]~~utility[36-38]. In a study of rheumatoid arthritis, Salaffi et al. found that the EQ-5D and SF-6D had equivalent discriminatory power, and that the EQ-5D more efficiently detected ~~change~~changes in health ~~status[41]~~status[39]. They further noted that median EQ-5D scores were significantly lower than median SF-6D scores in patients with worse health status, suggesting that these HRQoL scores assess quality of life differently and may not be ~~interchangeable[41]. As the SF-6D is derived from the SF-36, which assesses both physical function and mental health, the difference we found with regard to the SF-6D may relate to the fact that shoulder arthroplasty is highly successful in relieving chronic pain and physical dysfunction compared with these other conditions, which not only are chronic but also are not necessarily eliminated by the treatment[42].~~interchangeable[39].

It remains unclear if there are clinically relevant differences in the validity and effectiveness between the SF-6D and EQ-5D, despite the dissimilar methods for determining the HRQoL scores, the distinct questions, and the use of the standard gamble compared with the time trade-off. As the SF-6D is derived from the SF-36, which assesses both physical function and mental health, and the EQ-5D is more focused on functional activities, variations are expected, perhaps depending on the clinical problem or diagnosis and treatment[40]. The EQ-5D had significant correlations with the PROMs that demonstrated validity, as well as large effect size and standardized response mean values demonstrating responsiveness to ~~change~~changes in health. Although the EQ-5D has been used to assess outcomes in patients with a variety of shoulder conditions including reverse total shoulder arthroplasty, hemiarthroplasty, and anatomic total shoulder ~~arthroplasty[43-45]~~arthroplasty[41-43], to our knowledge, it has not been compared with other HRQoL measures in anatomic total shoulder arthroplasty. A systematic review of 23 studies that used the EQ-5D as an outcome measure for quality of life in shoulder conditions

found that it had at least moderate responsiveness and good validity and reliability[46].reliability[44]. Further, it was especially effective at recognizing improvement in shoulder conditions postoperatively in patients without trauma compared with those with traumatic injury[46]injury[44].

This study had several limitations. The study cohort was relatively small, which limited the validity of the statistical analysis. This was addressed through a post hoc power analysis to determine the correlation coefficients required for significance. The findings may not be generalizable as there may have been selection bias for treatment in the treating surgeon's practice. Future work should attempt to draw subjects from varying regions to include a broader spectrum of patients, surgeons, and health systems. The use of 1-year follow-up outcome data has limitations. However, at least 1 other study demonstrated that patients undergoing shoulder arthroplasty achieve much of their outcome at 6 months postoperatively and become stable at the 1-year follow-up[47]up[45]. Limiting the follow-up to 1 year after the surgical procedure was thought to minimize the effect of other conditions and events on the outcome and quality-of-life assessment. Although we did not have 1-year follow-up data on all subjects, comparison of the subjects without follow-up only identified a difference in the baseline DASH score. and SF-6D score. We acknowledge that this comparison of subjects with and without 1-year follow-up may be underpowered. Lastly, this study did not consider other HRQoL scores such as the PROMIS (Patient-Reported Outcomes Measurement Information System) score[48]score[46]. The PROMIS score uses computer adaptive testing to decrease assessment burden[44]burden[42]. Initial studies have demonstrated that, despite having fewer questions, the PROMIS score correlates well with EQ-5D and DASH scores and may even outperform ASES scores and the SST[48-50]SST[46-48].

In summary, this study found that patients with advanced glenohumeral osteoarthritis who are treated with anatomic total shoulder arthroplasty have substantial and significant improvements in HRQoL scores and PROMs. The VAS QoL and EQ-5D had greater validity and responsiveness to change compared with the EQ-5D and SF-6D. Based on these findings, the VAS QoL and EQ-5D appear tomay be a better HRQoL measuresmeasure for the evaluation of the outcome of anatomic total shoulder arthroplasty for the treatment of glenohumeral osteoarthritis compared with the ED-5Q and SF-6D and may be preferablepreferred for cost-effectiveness and value analysis of this procedure. The increased validity and responsiveness of the EQ-5D in comparison with theand SF-6D supportwere similar despite the idea that HRQoL scores that assess physical function (e.g., EQ-5D) have more relevance in the assessment of shoulder arthroplasty outcomedifferences in their derivation. Lastly, the data with regard to the HRQoL scores related to the treatment of advanced glenohumeral osteoarthritis with anatomic total shoulder arthroplasty derived in this study provide information that can be used for future cost-effectiveness and value analysis.

References

1. Wong SE, Zhang AL, Berliner JL, Ma CB, Feeley BT. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. J Shoulder Elbow Surg. 2016 Jun;25(6):913-9. Epub 2016 Mar 31.

2. Tashjian RZ, Hung M, Keener JD, Bowen RC, McAllister J, Chen W, Ebersole G, Granger EK, Chamberlain AM. Determining the minimal clinically important difference for the American Shoulder and Elbow Surgeons score, Simple Shoulder Test, and visual analog scale (VAS) measuring pain after shoulder arthroplasty. J Shoulder Elbow Surg. 2017 Jan;26(1):144-8. Epub 2016 Aug 18.

3. Torrens C, Guirro P, Santana F. The minimal clinically important difference for function and strength in patients undergoing reverse shoulder arthroplasty. J Shoulder Elbow Surg. 2016 Feb;25(2):262-8. Epub 2015 Sep 28.

4. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health. 2009 Mar;12(Suppl 1):S5-9.

5. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. Value Health. 2009 Mar;12(Suppl 1):S27-30.

6. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96(1):5-21. Epub 2010 Oct 29.

7. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990 Dec;16(3):199-208.

8. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271-92.

9. Shmueli A. The visual analog rating scale of health-related quality of life: an examination of end-digit preferences. Health Qual Life Outcomes. 2005 Nov 14;3:71.

10. Hernández Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. Rheumatology (Oxford). 2013 May;52(5):944-50. Epub 2013 Jan 21.

11. Obradovic M, Lal A, Liedgens H. Validity and responsiveness of EuroQol-5 Dimension (EQ-5D) versus Short Form-6 Dimension (SF-6D) questionnaire in chronic pain. Health Qual Life Outcomes. 2013 Jul 1;11:110.

12. Søgaard R, Christensen FB, Videbaek TS, Bünger C, Christiansen T. Interchangeability of the EQ-5D and the SF-6D in long-lasting low back pain. Value Health. 2009 Jun;12(4):606-12. Epub 2008 Nov 10.

13. Carter MJ, Mikuls TR, Nayak S, Fehringer EV, Michaud K. Impact of total shoulder arthroplasty on generic and shoulder-specific health-related quality-of-life measures: a systematic literature review and meta-analysis. J Bone Joint Surg Am. 2012 Sep 5;94(17):e127.

14. Cho CH, Song KS, Hwang I, Coats-Thomas MS, Warner JJP. Changes in psychological status and health-related quality of life following total shoulder arthroplasty. J Bone Joint Surg Am. 2017 Jun 21;99(12):1030-5.

15. StataCorp. Stata statistical software: release 13. College Station, TX: StataCorp; 2013.

16. Hunter JE, Schmidt FL. Methods of meta-analysis: correcting error and bias in research findings. 2nd ed. Thousand Oaks: Sage; 2004.

17. Evans J. Straightforward statistics for the behavioral sciences. Pacific Grove: Brooks/Cole Publishing; 1996.

18. 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 Mar 15;59(3):391-8.

19. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis indexOsteoarthritis Index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis. 2001 Sep;60(9):834-40.

20. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials. 1991 Aug;12(4)(Suppl):142S-58S.

21. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989 Mar;27(3)(Suppl):S178-89.

22. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care. 1990 Jul;28(7):632-42.

23. Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV. The handbook of research synthesis. New York: Russel Sage Foundation; 1994. p 231-44.

24. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. London: Routledge; 1988.

25. Matsen FA 3rd, Warme WJ, Jackins SE. Can the ream and run procedure improve glenohumeral relationships and function for shoulders with the arthritic triad? Clin Orthop Relat Res. 2015 Jun;473(6):2088-96. Epub 2014 Dec 9.

26. Matsen FA 3rd, Russ SM, Vu PT, Hsu JE, Lucas RM, Comstock BA. What factors are predictive of patient-reported outcomes? A prospective study of 337 shoulder arthroplasties. Clin Orthop Relat Res. 2016 Nov;474(11):2496-510. Epub 2016 Jul 25.

27. Kim SH, Wise BL, Zhang Y, Szabo RM. Increasing incidence of shoulder arthroplasty in the United States. J Bone Joint Surg Am. 2011 Dec 21;93(24):2249-54.

Copyright © by The Journal of Bone and Joint Surgery, Incorporated
Jain et al.
Patient-Reported Outcome Measures and Health-Related Quality-of-Life Scores of Patients Undergoing Anatomic
Total Shoulder Arthroplasty
http://dx.doi.org/10.2106/JBJS.20.00402
Page 9

26. Edwards SL, Lee JA, Bell JE, Packer JD, Ahmad CS, Levine WN, Bigliani LU, Blaine TA. Nonoperative treatment of superior labrum anterior posterior tears: improvements in pain, function, and quality of life. Am J Sports Med. 2010 Jul;38(7):1456-61. Epub 2010 Jun 3.

27. Bagheri F, Ebrahimzadeh MH, Moradi A, Bidgoli HF. Factors associated with pain, disability and quality of life in patients suffering from frozen shoulder. Arch Bone Jt Surg. 2016 Jun;4(3):243-7.

28. Adey-Wakeling Z, Crotty M, Shanahan EM. Suprascapular nerve block for shoulder pain in the first year after stroke: a randomized controlled trial. Stroke. 2013 Nov;44(11):3136-41. Epub 2013 Aug 22.

29. Abend R, Dan O, Maoz K, Raz S, Bar-Haim Y. Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. J Behav Ther Exp Psychiatry. 2014 Dec;45(4):447-53. Epub 2014 Jun 18.

30. Adamchic I, Langguth B, Hauptmann C, Tass PA. Psychometric evaluation of visual analog scale for the assessment of chronic tinnitus. Am J Audiol. 2012 Dec;21(2):215-25. Epub 2012 Jul 30.

31. Lukacz ES, Lawrence JM, Burchette RJ, Luber KM, Nager CW, Buckwalter JG. The use of visual analog scale in urogynecologic research: a psychometric evaluation. Am J Obstet Gynecol. 2004 Jul;191(1):165-70.

32. Wu J, Han Y, Zhao FL, Zhou J, Chen Z, Sun H. Validation and comparison of EuroQoL-5 Dimension (EQ-5D) and Short Form-6 Dimension (SF-6D) among stable angina patients. Health Qual Life Outcomes. 2014 Oct 25;12(1):156.

33. Chotai S, Parker SL, Sivaganesan A, Godil SS, McGirt MJ, Devin CJ. Quality of life and general health after elective surgery for cervical spine pathologies: determining a valid and responsive metric of health state utility. Neurosurgery. 2015 Oct;77(4):553-60; discussion 560.

34. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. J Shoulder Elbow Surg. 2010 Apr;19(3):342-8. Epub 2010 Mar 1.

35. Burton M, Walters SJ, Saleh M, Brazier JE. An evaluation of patient-reported outcome measures in lower limb reconstruction surgery. Qual Life Res. 2012 Dec;21(10):1731-43. Epub 2011 Dec 18.

36. Zhang F, Yang Y, Huang T, Zhang Y, Zhao L, Li S. Is there a difference between EQ-5D and SF-6D in the clinical setting? A comparative study on the quality of life measured by AIMS2-SF, EQ-5D and SF-6D scales for osteoarthritis patients. Int J Rheum Dis. 2018 Jun;21(6):1185-92. Epub 2016 Jul 27.

37. Dritsaki M, Petrou S, Williams M, Lamb SE. An empirical evaluation of the SF-12, SF-6D, EQ-5D and Michigan Hand Outcome Questionnaire in patients with rheumatoid arthritis of the hand. Health Qual Life Outcomes. 2017 Jan 24;15(1):20.

38. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, Lee MJ. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine (Phila Pa 1976). 2011 Oct 1;36(21)(Suppl):S54-68.

39. Salaffi F, Carotti M, Ciapetti A, Gasparini S, Grassi W. A comparison of utility measurement using EQ-5D and SF-6D preference-based generic instruments in patients with rheumatoid arthritis. Clin Exp Rheumatol. 2011 Jul-Aug;29(4):661-71. Epub 2011 Aug 31.

40. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol. 1998 Nov;51(11):1115-28.

41. Renfree KJ, Hattrup SJ, Chang YHH. Cost utility analysis of reverse total shoulder arthroplasty. J Shoulder Elbow Surg. 2013 Dec;22(12):1656-61. Epub 2013 Oct 14.

42. Dowdle SB, Glass N, Anthony CA, Hettrich CM. Use of PROMIS for patients undergoing primary total shoulder arthroplasty. Orthop J Sports Med. 2017 Sep 15;5(9):2325967117726044.

43. Levine WN, Fischer CR, Nguyen D, Flatow EL, Ahmad CS, Bigliani LU. Long-term follow-up of shoulder hemiarthroplasty for glenohumeral osteoarthritis. J Bone Joint Surg Am. 2012 Nov 21;94(22):e164.

44. Grobet C, Marks M, Tecklenburg L, Audigé L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. Arch Orthop Trauma Surg. 2018 Jul;138(7):953-61. Epub 2018 Apr 13.

45. Simovitch RW, Friedman RJ, Cheung EV, Flurin PH, Wright T, Zuckerman JD, Roche C. Rate of improvement in clinical outcomes with anatomic and reverse total shoulder arthroplasty. J Bone Joint Surg Am. 2017 Nov 1;99(21):1801-11.

46. Beckmann JT, Hung M, Bounsanga J, Wylie JD, Granger EK, Tashjian RZ. Psychometric evaluation of the PROMIS Physical Function Computerized Adaptive Test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. J Shoulder Elbow Surg. 2015 Dec;24(12):1961-7. Epub 2015 Aug 28.

47. Hung M, Voss MW, Bounsanga J, Gu Y, Granger EK, Tashjian RZ. Psychometrics of the Patient-Reported Outcomes Measurement Information System Physical Function instrument administered by computerized adaptive testing and the Disabilities of Arm, Shoulder and Hand in the orthopedic elbow patient population. J Shoulder Elbow Surg. 2018 Mar;27(3):515-22. Epub 2018 Jan 9.

48. Saad MA, Kassam HF, Suriani RJ Jr, Pan SD, Blaine TA, Kovacevic D. Performance of PROMIS Global-10 compared with legacy instruments in patients with shoulder arthritis. J Shoulder Elbow Surg. 2018 Dec;27(12):2249-56. Epub 2018 Jul 20.

49. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care. 2004 Sep;42(9):851-9.

50. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 Nov;35(11):1095-108.

51. Hudak PL, Amadio PC, Bombardier C; The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder and Hand) [corrected]. Am J Ind Med. 1996 Jun;29(6):602-8. Erratum in: Am J Ind Med. 1996 Sep;30(3):372.

52. Lippitt SB, Harryman DT II, Matsen FA III. A practical tool for evaluation function: the Simple Shoulder Test. In: Matsen FA III, Fu FH, Hawkins RJ, editors. The shoulder: a balance of mobility and stability. Rosemont, IL: American Academy of Orthopaedic Surgeons; 1993. p 503-18.

53. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, Iannotti JP, Mow VC, Sidles JA, Zuckerman JD. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg. 1994 Nov;3(6):347-52. Epub 2009 Feb 13.

54. The University of Sheffield. Measuring and valuing health: SF-6D. https://www.sheffield.ac.uk/scharr/sections/heds/mvh/sf-6d. Accessed 2019 May 30.

TABLE I Description of Each Measure Included in the Study[42,51-56]Study[40,49-54]

Measure	Description
PROMs	
ASES score	A score based on 10 functional Likert format questions that comprise 50% of the score and a VAS pain score that is 50%; measured on a scale from 0 to 100, with 100 representing full function
DASH score	A 30-question Likert format; measured on a scale of 0 to 100, with 100 indicating complete disability
SST	A 12-question binary (yes-or-no) test that is usually reported as the number of yes responses; for statistical analysis, we converted it to a scale of 0 to 100, with 0 representing complete inability and 100 representing 100% function
VAS function	A VAS from 0 to 100, with 0 representing no loss of shoulder function and 100 representing a complete loss of shoulder function
VAS pain	A VAS from 0 to 100, with 0 representing no shoulder pain and 100 representing disabling shoulder pain
HRQoL scores	
EQ-5D	A time trade-off score measured from 0 to 1, with 1 representing excellent quality of life
SF-6D	A standard gamble score measured from 0 to 1, with 1 representing excellent quality of life
VAS QoL	A VAS from 0 to 100, with 0 representing excellent quality of life and 100 representing worst quality of life

TABLE II Baseline Comparison of Subjects with and without Follow-up

Characteristics	Subjects with 1-Year Follow-up (N = ~~93~~92)	Subjects without 1-Year Follow-up (N = ~~52~~51)
Demographic		
Age*† *(yr)*	69.~~25~~24 (67.~~52~~51 to 70.~~99~~96)	69.~~24~~19 (66.~~32~~29 to 72.~~15~~09)
Sex		
Male		
Patients‡	49	28
Percentage*†	~~52.7 (42.4~~53.3 (43.0 to 63.~~0~~5)	~~53.8 (39.8~~54.9 (41.1 to ~~67.9~~68.7)
Female	~~44~~43	~~24~~23
Arm affected		
Dominant		
Patients‡	50	~~30~~29
Percentage*†	~~53.8 (43.4~~54.3 (44.1 to 64.~~1~~6)	~~57.7~~56.9 (43.~~8~~1 to ~~71~~70.6)
Nondominant‡	~~43~~42	22
BMI*† *(kg/m²)*	29.~~78~~76 (28.~~60~~59 to 30.~~95~~94)	28.~~08~~13 (26.~~48~~53 to 29.~~70~~73)
PROMs* *(points)*		
ASES score†	31.~~4~~3 (28.~~2~~1 to 34.~~7~~5)	~~33.0~~32.2 (27.~~6~~0 to ~~38~~37.4)
DASH score§	42.~~1~~2 (38.~~7~~8 to 45.~~6~~7)	~~50.8~~51.1 (46.~~1~~4 to 55.~~6~~8)
SST†#	27.~~2~~4 (23.~~5~~7 to 31.~~0~~1)	26.~~4~~0 (20.~~8~~5 to ~~32.0~~31.6)
VAS function†	74.~~25~~63 (70.~~47~~93 to 78.~~03~~32)	~~75.53 (70.46~~76.91 (72.69 to ~~80.60~~81.13)
VAS pain†	~~69.92 (65.97~~70.25 (66.36 to ~~73.87~~74.13)	~~73.38 (67.64~~74.42 (69.11 to 79.~~12~~97)
HRQoL scores~~*‡~~* *(points)*		
EQ-5D~~‡~~†	0.~~667~~665 (0.~~628~~626 to 0.~~705~~703)	0.~~586~~582 (0.~~527~~524 to 0.~~645~~641)
SF-6D~~§~~§	0.~~647~~645 (0.~~634~~622 to 0.~~661~~645)	0.~~627~~596 (0.~~607~~572 to 0.~~647~~619)
VAS QoL~~‡~~†	57.19 (52.06 to 62.32)	64.27 (58.25 to 70.33)

*The values are given as the mean, with the 95% CI in parentheses. †The ~~95%~~ CIs for these categories all overlap, suggesting a lack of significant difference in the values between the 2 groups. ‡The values are given as the number of patients. §~~This category has~~These categories have the only 95% CIs without overlap between those with and without follow-up. #There were ~~2~~3 yes responses for each group.

TABLE III Paired, 2-Sided T Test Results Comparing the Means of Each Measure at Baseline and 1 Year Postoperatively

	Baseline*	1 Year*	Change*†	T Statistic	P Value‡
PROMs *(points)*					
ASES score	31 (28 to 34)	89 (86 to 92)	~~57~~58 (54 to ~~61~~62)	~~29.5~~30.8	<0.001
DASH score	42 (39 to 46)	10 (7 to ~~13~~12)	32 (29 to ~~35~~36)	−20	<0.001
SST§	27 (~~23~~24 to 31)	85 (81 to 89)	58 (~~54~~53 to ~~63~~62)	25.~~7~~4	<0.001
VAS function	74.~~3~~6 (70.~~5~~7 to 78.~~0~~3)	10.~~3~~1 (6.~~46~~23 to 14.~~1~~0)	~~63.9 (58.8~~64.5 (59.4 to 69.~~1~~6)	−25.~~6~~2	<0.001
VAS pain	~~69.9~~70.3 (66.~~0~~4 to ~~73.9~~74.1)	9.~~34~~11 (5.~~77~~54 to 12.~~9~~7)	~~60.6 (55.7~~61.1 (56.3 to 65.~~5~~9)	−~~24.7~~25.3	<0.001
HRQoL scores *(points)*					
EQ-5D	0.~~667~~665 (0.63 to 0.~~71~~70)	0.~~887~~888 (0.86 to 0.~~91~~92)	0.~~221~~223 (0.18 to 0.26)	11.~~8~~	<0.001
SF-6D	0.~~647~~645 (0.~~63~~62 to 0.~~66~~67)	0.~~682~~769 (0.~~67~~74 to 0.~~70~~79)	0.~~035~~124 (0.~~02~~10 to 0.~~05~~15)	~~3.56~~10.4	<0.001
VAS QoL	57.~~2~~4 (52.~~1~~3 to 62.~~3~~5)	11.~~5~~3 (7.~~77~~55 to 15.~~3~~1)	~~45.6 (39.8~~46.02 (40.1 to 51.~~5~~9)	−15.~~5~~6	<0.001

*The values are given as the mean, with the 95% CI in parentheses. †Note that because the change in DASH and VAS scores was negative and negative changes indicated improvement in these measures, the absolute value of those changes is shown so that positive changes reflect improvement in the each of the measures. ‡Every measure showed significant improvement from baseline (p < 0.00625). §There were 3 yes responses at baseline and 10 yes responses at 1 year, a change of 7 yes responses.

JAIN ET AL.
PATIENT-REPORTED OUTCOME MEASURES AND HEALTH-RELATED QUALITY-OF-LIFE SCORES OF PATIENTS UNDERGOING ANATOMIC TOTAL SHOULDER ARTHROPLASTY
http://dx.doi.org/10.2106/JBJS.20.00402
Page 14

TABLE IV Bivariate Pearson Correlations of the Changes in Measures

	EQ-5D		SF-6D		VAS QoL	
	Correlation Coefficient*	P Value	Correlation Coefficient*	P Value	Correlation Coefficient*	P Value
ASES score	0.~~243~~213 (weak)	0.~~0191~~0411	0.~~199 (very~~247 (weak)	0.~~0563~~018	0.~~361~~340 (weak)	0.~~0004~~00009†
DASH score	0.~~329~~320 (weak)	0.~~0013~~00019†	0.~~147 (very~~322 (weak)	0.~~1597~~0002†	0.~~371~~363 (weak)	0.~~0003~~00004†
SST	0.~~257~~266 (weak)	0.~~0128~~0103	0.~~096 (very~~298 (weak)	0.~~3625~~004	0.~~171~~178 (very weak)	0.~~102~~089
VAS function	0.~~254~~231 (weak)	0.~~0141~~0267	0.~~172~~155 (very weak)	0.~~1~~139	0.~~530~~518 (moderate)	<0.0001†
VAS pain	0.~~324~~304 (weak)	0.~~0015~~00032†	0.~~223~~283 (weak)	0.~~031~~3006	0.~~584~~574 (moderate)	<0.0001†

*Using classifications reported by Evans, correlation magnitudes were designated as very weak (r = 0.00 to 0.19), weak (r = 0.20 to 0.39), moderate (r = 0.40 to 0.59), strong (r = 0.60 to 0.79), or very strong (r = 0.80 to 1.00). †Significant correlation with a significance threshold of p < 0.0033 determined by using the Bonferroni correction of 0.05 divided by the number of regressions (15).

TABLE V Effect Size and Standardized Response Mean for PROMs and HRQoL Scores

	ASES	DASH	SST	VAS Function	VAS Pain	EQ-5D	SF-6D	VAS QoL
Effect size*	3.~~679~~705	1.~~928~~926	3.~~187~~178	3.~~473~~567	3.~~156~~213	1.~~163~~186	~~0.530~~1.084	1.~~833~~843
Standardized response mean	3.~~053~~219	2.~~077~~084	2.~~673~~656	2.~~546~~624	2.~~557~~635	1.~~228~~230	~~0.372~~1.083	1.~~603~~622

*Using the Cohen criteria, an effect size of <0.50 was deemed small, an effect size of ≥0.50 to 0.80 was deemed moderate, and an effect size of >0.80 was deemed large. All effect sizes were large~~, except for SF-6D, which had a moderate effect size.~~.