The degree to which subjective patient-reported measures reflect objective findings or how well subjective and objective measures reflect patient satisfaction is not well established. The purpose of this study was to determine the correlation between such measures before and after shoulder arthroplasty. A group of 174 patients (93 total shoulder arthroplasty and 81 reverse shoulder arthroplasty) were prospectively evaluated pre- and postoperatively (mean follow-up, 49 months) with the following subjective measures: American Shoulder and Elbow Surgeons score, Simple Shoulder Test, Short Form 36 (SF-36) summary scores, and patient satisfaction. Objective measures included Biodex isometric strength and videotaped range of motion. The objective measures were combined to derive 1 number representative of the overall shoulder function. All measures improved from their preoperative statuses except the SF-36 physical component summary in patients undergoing revision and the SF-36 mental component summary in patients undergoing primary reverse shoulder arthroplasty. A patient satisfaction score of 5 or more was seen in 89% of patients. Preoperatively, a significant (P<.05) correlation existed between the American Shoulder and Elbow Surgeons and the Simple Shoulder Test (r=0.546), American Shoulder and Elbow Surgeons score and SF-36 physical component summary (r=0.407), and Simple Shoulder Test and SF-36 physical component summary (r=0.479). Objective measures had lower correlations (r<0.4) with subjective scores. Postoperatively, the correlation improved among all measures. Patient satisfaction correlated more with subjective than objective measures. Subjective measures had relatively low correlations with objective measures. Improvements in the current measures are necessary to provide evidence-based comparisons of the effectiveness of shoulder arthroplasty.
Clinical measures are frequently used to determine the effectiveness of surgical interventions, such as shoulder arthroplasty, in patients with advanced degenerative disease. Such assessments are critical in advancing the care of musculoskeletal conditions, allowing physicians to refine the procedures that have potential and moving beyond those that have less merit. However, the methods used to define a successful outcome are subject to some degree of contention and can vary. Traditional objective measures, such as strength and range of motion (ROM), have fallen out of favor with some physicians due to questions regarding their ability to accurately reflect patient-perceived outcome.\textsuperscript{1,2} The additional challenge in recording objective measures is the potential introduction of bias if treating physicians record the data. Patient-reported measures are thought to be more representative because they capture the social and personal context of the patient and the disease.\textsuperscript{3,4} However, the potential bias is shifted to the patient.

Although obvious merits exist in evaluating the patients’ perceptions, inherent limitations also exist. A conflicting body of literature exists for the performance of many of these measures, with previous studies demonstrating acceptable psychometric properties of many of the patient-derived scores. Kocher et al\textsuperscript{5} reported that the American Shoulder and Elbow Surgeons (ASES) score demonstrates sufficient responsiveness for use in multiple disease states, including osteoarthritis. However, another recent study raised concern regarding the clinometric properties of many of these scoring systems, including validity, reproducibility, responsiveness, and interpretability. Bot et al\textsuperscript{6} evaluated 16 shoulder questionnaires and reported that none demonstrated satisfactory results across all clinometric categories. They noted limitations in study designs that evaluated the validity of such measures in even the most extensively studied instruments, including the ASES score. In addition, they reported that clinometric properties might vary across settings, populations, and conditions, making the application of 1 scoring system to different disease states difficult.\textsuperscript{6} Other authors have reported similar difficulties regarding the universal applicability of 1 measure across multiple pathologic states despite their occurrence in a common anatomical region.\textsuperscript{7} Understanding how patients perceive and report their disability is critical to obtaining an accurate determination of treatment success. In particular, the psychological phenomenon of response shift, initially appreciated in chronic medical illnesses, is being increasingly recognized in chronic musculoskeletal conditions.\textsuperscript{8,11} Acceptance of chronic functional limitations might allow for higher patient satisfaction despite a persisting amount of disability, resulting in a discrepancy between patient-reported and objective measurements.

The introduction of an independent third-party evaluation of a patient’s disability by measuring objective outcomes, such as strength and ROM, may reduce the variability noted with patient questionnaires. These measures may be more indicative of the actual mechanical and functional change that occurs following surgery. However, relying solely on these measures to determine outcome shifts the focus from the patient to the shoulder and disregards subjective measures, such as pain, which is often the primary motivator for surgical treatment.

An improved understanding of how subjective and objective measures relate would allow for an enhanced determination of treatment success or failure in the currently available shoulder arthroplasty literature. The purpose of this study was to investigate the correlation between the commonly used subjective patient questionnaires (ASES, Simple Shoulder Test [SST], and Short Form 36 [SF-36]), and objective measures (by an independent third-party observer) in a shoulder arthroplasty population pre- and postoperatively.

The hypotheses were that objective and subjective measures would correlate well pre- and postoperatively and that postoperative patient satisfaction ratings would correlate with subjective and objective measures.

**Materials and Methods**

**Study Design and Patient Selection**

The procedures followed were approved by an independent institutional review board and were in accordance with the ethical standards of the responsible committee on human experimentation. All patients provided written informed consent for participation. These patients had participated in a prior prospective cohort study investigating objectively measured improvements in shoulder function after shoulder arthroplasty.\textsuperscript{12} Inclusion criteria were all patients undergoing anatomic total or reverse shoulder arthroplasty regardless of preoperative diagnosis between March 2004 and May 2006.

All patients had failed prior nonoperative treatment. Exclusion criteria were patients who were deceased, had relocated out of state, or could not be reached for follow-up. Also, in the case of bilateral shoulder arthroplasties, only the most recent shoulder was included.

**Patient Demographics**

A group of 174 patients (93 total shoulder arthroplasty, 81 reverse shoulder arthroplasty) fulfilled the inclusion criteria. Forty-one women and 52 men with a mean age of 66 years (range, 35-89 years) underwent total shoulder arthroplasty and had a mean follow-up of 48 months (range, 35-89 months). Fifty-five women and 26 men with a mean age of 69 years (range, 40-88 years) underwent reverse shoulder arthroplasty and had a mean follow-up of 48 months (range, 31-72 months) (Table 1).

**Data Collection**

Patients were prospectively evaluated using subjective patient-reported measures and objective strength and ROM measurements.
The initial assessment was performed at the routine preoperative visit 1 week preoperatively. Postoperative data collection was performed at a minimum 2-year follow-up. Patient-reported measures included the ASES score, SST, Patient Satisfaction on a 10-point visual analog scale (PTSAT), and version 2 of the SF-36.

Similar to the prior study, objective assessment included strength and ROM measurements. Isometric strength measurements were obtained for 6 shoulder positions using the Biodex System II dynamometer (Biodex Medical Systems, Shirley, New York). All measurements were performed by a physical therapist blinded to the purpose of the study. Range of motion measurements were obtained in 4 planes of motion (forward flexion, abduction, internal rotation, and external rotation) with the use of a digital goniometer on a standardized video-recorded ROM examination. An independent observer blinded to type of surgery and purpose of the study obtained the measurements. The preoperative video ROM examination was unavailable for 23 patients. Therefore, patient questionnaire data of the highest attainable ROM (indicated by selecting the position on a line drawing depicting varying amounts of ROM) was substituted. This method has been shown to have a high level of agreement with video-recorded ROM. The Objective Outcome Summary Score (OOSS) was developed for use in obtaining 1 number representative of overall functional status, including strength and ROM.

Statistical Methods
Pre- to postoperative change in subjective measures and the OOSS were assessed using the Wilcoxon signed-rank paired test. Patient groupings for analysis included primary total shoulder arthroplasty, primary reverse shoulder arthroplasty, and all revision procedures (Table 2). For the purposes of reporting, total (n=6) and reverse shoulder arthroplasty revisions (n=26) were grouped together because no difference existed when examining these populations independently.

Preoperative Correlation
Spearman’s ρ correlation coefficient was used to determine the extent of the correlation between the subjective and objective measures. The subjective measures, including ASES, SST, and SF-36 physical component summary (PCS) and mental component summary (MCS), were assessed for correlation with each other and with the OOSS pre- and postoperatively. The correlation of PTSAT scores was also evaluated with subjective and objective measures postoperatively. Patient satisfaction was graphically compared with pre- to postoperative change in all subjective and objective measures to determine the percentage of patients with a PTSAT of 5 or more who also had an improvement in outcome and the percentage of patients with a PTSAT less than 5 who also had a lack of improvement in outcome. Clinical significance was set at a P value less than .05.

RESULTS
Pre- Versus Postoperative Comparison
Pre- and postoperative outcome scores are shown in Table 2. A significant pre- to postoperative improvement occurred for nearly all measures for each group; only the SF-36 PCS in the revision group and the SF-36 MCS in the primary reverse shoulder arthroplasty group did not reach statistical significance.

Preoperative Correlation
Correlation among the subjective measures and between the subjective and objective measures in the preoperative setting are shown in Tables 3 and 4. All correlations were significant (P<.05). The highest correlation was between ASES and SST (ρ=0.546). The SF-36 PCS also correlated with ASES (ρ=0.407) and SST (ρ=0.479).
Compared with the correlation among the subjective scores, a weaker correlation existed between the OOSS and ASES ($\rho = 0.288$), SST ($\rho = 0.383$), SF-36 PCS ($\rho = 0.205$), and SF-36 MCS ($\rho = 0.193$).

**Postoperative Correlation**

Correlations among the postoperative measures are shown in Tables 3 and 4. All correlations were significant ($P < .05$). However, a strengthening in the degree of correlation occurred across all measures postoperatively, although the degree of change was not assessed for statistical significance. The highest correlation was between the ASES and SST ($\rho = 0.813$). The correlation of the SF-36 PCS with the ASES ($\rho = 0.537$) and SST ($\rho = 0.501$) also increased.

The correlation between the OOSS and the subjective measures increased for the ASES ($\rho = -0.566$) and SST ($\rho = -0.619$). However, compared with this increase, the correlation between the OOSS and SF-36 PCS ($\rho = -0.253$) and SF-36 MCS ($\rho = -0.245$) did not strengthen as much.

**Correlation With Patient Satisfaction**

Patient satisfaction had a higher correlation with subjective measures than with the OOSS ($\rho = -0.383$) (Table 5). The highest correlation of PTSAT was with ASES ($\rho = 0.589$), followed closely by SST ($\rho = 0.527$). The lowest correlations with PTSAT were with the SF-36 PCS ($\rho = 0.291$) and SF-36 MCS ($\rho = 0.385$).

Of 174 patients, 158 had both PTSAT and pre- and postoperative ASES available. Of these 158 patients, 144 had a PTSAT of 5 or more: 136 (94%) had an improved and 8 (6%) had an unchanged or diminished ASES score compared with the preoperative values. Similarly, 14 of 158 patients had a PTSAT less than 5: nine (64%) had an improved and 5 (36%) had an unchanged or diminished ASES compared with the preoperative values (Figure A). This trend was similar when graphically comparing PTSAT to change in SST (Figure B) and change in OOSS (Figure C). However, when graphically comparing PTSAT to the change in SF-36 PCS and MCS, a weaker association existed between a PTSAT of 5 or more and improvements in the scores (Figures D, E).

Of 154 patients who had a PTSAT of more than 5, fifty-two (34%) had improvements in all measures. Conversely, 1 of 15 (7%) patients with a PTSAT less than 5 had a lack of improvement in all measures.

**DISCUSSION**

Little consensus exists regarding the ideal method of determining and reporting clinical outcomes after shoulder arthroplasty. Patient-reported measures attempt to capture the surgical outcome from the patient’s perspective. However, with increasing recognition of the limitations of these scoring systems to accurately assess...
outcomes across multiple disease states, some renewed interest exists in using objective assessments. The current study examined the correlation between subjective measures (ASES, SST, and SF-36) and an objective measure (OOSS).  

The current literature is limited regarding the correlation of outcome data following shoulder arthroplasty. A previous study assessing the correlation of patient- and physician-assessed function after shoulder arthroplasty demonstrated a high level of agreement between patients and physicians. In that study, patients and physicians recorded the outcome by filling out different copies of the same questionnaire. The current study was unique because the authors investigated commonly used patient-reported questionnaires and objective measures obtained by third-party observers. The potential for increased correlation due to patient-physician communication has been minimized.

The current study found differences in the degree of correlation between the pre- and postoperative settings. The hypothesis was not supported preoperatively; little correlation seems to exist between how patients perceive their function and how well they perform in regard to objectively measured strength and ROM. Stronger correlations were observed postoperatively between subjective and objective measures. Although this change was not assessed for statistical significance, it is an interesting finding. Although the study design did not allow for a determination of the cause of this difference, it may suggest that the indications for treatment vary between patients and manifest as variability in scoring among the different measures preoperatively. However, the patients represent a more homogeneous group postoperatively. The postoperative results reflect a greater uniformity between the varying measures.

The lack of correlation between subjective and objective measures underscores the importance of understanding what each of these assessment methods captures. This study was not designed to determine why the discrepancy exists but rather to quantify the level of correlation. Patient-reported measures are likely influenced by multiple factors, only 1 of which is the efficacy of treatment. The level of expectations set by the surgeon, recall bias by the patient, and social context of the disease may be only a few of multiple variables that affect how a patient self-reports their pain and function. Although the ASES has had acceptable psychomet-

---

**Table 3**

Preoperative and Postoperative Correlation of Subjective Markers of Pain, Function, and Overall Health Status

<table>
<thead>
<tr>
<th>Score</th>
<th>Preoperative</th>
<th>Postoperative</th>
<th>Preoperative</th>
<th>Postoperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASES total</td>
<td></td>
<td></td>
<td>.546</td>
<td>.813</td>
</tr>
<tr>
<td>Simple Shoulder Test</td>
<td>.546</td>
<td>.813</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>.407</td>
<td>.537</td>
<td>.479</td>
<td>.501</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>.236</td>
<td>.337</td>
<td>.215</td>
<td>.321</td>
</tr>
</tbody>
</table>

Abbreviations: ASES, American Shoulder and Elbow Surgeons Score; MCS, mental component summary; PCS, physical component summary; SF-36, Short Form 36.

*Values reported as a correlation coefficient with all correlations reaching significance (P<.05).

**Table 4**

Correlation of Various Markers of Subjective Patient Pain, Function, and Overall Health Status with Preoperative and Postoperative Objective Outcome Summary Score

<table>
<thead>
<tr>
<th>Score</th>
<th>Preoperative</th>
<th>Postoperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASES total</td>
<td>-.288</td>
<td>-.566</td>
</tr>
<tr>
<td>Simple Shoulder Test</td>
<td>-.383</td>
<td>-.619</td>
</tr>
<tr>
<td>Short Form 36 PCS</td>
<td>-.205</td>
<td>-.253</td>
</tr>
<tr>
<td>Short Form 36 MCS</td>
<td>-.193</td>
<td>-.245</td>
</tr>
</tbody>
</table>

Abbreviations: ASES, American Shoulder and Elbow Surgeons Score; MCS, mental component summary; PCS, physical component summary.

*Values are reported as a correlation coefficient, with all correlations reaching significance (P<.05).

**Table 5**

Correlation of Postoperative Patient Satisfaction With the Subjective and Objective Outcome Measures

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Patient Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASES total</td>
<td>.589</td>
</tr>
<tr>
<td>SST</td>
<td>.527</td>
</tr>
<tr>
<td>OOSS</td>
<td>-.383</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>.291</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>.385</td>
</tr>
</tbody>
</table>

Abbreviations: ASES, American Shoulder and Elbow Surgeons Score; MCS, mental component summary; OOSS, Objective Outcome Summary Score; PCS, physical component summary; SF-36, Short Form 36; SST, Simple Shoulder Test.

*Measured on a scale of 1-10. Values are reported as a correlation coefficient, with all correlations reaching significance (P<.05).
ric properties for multiple disease states, including glenohumeral osteoarthritis, it may not be the most optimal measure for any 1 particular disease state. As a result, some of the observed lack of correlation may be due to the responsiveness of the measures to the pathology.

Many practitioners rely heavily on the PTSAT score to determine whether surgical intervention was successful. However, the satisfaction rating is subject to many of the same shortcomings as other patient-reported measures. In the current study, 34% of patients with a PTSAT of 5 or greater had improvements in all measures. Conversely, 7% of patients with a PTSAT less than 5 had a lack of improvement in all measures. This indicates the importance of looking at several additional markers when judging the success of an intervention. Patient satisfaction can be influenced by factors other than long-term subjective and objective improvements, such as emotional state, like or dislike of the physician, ongoing litigation, patient expectations, and worker’s compensation status. For example, patients with a failed arthroplasty revised multiple times may accept a lower ROM and less strength and be satisfied with pain relief alone.

In addition, PTSAT is better associated with changes in subjective and objective shoulder-specific measures than general health measures (eg, of patients with a PTSAT of 5 or more, 94% had an improvement in ASES, whereas 59% had an improvement in the SF-36 MCS) (Figures A, E). This suggests that a generic health measure is not as sensitive as shoulder-specific instruments to the effects of the procedures.

This study provides additional information on the clinical effectiveness of shoulder arthroplasty. Eighty-eight percent of patients had a PTSAT of 5 or more, and improvements were observed for subjective and objective measures. Patients who underwent total shoulder arthroplasty...
had the greatest overall OOS. The expla-
nation for the difference is the objective
strength data demonstrating that although
strength improves after reverse shoulder
arthroplasty, it does not recover to a simi-
lar level observed in the population un-
dergoing total shoulder arthroplasty. This
discrepancy is consistent with the notion
that patients undergoing reverse shoulder
arthroplasty have varying degrees of mus-
cular insufficiency, whereas patients un-
dergoing total shoulder arthroplasty only
have articular cartilage loss. Patients who
underwent revision shoulder arthroplasty
had more modest improvements.

One limitation of this study was the
lack of a validated system for calculating
1 number that represents overall objective
function. This study used the Florida Im-
pairment Guidelines because it is a purely
objective measure that does not include
patient-reported subjective measures. How-
ever, these guidelines give equal weight to
strength and ROM when determining the
overall function. As a result, it is not known
whether this provides an accurate overall
assessment of objective function. However,
in the absence of another available method,
this measure is an acceptable summariza-
tion of the objective data set. An additional
limitation was the inclusion of patients un-
dergoing both total and reverse shoulder ar-
throplasty. It is possible that the degree of
correlation in 1 group may differ from the
other. However, the focus of this study was
the subjective and objective assessment of
shoulder pain and function after arthro-
plasty regardless of the specific procedure
performed. Currently, no outcome instru-
ment that is specific to pathology exists. As
a result, incorporating all patients undergo-
ning shoulder arthroplasty best reflects the
current use of these measures.

This study did not demonstrate a high
level of correlation between the patient-
reported measures and the objective mea-
sure, particularly in the preoperative set-
ting. This may be due to an inadequate
sample size. No statistics were performed
prior to the study to determine the neces-
sary sample size. However, to the authors’
knowledge, this is the largest sample size
reported for this type of study in a shoul-
der arthroplasty population.

CONCLUSION

As an age of assessing cost-effectiveness of
care begins, it is critical to understand the
measures physicians use to determine a suc-
cessful outcome. This study highlights the
differences in using these various instru-
ments. Patient-reported measures did not
correlate with objective measures. This lack
of agreement does not necessarily indicate a
deficiency in 1 method or the other but may
reflect how these methods capture different
aspects of overall function. The develop-
ment of a universally applicable instrument
is not necessarily the solution because this is
neither feasible nor perhaps even desirable.
However, improvements in the current mea-
ures are necessary to provide evidence-
based comparisons of the clinical effective-
ness of total shoulder arthroplasty.

REFERENCES

1. Kirkley A, Alvarez C, Griffin S. The de-
velopment and evaluation of a disease-specific
quality-of-life questionnaire for disorders of
the rotator cuff: the Western Ontario Rotator
2. Kirkley A, Griffin S. Development of dis-
3. Higginson IJ, Carr AJ. Measuring quality of
life: using quality of life measures in the
4. Koran LM. The reliability of clinical meth-
5. Kocher MS, Horan MP, Briggs KK, Rich-
ardson TR, O’Holleran J, Hawkins RJ. Re-
liability, validity, and responsiveness of the
American Shoulder and Elbow Surgeons
Subjective Shoulder Scale in patients with
shoulder instability, rotator cuff disease, and
6. Bot SD, Terwee CB, van der Windt DA,
Bouter LM, Dekker J, de Vet HC. Clinimetric
evaluation of shoulder disability question-
naires: a systematic review of the literature.
7. Lippitt S, Harryman DT II, Matsen F. A prac-
tical tool for evaluating function: the simple
shoulder test. In: Matsen F, Fu F, Hawkins R, 
eds. *The Shoulder: A Balance of Mobility and
Stability.* Rosemont, IL: American Academy of
8. Finkelstein JA, Razmjou H, Schwartz CE.
Response shift and outcome assessment in
orthopedic surgery: is there a difference be-
tween complete and partial treatment? *J Clin
9. Razmjou H, Schwartz CE, Yee A, Finkelstein
JA. Traditional assessment of health outcome
following total knee arthroplasty was con-
fronted by response shift phenomenon. *J Clin
10. Razmjou H, Yee A, Ford M, Finkelstein JA.
Response shift in outcome assessment in pa-
ients undergoing total knee arthroplasty. *J
Bone Joint Surg Am.* 2006; 88:2590-2595.
11. Schwartz CE, Finkelstein JA. Understanding
inconsistencies in patient-reported outcomes
after spine treatment: response shift phenom-
12. Puskas B, Harrelk K, Clark R, Downes K,
Virani NA, Frankle M. Isometric strength,
rang-of-motion, and impairment before and
after total and reverse shoulder arthroplasty
[published online ahead of print January 9,
standardized method for the assessment of
14. Ware JE Jr, Sherbourne CD. The MOS 36-
item short-form health survey (SF-36). I. Con-
15. Mulieri P, Dunning P, Klein S, Pupello D,
Frankle M. Reverse shoulder arthroplasty for
the treatment of irreparable rotator cuff tear
without glenohumeral arthritis. *J Bone Joint
CM, Cummings JD, Coiffl RD. Patient and
physician-assessed shoulder function after ar-
throplasty. *J Bone Joint Surg Am.* 2006; 88:508-
513.