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 445143 patients (445143 shoulders) with glenohumeral osteoarthritis treated with anatomic total shoulder arthroplasty; 9392 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 large effect size in the VAS QoL (1.833843), 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 (1.228083), 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 Level IV. 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. 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. 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. The EuroQol-5 Dimensions (EQ-5D), Short Form-6 Dimensions (SF-6D), and Visual Analog Scale Quality of Life (VAS QoL) are the commonly used health utility indices. These metrics have not been widely used to assess HRQoL in patients with glenohumeral osteoarthritis.

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. 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.

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-three patients (145 shoulders), 77 male and 68 female, were treated by the senior author with a press-fit, short-stem anatomic humeral component (AEQUALIS ASCEND FLEXConvertible Shoulder Arthroplasty System; Wright Medical Group) and a hybrid, all-polyethylene, pegged glenoid (AFFINITI CORTILOC Glenoid Implant; Wright Medical Group). Ninety-three 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, 421 (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 of with 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 \((r = 0.00 \text{ to } 0.19)\), weak \((r = 0.20 \text{ to } 0.39)\), moderate \((r = 0.40 \text{ to } 0.59)\),
strong \((r = 0.60 \text{ to } 0.79)\), or very strong \((r = 0.80 \text{ to } 1.00)\)\(^\text{17}\).

The responsiveness of the PROMs and HRQoL scores was assessed by the effect size and
the standardized response mean\(^\text{18}\). 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\(^\text{18-23}\). Using the Cohen
criteria, an effect size of <0.5 was deemed small, ≥0.50 to 0.80 was deemed moderate, and
>0.80 was deemed large\(^\text{21,24}\).

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 not 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 \(r = 0.3287320\); \(p =
0.0043019\) and the VAS pain score \(r = 0.3240304\); \(p = 0.00450032\). The VAS QoL was
significantly and weakly correlated with the ASES score \(r = 0.3640340\); \(p = 0.00040009\) 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, 3.487178 for the SST, 3.473567 for VAS function,
and 3.56213 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 scores 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, 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. 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 changes in SF-6D was no significant correlated with the changes in any 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 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). 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. 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. 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. 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. Burton et al. found that the SF-6D was more responsive than the EQ-5D to outcome changes after lower-limb reconstruction surgery. 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. 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 changes in health status. 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. 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.

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. 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 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, 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...
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 to may be a better HRQoL measure 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 preferable for cost-effectiveness and value analysis of this procedure. The increased validity and responsiveness of the EQ-5D in comparison with the SF-6D support the idea that HRQoL scores that assess physical function (e.g., EQ-5D) have more relevance in the assessment of shoulder arthroplasty outcomes despite the idea that 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


<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROMs</td>
<td></td>
</tr>
<tr>
<td>ASES score</td>
<td>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</td>
</tr>
<tr>
<td>DASH score</td>
<td>A 30-question Likert format; measured on a scale of 0 to 100, with 100 indicating complete disability</td>
</tr>
<tr>
<td>SST</td>
<td>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</td>
</tr>
<tr>
<td>VAS function</td>
<td>A VAS from 0 to 100, with 0 representing no loss of shoulder function and 100 representing a complete loss of shoulder function</td>
</tr>
<tr>
<td>VAS pain</td>
<td>A VAS from 0 to 100, with 0 representing no shoulder pain and 100 representing disabling shoulder pain</td>
</tr>
<tr>
<td>HRQoL scores</td>
<td></td>
</tr>
<tr>
<td>EQ-5D</td>
<td>A time trade-off score measured from 0 to 1, with 1 representing excellent quality of life</td>
</tr>
<tr>
<td>SF-6D</td>
<td>A standard gamble score measured from 0 to 1, with 1 representing excellent quality of life</td>
</tr>
<tr>
<td>VAS QoL</td>
<td>A VAS from 0 to 100, with 0 representing excellent quality of life and 100 representing worst quality of life</td>
</tr>
</tbody>
</table>
### TABLE II Baseline Comparison of Subjects with and without Follow-up

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Subjects with 1-Year Follow-up (N = 93)</th>
<th>Subjects without 1-Year Follow-up (N = 52)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age*† (yr)</td>
<td>69.25 (67.52 to 70.99)</td>
<td>69.24 (66.32 to 72.15)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male Patients‡</td>
<td>49</td>
<td>28</td>
</tr>
<tr>
<td>Percentage*†</td>
<td>52.7 (42.45% to 63.05%)</td>
<td>53.8 (39.85% to 67.96%)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arm affected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant Patients‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage*†</td>
<td>53.8 (44.45% to 64.46%)</td>
<td>57.7 (43.40% to 71.70%)</td>
</tr>
<tr>
<td>BMI† (kg/m²)</td>
<td>29.28 (28.60 to 30.95)</td>
<td>28.08 (26.48 to 30.73)</td>
</tr>
<tr>
<td>PROMs* (points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASES score†</td>
<td>31.43 (28.21 to 34.75)</td>
<td>33.03 (27.60 to 38.37)</td>
</tr>
<tr>
<td>DASH score§</td>
<td>42.12 (38.78 to 45.67)</td>
<td>50.85 (46.14 to 55.68)</td>
</tr>
<tr>
<td>SST†#</td>
<td>27.24 (23.52 to 31.91)</td>
<td>26.40 (20.48 to 32.03)</td>
</tr>
<tr>
<td>VAS function†</td>
<td>74.25 (70.47 to 78.04)</td>
<td>75.83 (70.46 to 80.69)</td>
</tr>
<tr>
<td>VAS pain†</td>
<td>69.92 (65.97 to 74.32)</td>
<td>73.38 (67.64 to 79.13)</td>
</tr>
<tr>
<td>HRQoL scores*‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ-5D†</td>
<td>0.67 (0.64 to 0.70)</td>
<td>0.64 (0.60 to 0.68)</td>
</tr>
<tr>
<td>SF-6D§</td>
<td>0.67 (0.64 to 0.70)</td>
<td>0.64 (0.60 to 0.68)</td>
</tr>
<tr>
<td>VAS QoL†</td>
<td>57.19 (52.06 to 62.32)</td>
<td>64.27 (58.25 to 70.33)</td>
</tr>
</tbody>
</table>

*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 the only 95% CIs without overlap between those with and without follow-up. #There were 23 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

<table>
<thead>
<tr>
<th>PROMs (points)</th>
<th>Baseline*</th>
<th>1 Year*</th>
<th>Change†</th>
<th>T Statistic</th>
<th>P Value‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASES score</strong></td>
<td>31 (28 to 34)</td>
<td>89 (86 to 92)</td>
<td>58 (54 to 61)</td>
<td>58 (54 to 61)</td>
<td>29.530.8</td>
</tr>
<tr>
<td><strong>DASH score</strong></td>
<td>42 (39 to 46)</td>
<td>10 (7 to 13.1)</td>
<td>32 (29 to 35.36)</td>
<td>32 (29 to 35.36)</td>
<td>-20</td>
</tr>
<tr>
<td><strong>SST§</strong></td>
<td>27 (23.24 to 31)</td>
<td>85 (81 to 89)</td>
<td>58 (54.53 to 63.62)</td>
<td>58 (54.53 to 63.62)</td>
<td>25.74</td>
</tr>
<tr>
<td><strong>VAS function</strong></td>
<td>74.26 (70.57 to 78.82)</td>
<td>10.31 (6.46.23 to 14.49)</td>
<td>63.9 (58.964.5 to 59.4 to 69.46)</td>
<td>63.9 (58.964.5 to 59.4 to 69.46)</td>
<td>-25.62</td>
</tr>
<tr>
<td><strong>VAS pain</strong></td>
<td>69.970.3 (66.04 to 23.974.1)</td>
<td>9.4411 (5.7254 to 12.97)</td>
<td>60.6 (55.261.1 to 56.3 to 65.89)</td>
<td>60.6 (55.261.1 to 56.3 to 65.89)</td>
<td>-24.725.3</td>
</tr>
<tr>
<td><strong>HRQoL scores (points)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ-5D</td>
<td>0.662665 (0.63 to 0.7170)</td>
<td>0.882798 (0.86 to 0.9192)</td>
<td>0.221223 (0.18 to 0.26)</td>
<td>0.221223 (0.18 to 0.26)</td>
<td>11.8</td>
</tr>
<tr>
<td>SF-6D</td>
<td>0.642645 (0.64 to 0.6667)</td>
<td>0.682769 (0.67 to 0.7920)</td>
<td>0.065124 (0.0410 to 0.0815)</td>
<td>0.065124 (0.0410 to 0.0815)</td>
<td>3.5610.4</td>
</tr>
<tr>
<td>VAS QoL</td>
<td>57.24 (52.41 to 62.25)</td>
<td>11.94 (7.7775 to 15.31)</td>
<td>45.64 (39.846.02 to 40.1 to 51.59)</td>
<td>45.64 (39.846.02 to 40.1 to 51.59)</td>
<td>-15.56</td>
</tr>
</tbody>
</table>

*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 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.
### TABLE IV Bivariate Pearson Correlations of the Changes in Measures

<table>
<thead>
<tr>
<th></th>
<th>EQ-5D</th>
<th>SF-6D</th>
<th>VAS QoL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation Coefficient*</td>
<td>P Value</td>
<td>Correlation Coefficient*</td>
</tr>
<tr>
<td>ASES score</td>
<td>0.2472 (weak)</td>
<td>0.0191 (very weak)</td>
<td>0.189 (very weak)</td>
</tr>
<tr>
<td>DASH score</td>
<td>0.429 (weak)</td>
<td>0.0013 (very weak)</td>
<td>0.442 (very weak)</td>
</tr>
<tr>
<td>SST</td>
<td>0.3572 (weak)</td>
<td>0.0128 (very weak)</td>
<td>0.096 (very weak)</td>
</tr>
<tr>
<td>VAS function</td>
<td>0.25821 (weak)</td>
<td>0.0141 (very weak)</td>
<td>0.473155 (very weak)</td>
</tr>
<tr>
<td>VAS pain</td>
<td>0.324304 (weak)</td>
<td>0.0015 (very weak)</td>
<td>0.223283 (very weak)</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>ASES</th>
<th>DASH</th>
<th>SST</th>
<th>VAS Function</th>
<th>VAS Pain</th>
<th>EQ-5D</th>
<th>SF-6D</th>
<th>VAS QoL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect size</strong>*</td>
<td>3.679705</td>
<td>1.928926</td>
<td>3.102179</td>
<td>3.423567</td>
<td>3.156213</td>
<td>1.463196</td>
<td>0.5401084</td>
<td>1.824843</td>
</tr>
<tr>
<td><strong>Standardized response mean</strong></td>
<td>3.053219</td>
<td>2.037004</td>
<td>2.673656</td>
<td>2.546624</td>
<td>2.557635</td>
<td>1.228230</td>
<td>0.3721083</td>
<td>1.603622</td>
</tr>
</tbody>
</table>

*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.*